

POLITECNICO DI TORINO

Master's degree Thesis

Instrumentation biomedical engineering



Politecnico di Torino

Seeing beyond the surface: Quantifying Actinic Keratosis Progression with OCT and OCTA

Supervisors:

Prof. Kristen Mariko Meiburger

Ing. Giulia Rotunno

Prof. Mengyang Liu

Candidate:

Silisteanu Andrada

December 2024

ACKNOWLEDGMENTS

I would like to express my deepest gratitude
to my supervisor, Professor Kristen M. Meiburger, for giving me the opportunity to
work on this incredible thesis project.

to my co- supervisor, Engineer Giulia Rotunno, for guiding and supporting me in all
the choices, transferring all her precious knowledge to me: it was an honour to work
next to her.

to the whole team at the General Hospital in Vienna, for teaching me not only
everything there was to know about the world of OCT/OCTA, but also how
indispensable confrontation and teamwork are.

To my parents
thank you for not only giving me life, but also teaching me how to live it at its best.
Everything I do and will do will always be dedicated to you, with the hope of
making you proud every day.

To my boyfriend, my best friends, my long-standing buddies and my fellow travellers
Thank you for always believing in me, from start to finish, pushing me to aspire
high and never give up.

This project is for all of you.

ABSTRACT

Actinic keratosis, a precancerous skin lesion, can transform into squamous cell carcinoma if left untreated, passing through critical stages such as Bowen's disease. Diagnosis of actinic keratosis can be made clinically, by visually analyzing the shape and size of the lesion, or dermoscopically, by investigating the more superficial layer of the skin. However, since these are qualitative analyses, limited by the inability to explore the deeper layers of the skin, they cannot distinguish an actinic keratosis from a tumor. This is why biopsy, an invasive procedure that ensures certain diagnosis of cancerous and non-cancerous condition, is often used.

What if it were possible to diagnose actinic keratosis and follow its evolution without invasive intervention?

This thesis project proposes an alternative: it explores the use of OCT (optical coherence tomography) and OCTA (OCT angiography), non-invasive imaging techniques that, by exploiting light and the principle of interference, can capture high-resolution images and volumes of biological tissues structure and vasculature. The aim is to demonstrate how these techniques, already used in medical settings, can support the qualitative and quantitative distinction between actinic keratosis and its possible developmental stages.

Using a dataset of 13 patients recruited at the Vienna General Hospital, 48 OCT volumes were collected. From these, OCTA volumes were extracted and processed, including artifact reduction and adjustments to contrast and brightness to improve data visualization. Volume segmentation and subsequent skeletonization were then performed. The project runs through two pathways to achieve the presented goal: OCTA processing to obtain meaningful vascular parameters and OCT volume analysis to identify lesion depth. Considering that as the tumor stage progresses, the vascular network tends to be increasingly disorganized and to extend to the deep layers of the

skin, parameters such as tortuosity, vascular density, entropy, and fractal dimension were chosen to distinguish different vascular pattern, and were calculated separately for the superficial and deep layers, with the aim of demonstrating differences between the four patient groups: healthy, actinic keratosis, Bowen’s disease, and squamous cell tumor. To validate these differences, a statistical analysis, both on individual parameters and multiparametric, was carried out, and a logistic regression-based classifier was then introduced to predict the stage of pathology in output.

Parametric analysis effectively distinguished healthy from diseased groups, revealing statistically significant differences in parameters such as vascular density and entropy. However, in the differentiation between actinic keratosis and tumor, the results were partially limited by some critical issues, such as the unbalanced dataset and the presence of artifacts and noise. On the other hand, OCT volume analysis made it possible to determine the depth of lesions, showing that tumor lesions tend to involve deeper layers than non-tumor lesions, facilitating the distinction between actinic keratosis and later stages.

This thesis guides through each step from image acquisition to morphologic analysis of OCTA and OCT volumes, offering a noninvasive method to support the diagnosis and distinction between actinic keratoses and cancerous lesions.

TABLE OF CONTENTS

	Page
LIST OF TABLES	xi
LIST OF FIGURES	xii
CHAPTER	
PART I	2
1. ACTINIC KERATOSIS	2
1.1 About Actinic Keratosis	2
1.2 Causes and Risk Factors	3
1.3 Features and evolutions	4
1.4 Diagnosis	6
2. OCT SYSTEM	10
2.1 Overview of OCT Fundamentals and Methods	10
2.1.1 OCT Principles	10
2.1.2 OCT Techniques	13
2.2 Laboratory OCT System	16
2.2.1 Device components	16
2.2.2 Device's functioning	19
2.3 OCTA techniques	22
3. PURPOSE AND STRUCTURE OF THE THESIS	26
PART II	30
4. ACQUISITION PROTOCOL	30
5. DATA PROCESSING	34
5.1 Raw data processing	34

5.2	OCTA extraction	36
5.3	Artifact removal and image enhancement	40
5.3.1	White Line Correction	40
5.3.2	Projection artefact	41
5.3.3	Smoothing Filter and Contrast Enhancement	43
5.3.4	Depth Color Coding	45
5.4	Segmentation and skeletonization	48
5.4.1	Three-dimensional whole-volume processing	48
5.4.2	Superficial and deep layering for 2D analysis	52
6.	VASCULAR PARAMETERS EXTRACTION	56
6.1	Characteristics of the Tumor Vascular Network	56
6.2	Parameter List	59
6.3	3D Parameter Extraction	67
6.4	2D Parameters Extraction	70
6.4.1	Step1: Raw Dataset	74
6.4.2	Step2: Dataset and Data Deep Cleaning	75
6.4.3	Step3: Normalization with BodyIndex	77
6.4.4	Step4: Skeleton Improvement	80
7.	PARAMETER ANALYSIS	82
7.1	T-test and boxplot representation	82
7.2	MANOVA test	85
7.3	Logistic regression and multiclass ECOC classifier	88
8.	OCT PROCESSING	91
8.1	Lesion Depth Calculation	92
	PART III	96

9. RESULTS PRESENTATION	96
9.1 T-test	96
9.1.1 Constant Parameters	97
9.1.2 Range	100
9.1.3 Mean	106
9.1.4 Max	111
9.2 MANOVA-test	116
9.2.1 Range	118
9.2.2 Mean	121
9.2.3 Max	124
9.3 Binary and Multiclass classifiers	126
9.3.1 Range	128
9.3.2 Mean	129
9.3.3 Max	130
9.4 Depth of lesions	132
9.5 Combined Analysis	133
9.5.1 Range	135
9.5.2 Mean	135
9.5.3 Max	136
PART IV	139
10. CONCLUSIONS, LIMITATIONS AND FUTURE DEVELOPMENTS . .	139
BIBLIOGRAPHY	145

LIST OF TABLES

Table		Page
2.1	Comparison of properties between TD-OCT and FD-OCT.	15
2.2	Property comparison between SD-OCT and SS-OCT.	16
2.3	Comparison of advantages and disadvantages of the OCTA system.	22
6.1	DATASET	75
6.2	CLEAN DATASET	77
6.3	BODY INDEX	79
9.1	Balanced Accuracy for each built classifier	131
9.2	Lesion depth in pixels	133
9.3	Balanced Accuracy for new classifiers	137
10.1	Summary conclusions	142

LIST OF FIGURES

Figure	Page
1.1	Incision of actinic keratosis and progression through the various stages. 3
1.2	Progression of actinic keratosis. From left to right: AKI, AKII, AKIII, squamous cell carcinoma(SCC). It is shown in the figure how pathology by differential pathology can also evolve directly into cancer.[4] 5
1.3	Examples of lesions for actinic keratosis and its progression(BCC and SCC). 6
1.4	Schematic rapresentation of diagnosis of AK and its evolution. 9
2.1	Schematic image of the Michelson interferometer.[6] 12
2.2	Schematization of TD-OCT applied to a human eye.[5] 14
2.3	Schematization of the set-up of SD-OCT on the left, SS-OCT on the right.[5] 15
2.4	Schematic representation of the SS-OCT system. 17
2.5	OCT system in place at the Vienna Laboratory. 19
2.6	Data format. in red is shown the A-line,; in orange the B-scan. 20
2.7	Set of B-scans constituting OCT volume. 21
2.8	Left: example of OCT-en-face. Right:example of cross-section. 21
2.9	Graphical representation of the transition from OCT volumes to OCTA volumes.[7] 23
4.1	OCT system slide present at the Vienna Laboratory. 31
5.1	Example of B-scan and Cross-Section with correctly selected ROI. The red line represents the cut that must be made to remove the slide artifact, white line present in the upper portion. 36
5.2	B-scan and Cross-section with incorrect ROI. The missing background does not allow the MATLAB code to work properly. 37
5.3	OCT en face: examples from the acquired Dataset. Left: more superficial OCT. Right: OCT at a deeper layer. 38

5.4	OCTA en face: examples from the acquired Dataset. Left: more superficial vasculature. Right: OCTA at a deeper layer.	38
5.5	Flow chart of OCT processing and OCTA extraction. Left: description of raw data processing and extraction of complex volume. Right: description of OCT processing and intensity-based-method to obtain the OCTA.	39
5.6	Example of motion artifact seen as white vertical lines in the two OCTA images at two different depths.	40
5.7	The same images as in the previous figure are shown after applying the correction code for the motion artifact.	41
5.8	The figure shows on the left an en face OCTA for a more superficial skin layer; on the right a deeper OCTA. By comparing the pre and post-correction figures, the decrease in projection artefact is highlighted.	42
5.9	Example of an OCTA image pre and post application of a median filter.	43
5.10	Example of an OCTA image pre and post the contrast enhancement.	44
5.11	Left-to-right presentation of all generally performed editing steps. It can be seen how the final image obtained highlights the vascular structures more than the background.	44
5.12	Example of representation with color map for different depths.	45
5.13	Flow chart of the OCTA data improvement process.	47
5.14	Amira interface in the volume segmentation section. The main tools used to carry out segmentation have been squared off in orange. The parts in green are the segmented parts of the volume.	50
5.15	Flow chart of the three-dimensional skeleton segmentation and extraction process	51
5.16	From left to right: MIP- segmentation (binary mask with white-vessel, black-background)- skeleton. Top: images related to the surface laser. Bottom: Images referring to the deep laser.	53
5.17	Visualization of overlap of binary mask- in white- and skeleton- in red. Left: surface layer. Right: deep layer.	54

5.18	Flow chart of the two-dimensional skeleton segmentation and extraction process for superficial and deep layer.	55
6.1	Tumor angiogenesis process.	57
6.2	Left: graphic representation and dermoscopic image of a vascular structure with arborizing vessels. Right: graphic representation and dermoscopic image of glomerular vessels.	58
6.3	Dermoscopic image of a lesion characterized by Milky-pink background.	59
6.4	Working diagram of the Box Counting Method: notice how from (a) to (e) the boxes into which the image is divided become smaller and smaller. Also, boxes containing useful pixels of the image are shown in pink: these are the boxes that are counted.	62
6.5	An example of the log-log graph described above is illustrated in the figure, showing on the x-axis the box size r , on the y-axis the number of boxes containing useful pixels $N(r)$. It can be seen that in the graph on the left, the linear behavior is satisfied by 6 out of 10 points, indicating a higher complexity of the structure studied. In the right graph, only 4 out of 10 points are linear, which means that the structure will have a lower fractal dimension.	63
6.6	Example of acquisition in which, due to the crust of the lesion not allowing visualization of the vasculature, there is a black area in the image.	65
6.7	Flow chart of 3D parameter extraction with Amira and MATLAB.	69
6.8	Left: representation of an entire skeleton from the constructed dataset. Right: subdivision of the skeleton into 16 2D ROIs for parameter calculation.	71
6.9	Flow chart of 2D parameter extraction with MATLAB. The described procedure was repeated for surface and deep volumes respectively.	72
6.10	Visualization of heat maps constructed for each parameter. In this case the maps are referred to a superficial layer.	73
6.11	Example of noisy and artefact-rich data removed from the Dataset.	75
6.12	Example MIP pre and post elimination of the noisiest deep slice in the volume. It can be seen from the image that the vessels are better defined.	76

6.13	Top: heat map corresponding to a lesion on the leg. Bottom: heat map of a head lesion. The two lesions both belong to the class ‘metastatic-cancer’. From the figure it can be seen that, especially for vascular density, NB, entropy, the values are higher in the lesion on the head, where a denser and richer vasculature is expected due to its anatomical position.	78
6.14	Example of the result of skeleton improvement. It can be seen from the image that previously the skeleton tended to overrepresent vessels. By adjusting the bwskel parameter, a more accurate result can be obtained.	81
7.1	Schematisation of T-test steps.	84
7.2	Example of the correlation matrix with colour coding. Intense blue or very light colours, towards white, indicate a high correlation between the parameters.	86
7.3	Flow chart of MANOVA test.	87
8.1	Example of the depths for lesions in the four groups in the study. Depth propagates along the horizontal direction; the maximum depth reached by the lesion is indicated in red. It can be seen that the more invasive the tumour, the more the lesion tends to involve the deep layers of the skin.	92
8.2	Right: the MIP obtained by isolating the B-scans containing the lesion. Left: the binary segmentation mask obtained by setting a threshold on the intensity of the MIP. The red line indicates the maximum depth reached by the lesion, shown along the horizontal direction.	94
9.1	Boxplot Range of % Avascular Area in the 4 groups for superficial and deep layers respectively.	97
9.2	Boxplot of Maximum Radius in the 4 groups for superficial and deep layers respectively.	98
9.3	Boxplot of Fractal Dimension in the 4 groups for superficial and deep layers respectively.	98
9.4	Boxplot Range of Vascular Density calculated on the segmentation mask in the 4 groups for superficial and deep layers respectively. . .	100
9.5	Boxplot Range of Vascular Density calculated on the skeleton in the 4 groups for superficial and deep layers respectively.	101

9.6	Boxplot Range of Number of Trees in the 4 groups for superficial and deep layers respectively.	101
9.7	Boxplot Range of Number of Branches in the 4 groups for superficial and deep layers respectively.	102
9.8	Boxplot Range of Distance Metric in the 4 groups for superficial and deep layers respectively.	102
9.9	Boxplot Range of Inflection Count Metric in the 4 groups for superficial and deep layers respectively.	103
9.10	Boxplot Range of Sum of Angles Metric in the 4 groups for superficial and deep layers respectively.	103
9.11	Boxplot Range of Entropy calculated on the MIP in the 4 groups for superficial and deep layers respectively.	104
9.12	Boxplot Range of Entropy calculated on the skeleton in the 4 groups for superficial and deep layers respectively.	104
9.13	Boxplot of the Mean of the Vascular Density calculated on the segmentation mask in the 4 groups for superficial and deep layers respectively.	106
9.14	Boxplot of the Mean of the Vascular Density calculated on the skeleton in the 4 groups for superficial and deep layers respectively.	107
9.15	Boxplot of the Mean of the Number of Trees in the 4 groups for superficial and deep layers respectively.	107
9.16	Boxplot of the Mean of the Number of Branches in the 4 groups for superficial and deep layers respectively.	108
9.17	Boxplot of the Mean of the Distance Metric in the 4 groups for superficial and deep layers respectively.	108
9.18	Boxplot of the Mean of the Inflection Count Metric in the 4 groups for superficial and deep layers respectively.	109
9.19	Boxplot of the Mean of the Entropy calculated on the MIP in the 4 groups for superficial and deep layers respectively.	109
9.20	Boxplot of the Mean of the Entropy calculated on the skeleton in the 4 groups for superficial and deep layers respectively.	110

9.21	Boxplot of the Maximun values of the Vascular Density calculated on the segmentation mask in the 4 groups for superficial and deep layers respectively.	111
9.22	Boxplot of the Maximun values of the Vascular Density calculated on the skeleton in the 4 groups for superficial and deep layers respectively.	112
9.23	Boxplot of the Maximun values of the Number of Trees in the 4 groups for superficial and deep layers respectively.	112
9.24	Boxplot of the Maximun values of the Number of Branches in the 4 groups for superficial and deep layers respectively.	113
9.25	Boxplot of the Maximun values of the Distance metric in the 4 groups for superficial and deep layers respectively.	113
9.26	Boxplot of the Maximun values of the Inflection Count Metric in the 4 groups for superficial and deep layers respectively.	114
9.27	Boxplot of the Maximun values of the Sum of Angles Metric in the 4 groups for superficial and deep layers respectively.	114
9.28	Boxplot of the Maximun values of the Entropy calculated on the MIP in the 4 groups for superficial and deep layers respectively.	115
9.29	Boxplot of the Maximun values of the Entropy calculated on the skeleton in the 4 groups for superficial and deep layers respectively.	115
9.30	Results MANOVA test performed on uncorrelated parameter range values on the surface layer. Left: canonical loads of analysis without interaction. Right: canonical loads of the analysis with interaction.	118
9.31	Results MANOVA test performed on the values for the range of uncorrelated parameters, with an increasing visual trend in the boxplots from healthy to metastatic tumour condition, on the superficial layer.	119
9.32	Results MANOVA test performed on uncorrelated parameter range values on the deep layer. Left: canonical loads of analysis without interaction. Right: canonical loads of the analysis with interaction.	120
9.33	Results MANOVA test performed on uncorrelated parameter mean values on the surface layer. Left: canonical loads of analysis without interaction. Right: canonical loads of the analysis with interaction.	121

9.34	Results MANOVA test performed on the values for the mean of uncorrelated parameters, with an increasing visual trend in the boxplots from healthy to metastatic tumour condition, on the superficial layer. Left: canonical loads of analysis without interaction. Right: canonical loads of the analysis with interaction.	122
9.35	Results MANOVA test performed on uncorrelated parameter mean values on the deep layer. Left: canonical loads of analysis without interaction. Right: canonical loads of the analysis with interaction.	123
9.36	Results MANOVA test performed on uncorrelated parameter maximum values on the surface layer.	124
9.37	Results MANOVA test performed on the values for the maximum of uncorrelated parameters, with an increasing visual trend in the boxplots from healthy to metastatic tumour condition, on the superficial layer. Left: canonical loads of analysis without interaction. Right: canonical loads of the analysis with interaction.	125
9.38	Results MANOVA test performed on uncorrelated parameter maximum values on the deep layer.	126
9.39	Confusion Matrix representing the results of the multiclass classifier, taking parameter ranges as input. Above are the classifier outputs considering only the surface parameters; below is the classifier obtained using the deep parameters.	128
9.40	Confusion Matrix representing the results of the binary classifier, taking parameter ranges as input. Right: output of the classifier considering only surface parameters; left: results of the classifier with only deep data as input.	128
9.41	Confusion Matrix representing the results of the multiclass classifier, taking mean parameters as input. Above are the classifier outputs considering only the surface parameters; below is the classifier obtained using the deep parameters.	129
9.42	Confusion Matrix representing the results of the binary classifier, taking mean parameters as input. Right: output of the classifier considering only surface parameters; left: results of the classifier with only deep data as input.	129
9.43	Confusion Matrix representing the results of the multiclass classifier, taking maximum parameters as input. Above are the classifier outputs considering only the surface parameters; below is the classifier obtained using the deep parameters.	130

9.44	Confusion Matrix representing the results of the binary classifier, taking maximum parameters as input. Right: output of the classifier considering only surface parameters; left: results of the classifier with only deep data as input.	130
9.45	Results PCA analysis conducted: left - before insertion of lesion depth as a new parameter; right - after insertion of the new parameter. The dotted lines show how there is a difference in the distribution in space between the 4 groups.	134
9.46	Confusion Matrix representing the results of the multiclass classifier, taking range parameters as input and the depth lesion. Above are the classifier outputs considering only the surface parameters; below is the classifier obtained using the deep parameters.	135
9.47	Confusion Matrix representing the results of the multiclass classifier, taking the mean of the parameters as input and the depth lesion. Above are the classifier outputs considering only the surface parameters; below is the classifier obtained using the deep parameters.	135
9.48	Confusion Matrix representing the results of the multiclass classifier, taking maximum parameters as input and the depth lesion. Above are the classifier outputs considering only the surface parameters; below is the classifier obtained using the deep parameters.	136

PART I

INTRODUCTION

Chapter 1

ACTINIC KERATOSIS

1.1 About Actinic Keratosis

Actinic keratosis is a form of pre-cancer of the skin, which usually develops due to prolonged and unprotected exposure to ultraviolet rays. It affects about 60% of fair-skinned people over 40 years old and 80% of elderly people over 60 years old[1, 2]. Actinic keratosis is caused by cumulative damage: over the years, ultraviolet radiation accumulates in the skin, which acts like a sponge, and the disease can occur suddenly. UV radiation is divided into UVA, UVB and UVC[1]:

- The **UVA** represent 94% of the rays that reach the earth's surface: these penetrate deeply into the skin, stimulating the production of reactive oxygen species that damage cells; they also promote mutations in DNA that lead to abnormal cell proliferation;
- The **UVB**, partially filtered by the ozone layer, make up about 4% of the UV radiation that reaches Earth. However, they have an effect on DNA, leading mainly to a mutation in the p53 protein that regulates the cell cycle;
- Finally, the **UVC** are completely absorbed by ozone, not reaching the Earth's surface.

Mechanisms contributing to the development of actinic keratoses include inflammation, oxidative stress, immunosuppression, and deregulation of the cell cycle and cell

proliferation. Depending on the stage of the lesion, these processes may affect different layers of the epidermis.

Given the high prevalence of actinic keratoses and the risk of progression to squamous cell carcinoma[1, 2], as highlighted in Figure 1.1, timely diagnosis and appropriate treatment is essential[3].

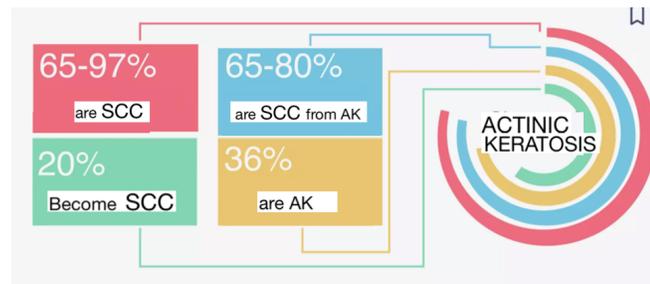


Figure 1.1: Incision of actinic keratosis and progression through the various stages.

1.2 Causes and Risk Factors

Both experimentally and epidemiologically, ultraviolet rays are identified as the main cause of the manifestation of actinic keratosis. This is evident as the affected skin exhibits not only actinic keratosis but also visible signs of solar damage[2], such as wrinkles and freckles. Exposure to UV rays can occur voluntarily, as in the case of summer tanning and the use of tanning lamps, but also unintentionally during walks and predominantly outdoor work activities. The main risk factors[3, 1, 2] include:

- **Age:** Keratosis is indeed common in the older people;
- **Clear skin;**
- **Genetics:** whether you have a family history increases the probability of developing the disease;

- **Immunosuppression:** alteration of the immune system can accelerate the evolution of keratosis.
- **Male gender.**

1.3 Features and evolutions

The lesions occur in the areas most exposed to radiation, such as the face, forehead, neck, and hands. They manifest as spots, pustules, with undefined contours that may sometimes have small plaques, and the color ranges from pink, bright red to brownish[3, 2]. Their detection can be much easier by touch than by sight, especially at the beginning of pathology, when the lesions are very small but the scales are rough and the area has raised papules[1].

Most often they are asymptomatic, although they may give itching, burning or bleeding[1]. According to the best-known three-level classification systems, actinic keratoses have three main stages, which are differentiated according to the extent of cellular atypia in the skin layers[4]:

- **AK I:** at this stage the atypia of keratinocytes (epidermal cells) is limited to the lower third of the epidermis;
- **AK II:** here, keratinocyte atypia affects the lower two thirds of the epidermis;
- **AK III:** at this advanced stage, keratinocyte atypia extends to the upper layers of the epidermis.

It is often thought that disease progression occurs by moving in an orderly fashion from stage 1 to stage 2 to stage 3; yet, recent studies show that AKs with atypical cells present in the deeper layers are the most common precursors of SCC arising

through the differentiated pathway.

Therefore, it is not possible to prevent which AKs will progress, independent of grade.

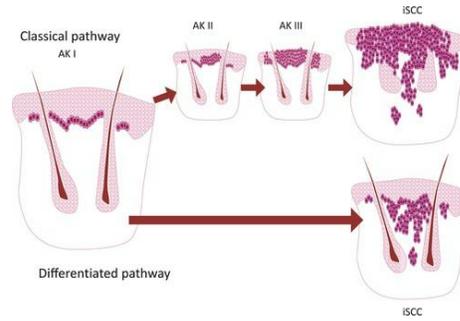


Figure 1.2: Progression of actinic keratosis. From left to right: AKI, AKII, AKIII, squamous cell carcinoma(SCC). It is shown in the figure how pathology by differential pathology can also evolve directly into cancer.[4]

Actinic keratosis represents a potentially evolutionary lesion, with the possibility of transforming into more advanced forms of carcinoma.

Therefore, the study conducted was concerned with going to discriminate between the groups presented below:

- **Actinic keratosis (AK)** -Pre-cancerous condition: involves only a lower third of the epidermis, with alteration of keratinocytes (epidermal cells) limited in this area.

Atypical keratinocytes are called in this way because these cells show enlarged and irregular nucleuses, disorder in cellular organization, increased proliferation, and abnormal mitoses.

The skin appears rough, scaly, and slightly reddened, characterized by very small lesions.

- **Bowen’s disease (superficial stage of squamous cell carcinoma):** this is an early stage of SCC in which cancer cells remain confined to the lower

two-thirds of the epidermis. Atypical keratinocytes increase in number, and the lesions become more obvious, manifesting as a thickened, wrinkled surface. Bowen's disease can persist for years and, if untreated, evolves into SCC.

- **Basal cell carcinoma (BCC):** is a malignant skin tumor, generally less aggressive than SCC, but with a significant risk of local recurrence. Unlike Bowen's disease, BCC remains limited and invasive only locally. It rarely metastasizes, but can cause significant damage to surrounding structures.
- **Squamous cell carcinoma (SCC):** is the invasive form of skin carcinoma. Atypical cells spread through the entire thickness of the epidermis, invading the basal layer and, in some cases, the underlying tissues. Lesions are thick, sometimes ulcerated, and have a risk of metastasis, especially in advanced stages.

Because Bowen's disease and basal cell carcinoma (BCC) are intermediate and less invasive conditions than squamous cell carcinoma (SCC), they were considered together in the study and treated as one category to be discriminated.

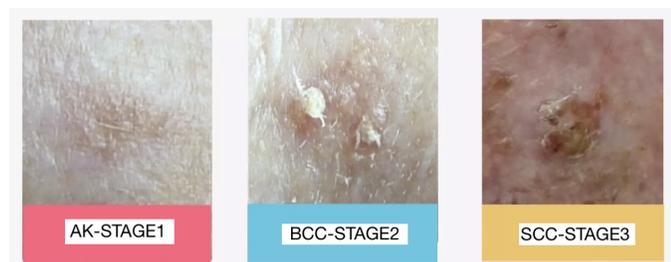


Figure 1.3: Examples of lesions for actinic keratosis and its progression(BCC and SCC).

1.4 Diagnosis

The main steps of how the diagnosis of actinic keratosis is made are given in Figure 1.4. The diagnosis is frequently **clinical**. The doctor examines the patient's medical

history and symptoms potentially associated with the condition, carefully evaluating the lesions based on features such as color, texture, shape, and size. However, visual observation alone does not provide a definitive diagnosis.

For this reason, **Dermoscopy**[1], a noninvasive diagnostic examination that allows the different layers of the skin to be explored and analyzed for abnormalities, is often used. A distinction can be made between:

- **Normal Dermoscopy:** uses normal light and allows observation of the skin with magnification, without exploring internal structures;
- **Epiluminescence Dermoscopy:** uses a polarized light or techniques to reduce glare, and allows penetration of the skin by investigating internal structures. After disinfecting the lesion and applying a thin layer of oil to increase the transparency of the surface layers, the doctor uses a dermatoscope, equipped with special lenses and illumination systems, to examine the deeper layers of the skin.

These techniques are painless, require no special preparations, are rapid, and can be performed in real time. However, it has some limitations:

- Initial lesions may be very small and not very obvious with the dermatoscope; in addition to this, they may have features very similar to other conditions, making certain diagnosis complex.
- Using the dermatoscope in some areas, such as the scalp or where the skin is very thick may not be easy, making it difficult to examine such areas.
- It does not provide definitive diagnosis.
- It is not sufficient to distinguish with certainty the developmental stages of keratosis, and therefore a cancerous condition from a noncancerous one.

For all these reasons, in order to have a definite diagnosis of keratosis and its later stages, thus identifying the severity of the pathology, it is often necessary to have recourse to **Biopsy**.

During this procedure, a sample of the lesion is taken that will be submitted to microscopic analysis in the laboratory. The biopsy is not a dangerous examination; however, it is an invasive examination, and considering that keratosis very often does not present as a single lesion, but as multiple lesions[3], the number of which may increase according to the severity of the condition, the examination may be uncomfortable. In addition, compared to dermoscopy, it has much higher costs, but more importantly, it is not a real time analysis: it takes time to study the specimen in the laboratory.

Despite this, up to now it is shown to be the most accurate method to diagnose actinic keratosis and especially to understand if it is evolving to SCC.

However, in recent years a technique has taken over that allows the advantages of the analyses presented above to be merged: **Optical Coherence Tomography (OCT) and OCT Angiography (OCTA)** use near-infrared light to create images of cross-sections of the skin in real time. In addition to visualizing the architecture of the skin, OCTA also allows visualization of blood vessels and their structure. These techniques are noninvasive, and they could allow not only to make diagnoses of actinic keratosis, but also to follow its evolution, distinguishing between noncancerous keratoses and successive cancerous conditions.

Keratoses Actinics Diagnostic Algorithm

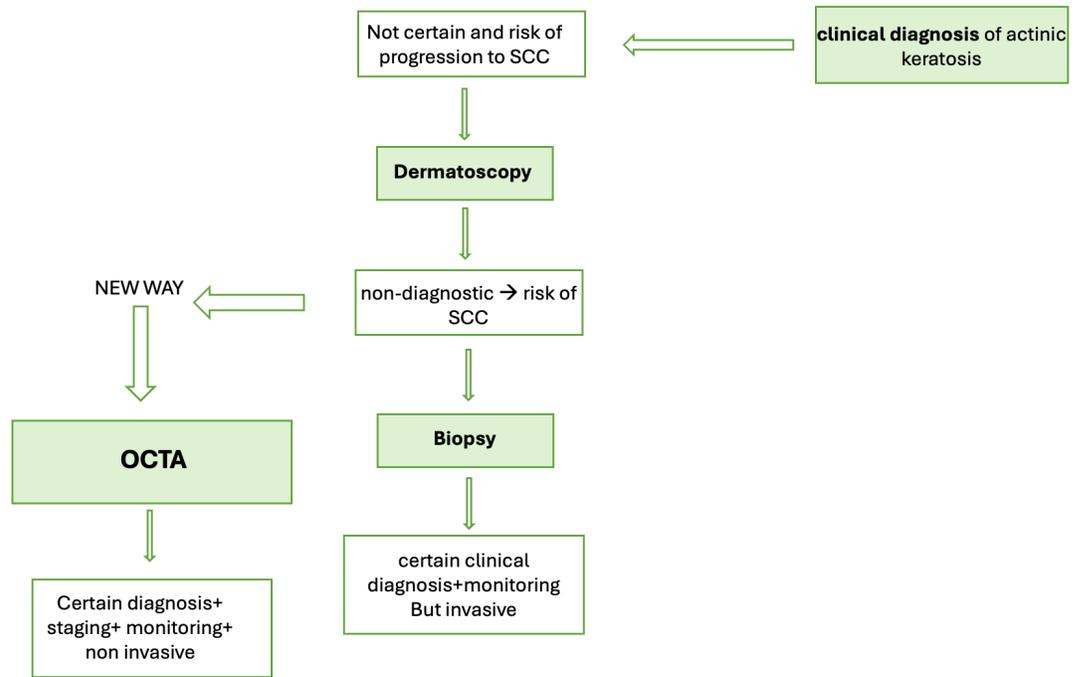


Figure 1.4: Schematic representation of diagnosis of AK and its evolution.

Chapter 2

OCT SYSTEM

OCT represents an advanced imaging technology that uses light to acquire high-resolution images at both the microscopic and submicroscopic levels, including three-dimensional images [5]. This technique makes it possible to obtain detailed cross-sections of biological tissues, penetrating materials down to a depth of a few millimeters. OCT has wide application in the medical field, particularly for ophthalmic and dermatologic imaging, due to its ability to provide structural information noninvasively.

OCT imaging is based on similar principles to those on which ultrasonography is based; however, instead of sound waves, light waves are used, which propagate at a much higher speed and allow measurement of the intensity and time delay of the reflected light signal. This procedure makes it possible to obtain information on the depth of the different tissue layers being analyzed.

2.1 Overview of OCT Fundamentals and Methods

2.1.1 OCT Principles

At the basis of OCT systems is definitely the **interference**, a physical phenomenon that will be discussed briefly.

Interference occurs when two waves overlap, resulting in a new wave. This overlapping of amplitudes can be constructive or destructive, depending on their relative phases:

- When they are in phase, the resulting wave has amplitude equal to the sum of the respective two amplitudes, and the **interference** will be **constructive**;
- When they are out of phase, the amplitude of the resulting wave will be smaller;
- When they are in counterphase, the result is a wave with zero amplitude, and the **interference** will be **destructive**.

Consider two monochromatic waves of light, with equal frequency, propagating in space:

$$E_1(x, t) = E_0 \cos(kx - \omega t) \quad (2.1)$$

$$E_2(x, t) = E_0 \cos(kx - \omega t + \varphi) \quad (2.2)$$

The wave amplitude is E_0 , k is the wave number (which depends on the wavelength of light), ω is the angular frequency, t is the time, x is the position along the direction of propagation, and φ is the phase difference between the two waves.

When these two waves overlap, the resulting intensity of light (proportional to the square of the total amplitude) will be given by:

$$I \propto (E_1 + E_2)^2 = E_0^2[2 + 2 \cos(\varphi)] \quad (2.3)$$

The total intensity depends on the phase difference φ , which determines whether the interference is constructive or destructive.

One of the most famous interferometers is the **Michelson Interferometer**[6], reported in Figure 2.1.

The system consisted of two reflecting mirrors and a separating mirror, with the objective of producing an interference wave for analysis and study. When the light beam emitted from the source impacts the splitter, two beams are created, which, traveling through L1 and L2 respectively, reach the two mirrors M1 and M2 and are reflected,

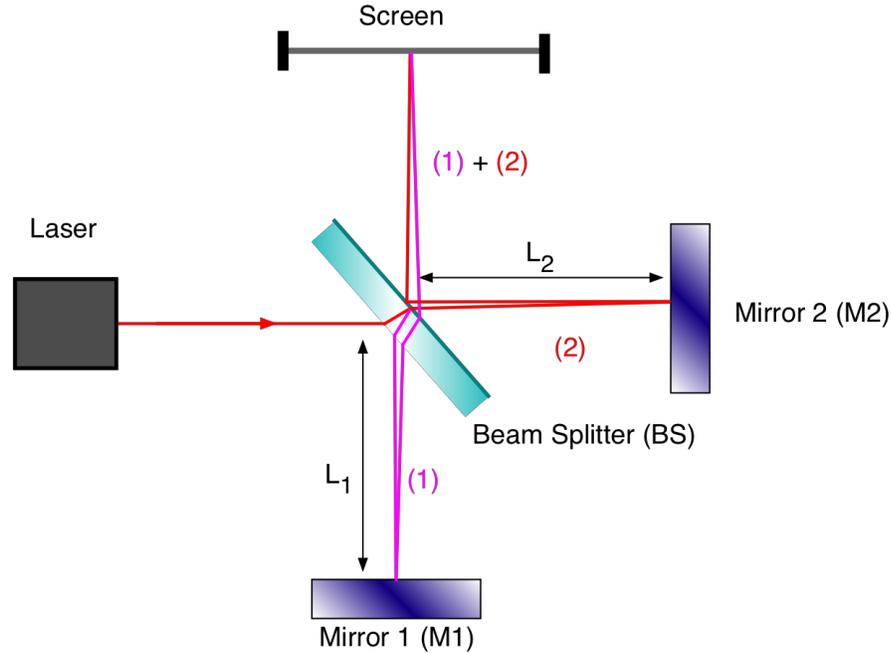


Figure 2.1: Schematic image of the Michelson interferometer.[6]

traveling back along the same path and combining in the splitter. In OCT, one mirror is our sample so, depending on the refractive indices of the various tissue layers, partial reflections are generated. For each reflection in the sample, the optical path difference produces a different phase in the interferometric signal. This phase difference contains information about the depth from which the reflection originated.[6]

The resulting signal is detected by a photodiode array, which measures the amplitude of the interference signal.

However, interference occurs only when the optical path difference between the two beams is less than the longitudinal coherence of the light.

The coherence of a light source is related to the ability of light waves to maintain a constant phase over time. In the case of temporal coherence, there is a close correlation with the spectral width of the source: the more wavelengths emitted, the lower the temporal coherence. We therefore distinguish:

- **Long-coherence light:** Monochromatic light sources (a single wavelength) are characterized by very long coherence, so waves can interfere even if the beams travel very different optical paths.
- **Low-coherence light:** Sources with a large spectrum (e.g., broadband light) have a very short time coherence, so the waves can interfere only if the optical path difference between the two beams is very small.

OCT exploits **low coherence interference**. The principle is that, using low coherence light, interference between the two beams (sample and reference) occurs only when the optical path difference between them is equal to or less than the coherence length L_c of the source. The coherence length is defined as:

$$L_c = \frac{\pi}{2 \ln(2)} \cdot \frac{\Delta\lambda}{\lambda_0^2} \quad (2.4)$$

where:

λ_0 is the central wavelength of the source, $\Delta\lambda$ is the spectral width (bandwidth) of the source. A low-coherence source (with wide λ_0) has a very short coherence length, which implies that interference will be observable only when **the optical path difference between the reference and sample arms is very small**. This allows very high axial resolution, insulating reflections from specific depths.

2.1.2 OCT Techniques

The first implementation of OCT was the **OCT in the time domain**, also called **TD-OCT**, whose working scheme is shown in Figure 2.2.

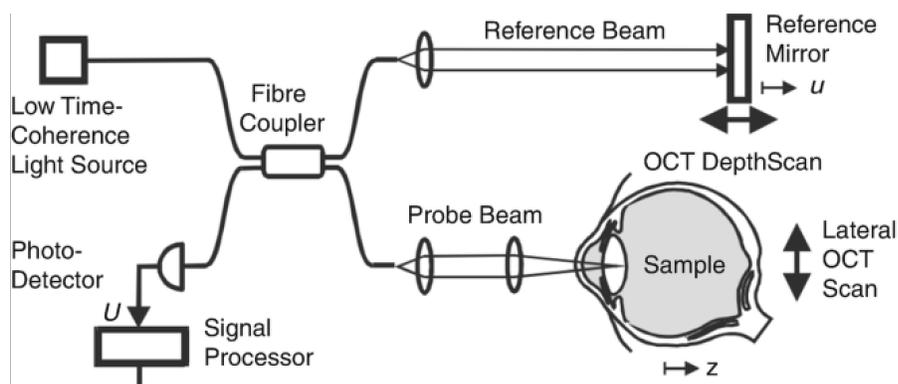


Figure 2.2: Schematization of TD-OCT applied to a human eye.[5]

The system uses a low-coherence light source, and the beam produced through a beam splitter is split into two arms: one to the sample, one to a reference mirror. The two beams, after being reflected, are combined and interfere only if the optical path lengths coincide and thus the time traveled by the light is almost equal in both arms.[5]

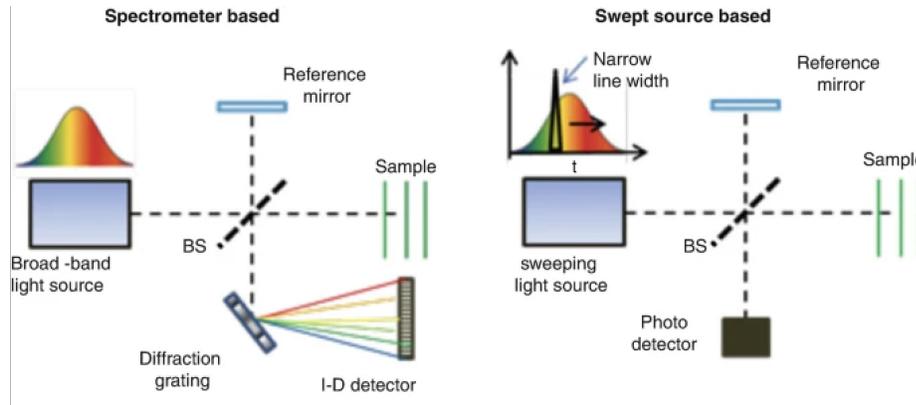
The reference arm is mechanically shifted to vary the optical path difference between the two beams, and the interferometric signal strength is measured precisely as a function of this shift.

The **OCT with Fourier domain (FD-OCT)**, is the second generation of OCT technology. A low-coherence light source is also used in FD-OCT, but a device is used to acquire the spectrum of the interferometric signal rather than mechanically moving the reference arm. The mechanics are the same as in TD-OCT, except that the Fourier Transform is to be applied to the interferometric signal to derive its spectrum. Once acquired, an inverse Fourier transform is applied to this to convert the interferometric signal from the frequency domain to the depth domain.[5]

To acquire the spectral information of the interferometric signal in the FD-OCT mode, two methods have been defined[5].

Table 2.1: Comparison of properties between TD-OCT and FD-OCT.

Characteristic	TD-OCT	FD-OCT
Acquisition method	mechanical motion of the reference arm	spectrometry- no mechanical motion
Acquisition rate	limited by the speed of movement	very high
Sensitivity	low	greater for simultaneous recording
Complexity	easier	more complex and expensive

**Figure 2.3:** Schematization of the set-up of SD-OCT on the left, SS-OCT on the right.[5]

Spectral-Domain OCT (SD-OCT)

The set up is very similar to TD-OCT, but the point detector is replaced by a spectrometer, which separates the different wavelengths of reflected light. One commonly goes to use a source that has a wide bandwidth and high power. Interference between the reference beam and the sample beam produces a spectral interferogram, recorded by a line-scanning camera-Figure 2.3 at left.

Swept-Source OCT (SS-OCT)

The operation is very similar to SD-OCT, with the difference that:

1. There is a double balanced detector (DBD) at the output of the interferometer

instead of a spectrometer.

2.A variable frequency laser is used as the light source. It emits a single wavelength at a time, covering a wide range of wavelengths in total. Figure 2.3 at right.

Table 2.2: Property comparison between SD-OCT and SS-OCT.

Characteristic	SD-OCT	SS-OCT
Configuration	point-detector spectrometer	light source scanning-sweep
Signal detection	simultaneous spectrum acquisition	sequential acquisition during sweep
Acquisition rate	limited by the recording speed	High, >100kHz
Depth reached in the tissues	limited by the used wavelengths	greater-longer wavelengths
Resolution	High, due to simultaneous recording	high, but it depends on the speed of the sweep

2.2 Laboratory OCT System

This section presents the SS-OCT system built at the Center for Medical Physics and Biomedical Engineering at the Medical University of Vienna and used for all acquisitions in the study.

2.2.1 Device components

First, the main components of the system are examined, also shown in the overall OCT device diagram in Figure 2.4:

- The **Laser Source** is the heart of the system, because it provides the light used to scan tissue.

It is a swept source (SS-OCT-1310, Insight Photonic Solutions, Inc, USA). The

central wavelength used for our purpose is 1310 nm-used to be able to penetrate tissues without damage-and the bandwidth is 29 nm. The scanning frequency is 222.22 kHz, and the power emitted by the laser is 70 mW. During the path some of the power is dispersed, so about 25 % of the total power is delivered to the patient's skin. The laser emits light that is divided into the two arms of the system (sample and reference). In the device analyzed, 75 % of the laser power goes into the sample arm and 25 % into the reference arm.

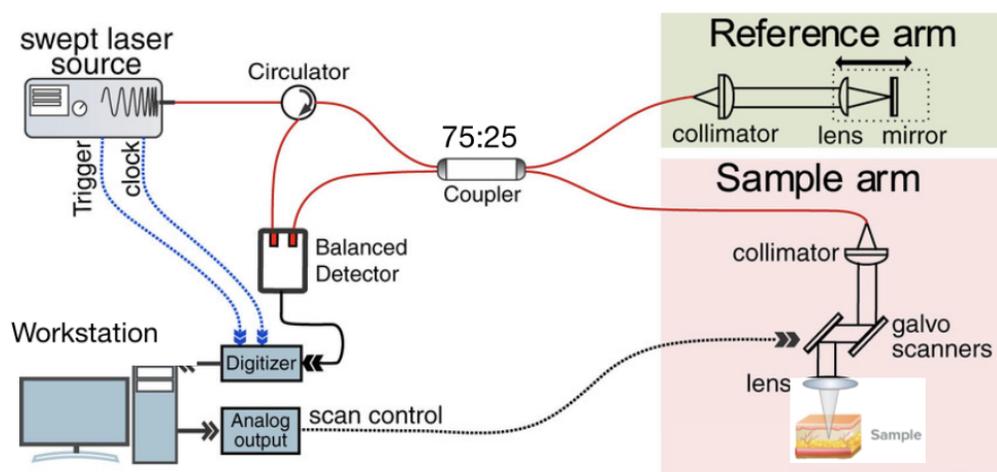


Figure 2.4: Schematic representation of the SS-OCT system.

- The **Reference Arm** consists of a collimator and 3 pedals for polarization adjustment, a power adjustment roller, and 2 mirrors. Light from the laser source is sent into this arm, travels along a fixed optical path, at the end of which is a mirror. Having encountered this, it is reflected back to the beam splitter. We thus obtain a signal that can be used as a comparison with the signal coming from the sample arm.
- The **Sample Arm** is the path through which part of the light is sent to the sample. It is characterized by a galvanometric scanner, which moves the laser

beam toward the sample, and lenses, which focus the beam into the portion of the tissue you are interested in analyzing. A series of knobs are available to manually focus and orient the beam on the sample.

The light penetrates the skin tissues and is reflected by the various layers and vessels. The reflected signal thus comes back following the same path, reaching the beam splitter.

- The **Beam splitter** is fundamental to the interference principle on which the OCT device is based. It is in fact responsible for splitting the light source into the two arms, but also for receiving the two reflected signals and recombining them: thus, the interference phenomenon is generated, which will allow the extraction of information about the characteristics and depth of the penetrated tissues.
- After recombination, the resulting signal is detected by the **Dual Balance Detector**. This is able, by means of two photodiodes, to detect the phase difference between the sample and reference signals to produce an accurate signal, also reducing noise and thus increasing its quality. The signal is then sent to the workstation.
- In the **PC Workstation**, data are collected and processed. The raw signals are transformed into useful images, including operations such as the Fourier Transform.

To summarize, the laser delivers light that is split into the two arms: a reference one, useful for comparison, and one that reaches the sample, penetrating the tissues. The two reflected beams are collected by the beam splitter, generating the interferometric signal collected by the Dual balance Detector and processed in the workstation. In

Figure 2.5 is reported the system present at the Vienna Laboratory and used for all the scans in the study.



Figure 2.5: OCT system in place at the Vienna Laboratory.

2.2.2 Device's functioning

Describing the parts involved and their roles, let's go on to understand how, by measuring the intensity of reflected light as a function of depth using the phenomenon of interference, an OCT volume can be reconstructed.

In Figure 2.6 is shown how the scanning takes place by moving on the three axes, which is characterized by:

- An **A-line**, representing the optical reflection signal along a depth-axis z-axis within the tissue. The A-Line can be seen as the basic unit of OCT volume: in fact, combining multiple A-Lines results in a B-scan.

- A **B-scan** is a two-dimensional image formed by combining a set of A-lines acquired as the laser beam is moved along the x-axis of the sample. Each A-line represents a column of the image, while lateral movement allows a cross section (2D) to be constructed in the xz plane of the tissue under examination.

In the systems under analysis, each A-scan collects 2048 points even though only about 1500 of them are valid, each B-scan features 512 A-scans, and each volume is constructed using 512 B-scans.

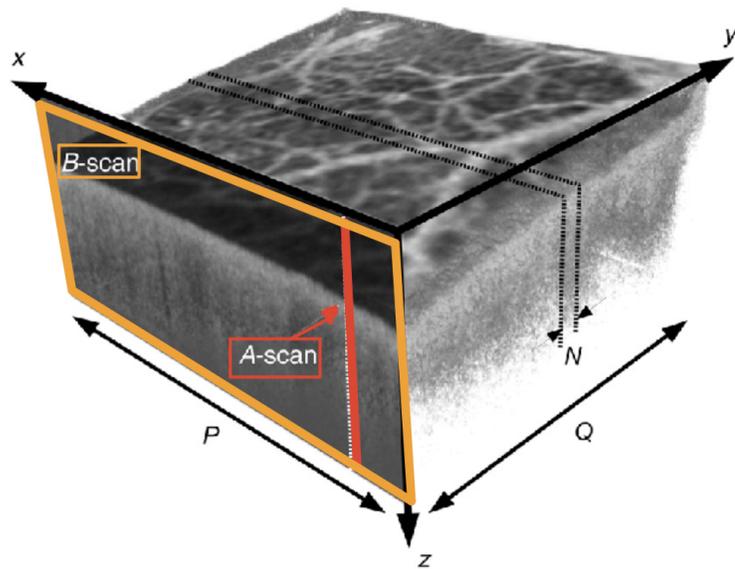


Figure 2.6: Data format. in red is shown the A-line,; in orange the B-scan.

By scanning the sample along the y-direction, the OCT system can collect a series of B-scans to reconstruct a three-dimensional volume of the tissue-Figure 2.7 . Each B-scan represents a two-dimensional “slice” of the tissue, and putting these slices together provides a detailed volumetric representation.

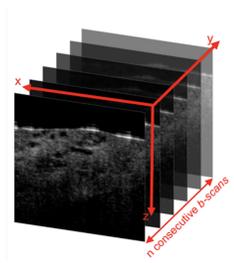


Figure 2.7: Set of B-scans constituting OCT volume.

After collecting all the B-scans and obtaining the OCT volume, it can be informative to look at the **cross-section**, two-dimensional images in the yz plane, and the **OCT-en-face**, in the xy plane, which gives a view of the volume from above. These can be informative to visualize the layering of the skin, but also the morphology of the more superficial and deeper layers of the samples analyzed. We report one example for each in Figure 2.8.

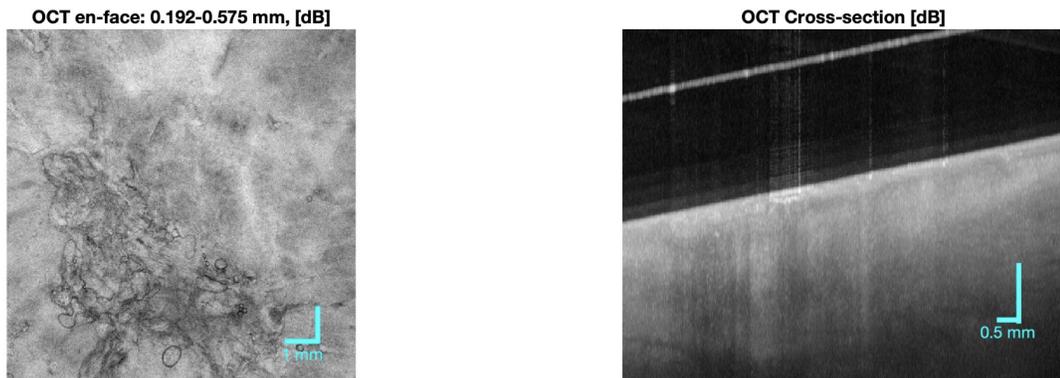


Figure 2.8: Left: example of OCT-en-face. Right: example of cross-section.

From OCT, our goal is to move to OCTA, a useful technique for conducting our study.

2.3 OCTA techniques

OCT provides structural images of tissue, but given our purpose of diagnosing and differentiating between a cancerous condition and not, this information is not enough for us. Tumors have the well-known characteristic of greatly changing the vasculature of tissue: they need to grow, and to do this they need blood that they receive by creating an increasingly dense but also disorganized network of vessels.

That is why, for the purposes of this study, talking only about tissue is not enough. It is necessary to talk about vasculature to observe the architecture of the vessels, and this, with an imaging technique like OCT that only collects static information, is not exhaustive.

Therefore, to extract the dynamic information related to blood flow in the vessels, and thus to study the vasculature, we introduce OCTA.

Over the past decade, OCTA has developed considerably. Initially proposed in 2006, it was approved for clinical use in 2016[7]. OCTA has rapidly gained popularity in the medical field because, while similar to fluorescein angiography (FA), it offers detailed three-dimensional visualization of the microvasculature, down to the capillary level, noninvasively and without the need for contrast agents.

Table 2.3: Comparison of advantages and disadvantages of the OCTA system.

OCTA ADVANTAGES	OCTA DISADVANTAGES
✓ Vascular images without contrast agent	✗ Sensitivity to movement artifacts
✓ Non-invasive	✗ Lack of information about flow velocity
✓ High resolution	✗ Susceptibility to projection and shadowing artifacts
✓ Segmentation of different vascular layers	✗ Limited field of view

The basic principle of OCTA is to scan the same area at different times. Specifically, in the device used, four volumes of OCT are acquired. The only variable among these images is the movement of blood within the vessels, as the tissues remain constant. From the interferogram generated by the OCT, which provides a spectral profile, it is possible, by applying the Fourier Transform to obtain a signal in the spatial domain, to extract information regarding intensity and phase.

$$\mathcal{FFT}[S(x, k)] = T(x, z) \exp[-i\phi(x, z)], \quad (2.1) \quad (2.5)$$

where $S(x, k)$ is the spectral profile of an A-scan in the space k , $T(x, z)$ and $\phi(x, z)$ are the magnitude and phase of its FFT, respectively. It is known that signals from stationary tissues and those from red blood cells, or otherwise moving matter, have different temporal characteristics. These temporal differences make it easy to identify areas where blood flow is active, thus distinguishing between moving and static tissues. A pipeline of the process is shown in Figure 9.29.

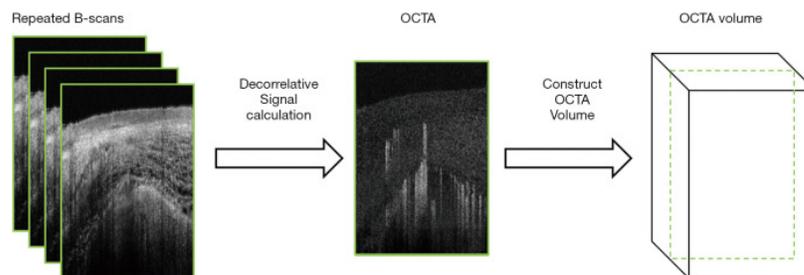


Figure 2.9: Graphical representation of the transition from OCT volumes to OCTA volumes.[7]

Depending on the signal we use (amplitude or phase), decorrelation can be done in two different ways, which we will now describe. In addition, to improve the noise on angiography, an averaging of the decorrelated signals is done.

Intensity-based OCTA

The technique is based on using the Fourier transform intensity of the interference

signal: when there is motion, as in blood flow, the signal intensity changes[7]. The contrast volume $A(x,y,z)$ is obtained by averaging the differences between pairs of consecutive tomograms, from the same set of a y -position, after applying the logarithm to each tomogram $T(x,z)$, to improve image quality and reduce unhelpful variations[8]:

$$A(x, y, z) = \frac{1}{N-1} \sum_{i=0}^{N-1} |\log(T(x, z)_{i+1}) - \log(T(x, z)_i)|. \quad (2.6)$$

where N is the number of B-scans acquired at different times.

The technique is simple to implement and is useful for detecting vessels where there is a high contrast between flow and tissue. However, the method is very sensitive to motion artifacts.

Phase-based OCTA

In this case, the technique goes to exploit the phase of the Fourier transform of the interference signal, which changes when the red blood cells move[7]. The volume contrast $A(x,y,z)$ is in this case obtained by averaging the phase changes[8]:

$$A(x, y, z) = \frac{1}{N-1} \sum_{i=0}^{N-1} |\Delta\phi(x, z)_i|, \quad (2.7)$$

with $\Delta\phi(x, z)_i$ phase difference between two A-scans performed at the same y -position but at different times, calculated as:

$$\Delta\phi(x, z)_i = \arg \{ \exp [-i\phi(x, z)_{i+1}] \exp [i\phi(x, z)_i] \}. \quad (2.8)$$

The method, compared with Intensity-based, is more sensitive to small movements, making it useful for obtaining more detailed images characterized by many microvessels. On the other hand, however, it is more difficult to implement and it is prone to bulk motion and projection artifacts[7].

The method adopted for our purposes was the **Intensity-based method**, views:

- its simple implementation;
- Its effectiveness in detecting the density and structure of the vasculature, as tumors are often associated with an abnormal network;
- the ability to visualize the superficial and deep vasculature, which may be altered depending on the grade of the lesion;
- little more robustness to artifacts than the other technique.

Chapter 3

PURPOSE AND STRUCTURE OF THE THESIS

In the previous chapters, the two protagonists of this thesis work have been described: **actinic keratosis and its pathological evolutions** on the one hand, and imaging techniques such as **OCT and OCTA**, based on optical interference, on the other.

Actinic keratosis is an increasingly common condition; it represents a precancerous lesion of the skin, which can progress over time to more severe forms, such as carcinomas in situ or metastatic carcinomas. This pathological progression makes early and safe diagnosis important, to allow rapid intervention and reduce the risk of tumour progression.

Currently, for the diagnosis of actinic keratosis, which mainly affects the more superficial layers of the skin, non-invasive techniques such as dermoscopy are available, which allow almost certain detection of pre-tumour. However, when the pathology evolves into more advanced and deeper conditions, the only safe diagnostic method is biopsy, an invasive procedure that involves pain for the patient and is not always easily repeated.

In this context, the **thesis work aims to propose a non-invasive diagnostic method capable of discriminating not only actinic keratosis, but also its subsequent pathological progressions**. This method should combine safety and the ability to detect the distinctive features of the different pathological stages, thus overcoming the limitations of currently available diagnostic techniques.

OCT and OCTA, described in *Chapter 2*, have demonstrated to be promising tools for this purpose. These non-invasive optical techniques allow detailed information on tissue structure and vascularisation to be obtained, representing a valid alternative

for the diagnosis of these dermatological pathologies that combines the advantages of the methods presented previously and used these days.

After this introduction, the **Materials and Methods** section will describe the approach followed to achieve the goals. After the extraction and initial processing of the data, the thesis work was divided into two main directions of analysis:

- *OCTA analysis*: It is known from the literature [9, 10, 11, 12, 13] that the vascular structure in a pathological condition differs in certain characteristics from a healthy architecture. For this reason, an OCTA-based analysis was considered: an accurate description of the vasculature in each acquisition was made by extracting parameters from the angiography. The idea was to, by merging the data from the constructed dataset belonging to each study group, find one or more parameters that would allow discrimination between the various groups. Once the parametric data had been obtained, various statistical analyses were conducted to check the validity of the results obtained:
 - T-test to assess the significance of each parameter;
 - Multi-parametric analyses to examine the interaction between parameters;
 - Construction of binary and multiclass classifiers to test the overall discriminative capacity of the extracted parameters.
- *OCT analysis*: Considering that the progression of actinic keratosis towards more severe forms increasingly involves the deep layers of the skin, it was thought to exploit OCT morphology to detect the depth of each lesion. This information could be useful to safely discriminate between the three pathological stages, which are characterised by a different type of invasiveness.

The results of each analysis will be presented in the **Results** section, and finally summed up in the **Conclusions** section, highlighting the objectives achieved with the work, the limitations met and possible future developments of the project.

PART II

MATERIALS AND METHODS

Chapter 4

ACQUISITION PROTOCOL

The study involved a total of **14 patients**, who were clinically diagnosed with actinic keratosis, or its evolution into a more cancerous condition, through a dermatological examination. Subsequently, patients with suspected pathology were submitted to imaging by OCT at the department of *Vienna General Hospital*, resulting in a total dataset of **51 OCT volumes**. The diagnosis was later confirmed by biopsy, with removal of the lesion in the most severe cases.

Following diagnosis, patients were divided into **four main groups** according to the severity of their condition:

- a control group consisting of **healthy** subjects ;
- **pre-cancerous group**: patients with actinic keratosis;
- **in situ cancer group**: patients suffering from BCC and Bowen's disease;
- **metastatic cancer group**: patients suffering from SCC.

These groups represent the classes that the study aims to discriminate through the analysis of OCT and OCTA data. The main steps of each acquisition are described below:

- **Initial calibration**: At system startup, a laser calibration is performed by dedicated software, which identifies valid laser sweep points. These data are saved in a DVV file, which is necessary for subsequent processing and extraction of the angiography from the OCT volume. If multiple acquisitions are made on the same day, calibration is only needed once, when the system is turned on.

- **Environment preparation:** As the acquisition is performed in a clinical setting, it is essential to disinfect all surfaces in contact with the patient and operator preventively, including:
 - lounge,
 - device slide,
 - safety glasses.
- **Security measures:** Although the device is noninvasive, the use of a laser source requires protective glasses for patients, especially in facial area acquisitions. The laser power used is well below the maximum allowable limit (1 W/cm² according to the IEC 60825-1 standard), thus being a precautionary measure.
- **Patient preparation:** Before the acquisition, the patient is clearly explained the procedure to avoid anxiety or worries that could compromise the quality of the acquisition. The patient is placed on the lounge, with the area of the lesion exposed. A slide(Figure 4.1), 1 mm thick, is placed between the probe and the skin to flatten the region of interest and minimize unwanted reflections.

A thin layer of distilled water is applied between the slide and the probe. This

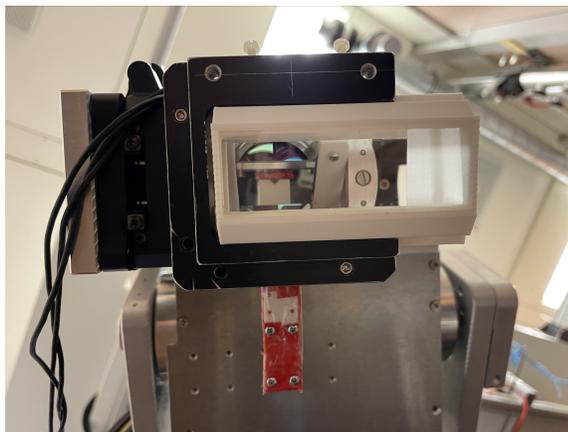


Figure 4.1: OCT system slide present at the Vienna Laboratory.

liquid is chosen for:

- its refractive index similar to that of biological tissues, which minimizes laser reflections and promotes laser penetration,
- Its transparency to the wavelength of the laser used,
- Its biocompatibility and simplicity of use.

The alignment of the laser beam is checked with an infrared card, making sure it is centered on the region of interest.

- **B-scan preview:** Before final acquisition, a preview B-scan is displayed on the workstation to verify image quality and proper contact between slide, water and skin. A balance must be found in the pressure applied so that the slide adheres completely to the skin without creating bubbles in the water. The positioning of the device can be adjusted using knobs to improve adherence and angle, and this process can be more or less complex depending on the area to be investigated (e.g., forehead or limbs).
- **Acquisition optimization:** Image quality can be improved by varying the tilt of the laser beam relative to the perpendicular, thus optimizing the visualization of subcutaneous details. Adjustments to the depth of focus and the distance of the probe from the skin allow for better images of blood vessels. Each modification is verified by observing the B-scan preview.
- **Acquisition and saving data:** Each acquisition takes about 20 seconds, and considering the time to save the data, the overall process can take up to 1-2 minutes. During the study, the following were generally performed:
 - Acquisitions of areas without lesions to have healthy tissue for reference,
 - multiple lesion acquisitions, varying the position of the probe according to the complexity of the anatomical area (e.g., forehead, sternum, shoulder).

- **Photo-documentation:** Whenever possible, a photograph of the lesion is taken to facilitate correlation between the acquired image and the anatomic area.
- **Final disinfection:** At the end of each acquisition, all equipment used is disinfected to prepare it for the next patient.

Chapter 5

DATA PROCESSING

As reported in the previous chapter, our dataset consists of 51 acquisitions that were individually processed. We also recall that, to reconstruct an OCTA volume, a multiple acquisition of 4 OCT volumes at the same location is exploited. An OCT volume consists of 512 B-scans, where each B-scan is characterized by 512 A-lines, each consisting of 2176 pixels. As each acquisition, however, consists of 4 OCT volumes, we total 2048 B-scans.

To handle this large data, the acquisition is divided into 32 packets, each containing 64 B-scans, which are individually processed to create an angiographic volume, allowing detailed visualization of the vascular structure.

5.1 Raw data processing

The first step in Processing is the loading of configuration parameters, including data on dimensions (x, y, z) and settings for the **Region OF Interest (ROI)**, which represents a selection of data with the most relevant information. Because processing the entire 3D data set requires a lot of memory, selecting a ROI reduces the computational cost and allows to focus on the areas of greatest interest.

At this stage, the loading of the **DVV file**, generated during the initial calibration of the system, is also performed, allowing processing to be limited to valid data, improving efficiency and signal quality.

With the data loaded, the function *ppOCT3D-angio* starts processing each packet:

- **Data reshape:** a reshape of the data is performed to organize the A-line and B-scan pixels into a three-dimensional matrix that simulates the acquired volume;
- **Selection of valid pixels:** the DVV file is used to select the pixels that are considered reliable by the calibration.

Then, for each B-scan, signal enhancement techniques are applied:

- **Background Subtraction:** the mean value of the horizontal line (background) is subtracted. This operation reduces the noise present along the A-line axis, improving the contrast of vascular structures;
- **Zero-Padding:** technique involving the addition of zeros at the beginning and end of the data to increase the resolution of the image, resulting in a more accurate representation;
- **Windowing:** a window is applied to the data to reduce the effect of edges that could result in distortions. This improves the clarity of the images.

Each A-line is then transformed using the FFT, allowing depth information to be obtained and a spectral representation of the data to be made, which is useful for highlighting vascular structures.

After this, all A-lines of each B-scan are concatenated together, allowing the B-scans to be merged into a single volumetric matrix called **COMPLEX**, where each B-scan is stored at a specific temporal position.

5.2 OCTA extraction

The COMPLEX volumetric matrix is the input of a new MATLAB code that allows the OCTA to be reconstructed from the OCT using the **Intensity-based-method**. In this approach, only the Fourier transform module of complex data is used, reducing computational complexity.

After the initial reconstruction, the OCT data are divided into four separate volumes. The last of these volumes is observed to be shifted by one B-scan, then a correction is applied to realign it, ensuring that all volumes are consistent with each other and ready for further analysis.

To enhance visual understanding and correct interpretation of structural information, a B-scan and a cross-section are displayed on a logarithmic scale of the first volume. There are two main observations to be extracted from these plots:

- **Glass presence:** During acquisition, the use of a glass slide is necessary to flatten the area of interest and avoid direct reflection of the light beam. However, the presence of the glass slide results as a white artefact line at the top of the image (Figure 5.2), which must therefore be removed.

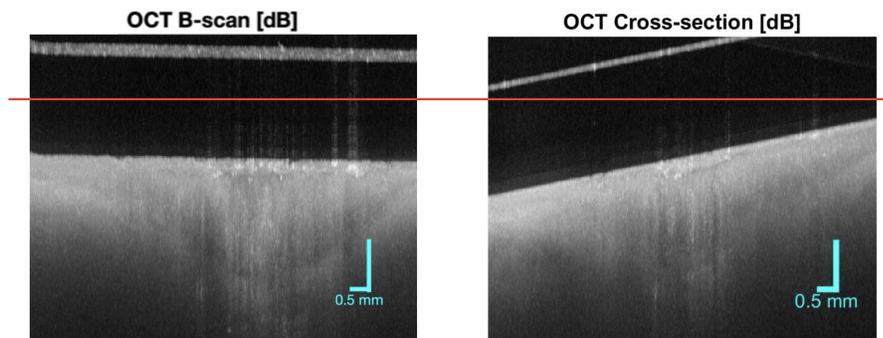


Figure 5.1: Example of B-scan and Cross-Section with correctly selected ROI. The red line represents the cut that must be made to remove the slide artifact, white line present in the upper portion.

Therefore, the depth in pixels of the cut to be made to remove the surface portion of the volumes with the artefact is entered manually.

- **Selection of ROI:** it is important to ensure that the selected ROI contains a portion of the background, as in (Figure 5.2), to allow the code to run correctly and reconstruct accurately.

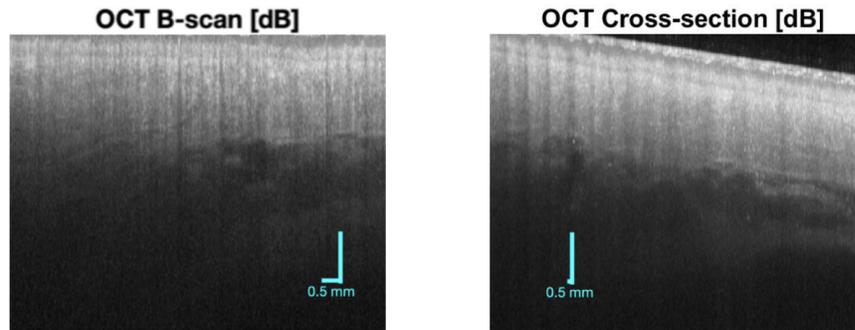


Figure 5.2: *B-scan and Cross-section with incorrect ROI. The missing background does not allow the MATLAB code to work properly.*

The B-scan also shows an inclination of the skin surface, which additionally needs to be corrected. Aligning all the B-scans at the same horizontal level ensures that the en face sections are parallel to the skin surface: this is important to obtain a correct visualisation of the cross-sections. Using each A-line of each B-scan, the beginning of the skin is identified based on an intensity threshold. These points, representing the position of the surface, are input to a linear fit to obtain a homogenous line. This line is used as a reference to realign all B-scans horizontally.

The OCTA is finally calculated by *the difference between B-scans of the same acquisition but recorded at different time points*. These differences are then averaged to obtain the vascular representation.

Once the angiographic reconstruction is complete, it is visualised:

- **OCT en face:** shows one of the four reconstructed OCT volumes.

- **Intensity-Angio en face:** showing the vessels highlighted in the differences between the volumes acquired at different time instants;

By varying the displayed slice interval, different vascular depths can be explored, as shown in Figure 5.3 and Figure 5.5.

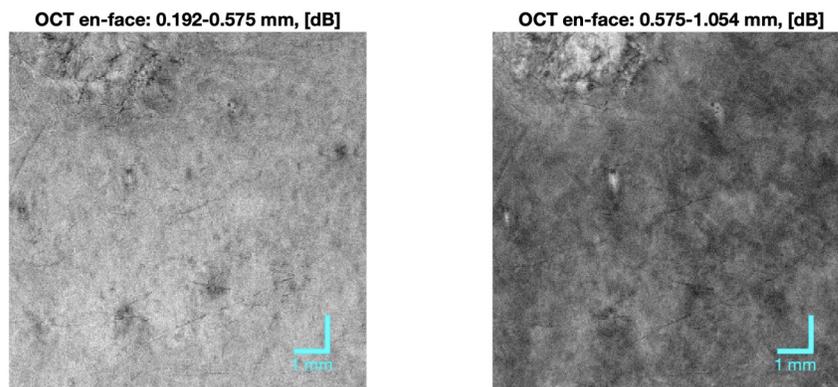


Figure 5.3: OCT en face: examples from the acquired Dataset. Left: more superficial OCT. Right: OCT at a deeper layer.

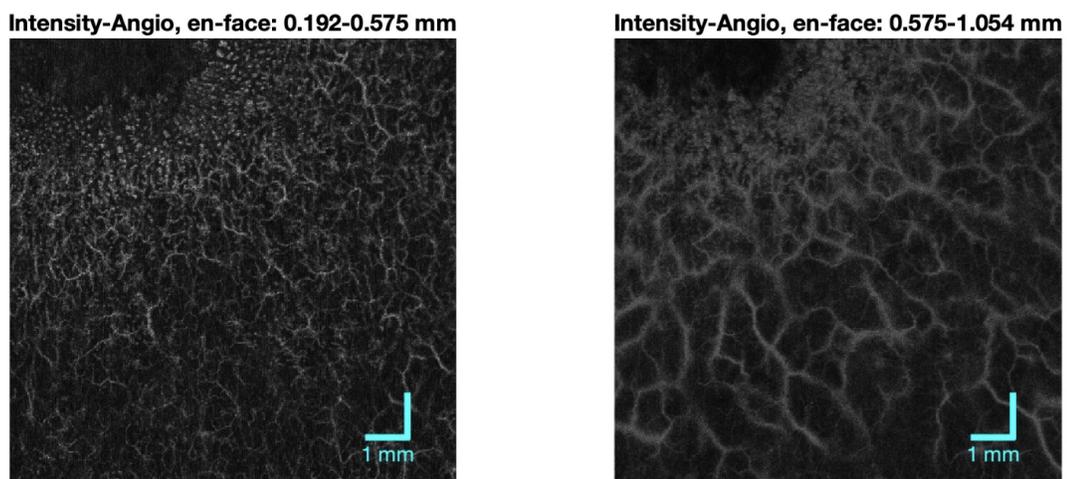


Figure 5.4: OCTA en face: examples from the acquired Dataset. Left: more superficial vasculature. Right: OCTA at a deeper layer.

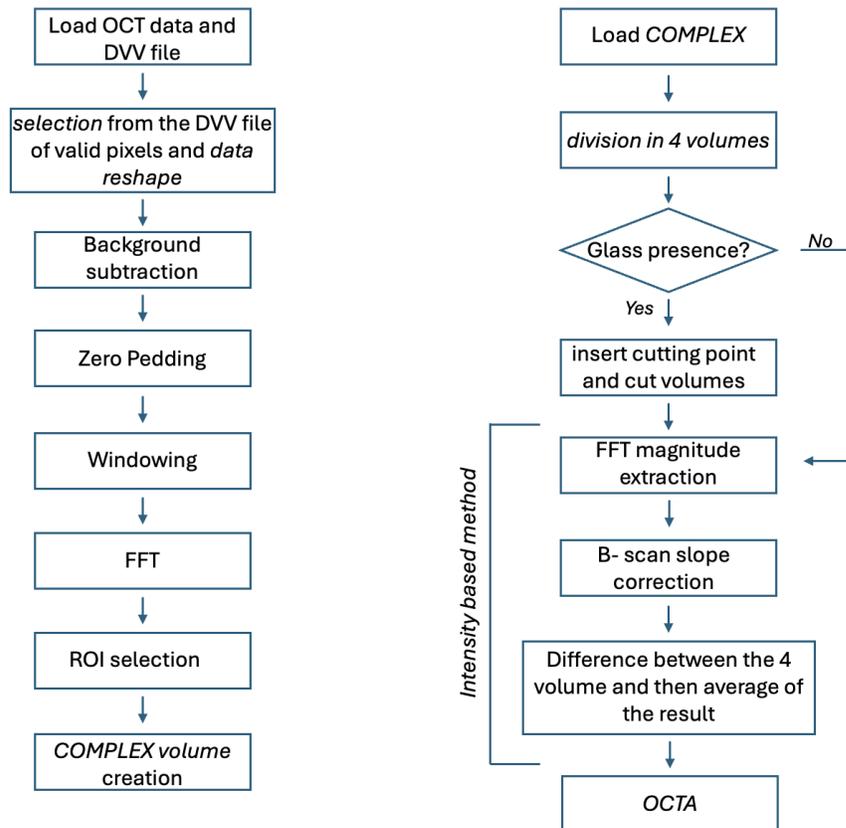


Figure 5.5: Flow chart of OCT processing and OCTA extraction. Left: description of raw data processing and extraction of complex volume. Right: description of OCT processing and intensity-based-method to obtain the OCTA.

5.3 Artifact removal and image enhancement

The OCTA can be subject to two main types of artefacts [14], which were carefully treated to avoid misinterpretation of the vasculature in the subsequent analysis.

Image editing operations were also performed to improve the visualisation of structures of interest.

5.3.1 White Line Correction

White line artefacts can have two main causes:

- **Movement:** OCTA requires a scan time to acquire images, so it may happen that the acquisition is affected by voluntary movements of the patient, or involuntary movements related to breathing or heartbeat.
- **Laser:** the laser power during acquisition is not always the same but varies, and this variation could generate an error.

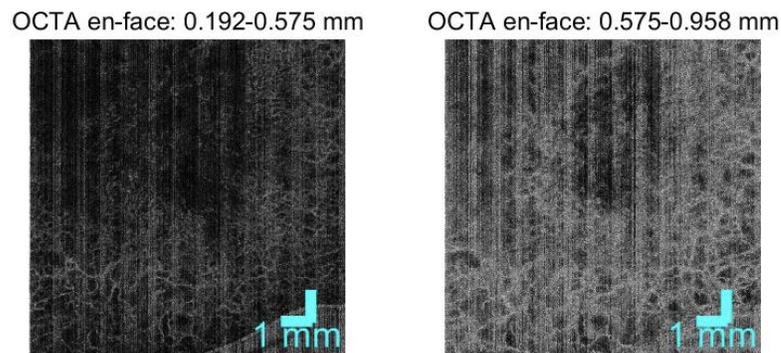


Figure 5.6: Example of motion artifact seen as white vertical lines in the two OCTA images at two different depths.

This artefact is visible as a bright white line, as shown in Figure 5.9, which can mask significant anatomical details in the image, affecting quality and making interpretation

more complex. A correction was therefore implemented in MATLAB. The code runs through each B-scan of the volume and calculates the average of the intensity values of all pixels in the B-scan considered. The current B-scan is divided by this calculated average, thereby normalising the intensity of each pixel to the average intensity of the B-scan.

Since artefact induces sudden variations in intensity (white lines), dividing by the average helps attenuate the artefact by making the image more homogeneous.

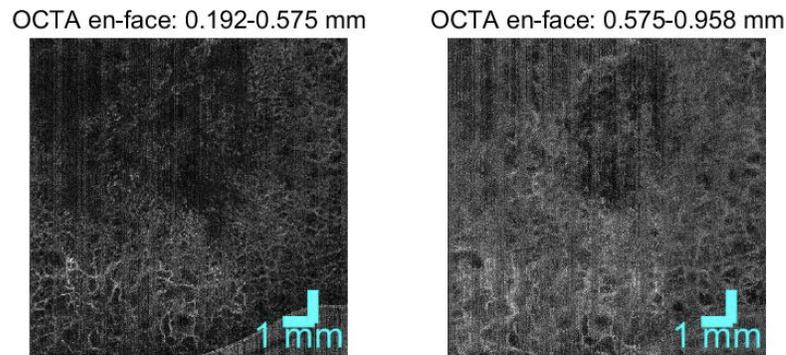


Figure 5.7: The same images as in the previous figure are shown after applying the correction code for the motion artefact.

5.3.2 Projection artefact

Projection artefact occurs when vessels in the surface layer project shadows or reflections in the deeper layers[14]: this leads to the appearance of false vessels in the lower layers of the image. An **Exponential filter** was used to correct the artefact.

Using a MATLAB code, for each pixel of the A-line of each B-scan, the cumulative sum of the intensities of the pixels located above the point considered was calculated. This sum represents the amount of information projected downwards from the most superficial layers. The intensity of the current point is attenuated using the following formula:

$$\text{Intensity}_{\text{attenuated}} = \text{Intensity}_{\text{original}} \times e^{-\frac{\text{cumulative sum}}{\gamma}}$$

where attenuation depends on the *cumulative sum* and a parameter *gamma*, which governs the intensity of artefact removal.

This process reduces the impact of superficial vascular structures as one descends to deeper layers, because the cumulative sum increases and exponential attenuation decreases in intensity.

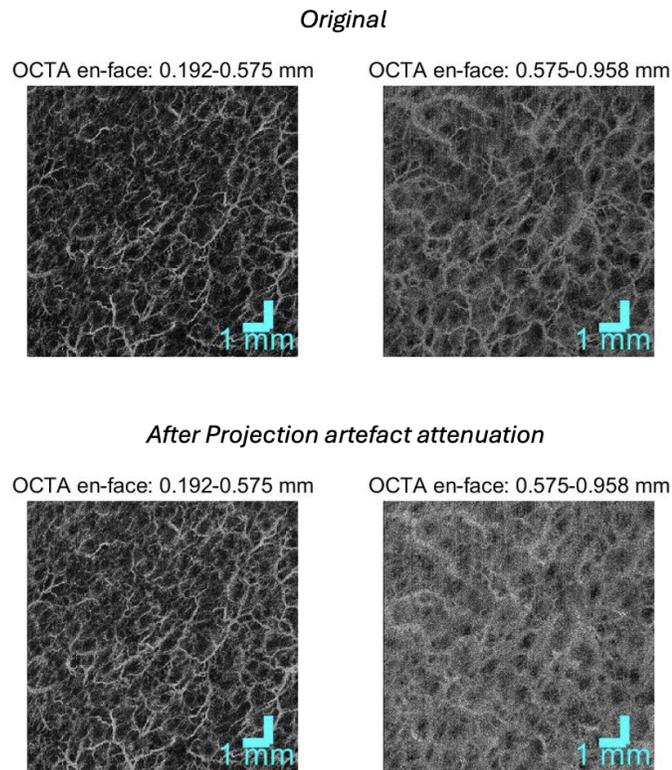


Figure 5.8: The figure shows on the left an en face OCTA for a more superficial skin layer; on the right a deeper OCTA. By comparing the pre and post-correction figures, the decrease in projection artefact is highlighted.

5.3.3 Smoothing Filter and Contrast Enhancement

In order to further improve the quality of the volumes to be analysed, two further modifications were applied in cascade, with the aim of further accentuating the structures of interest.

Smoothig Filter

A median filter was applied in three dimensions. Specifically, a 3x3x3 cube runs over each voxel of the OCTA volume, replacing the value of the considered voxel with the median value of the surrounding voxels. Median filtering reveals itself to be useful for reducing noise, while maintaining the contours and details of the OCTA's vascular structures.

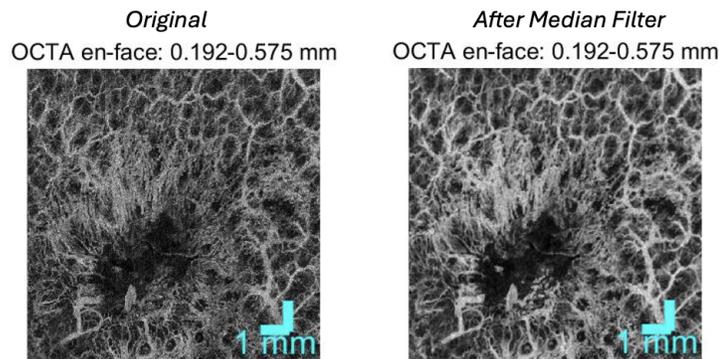


Figure 5.9: Example of an OCTA image pre and post application of a median filter.

Contrast Enhancement

Contrast modification can be very useful in highlighting the structures of interest (for us, the vessels) more than the background (surrounding tissues). Higher contrast images will also ensure greater accuracy in subsequent parameter extraction, as we will see in the following chapters.

The change was made by taking the average brightness of the image as a threshold:

- for pixels with an intensity greater than the threshold, the contrast is increased by adding a percentage of the pixel's original intensity;
- for pixels with intensities below the threshold, the contrast is always increased by decreasing the original intensity of these pixels.

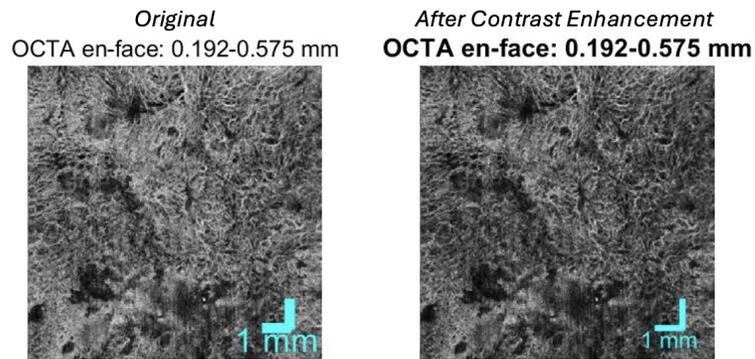


Figure 5.10: Example of an OCTA image pre and post the contrast enhancement.

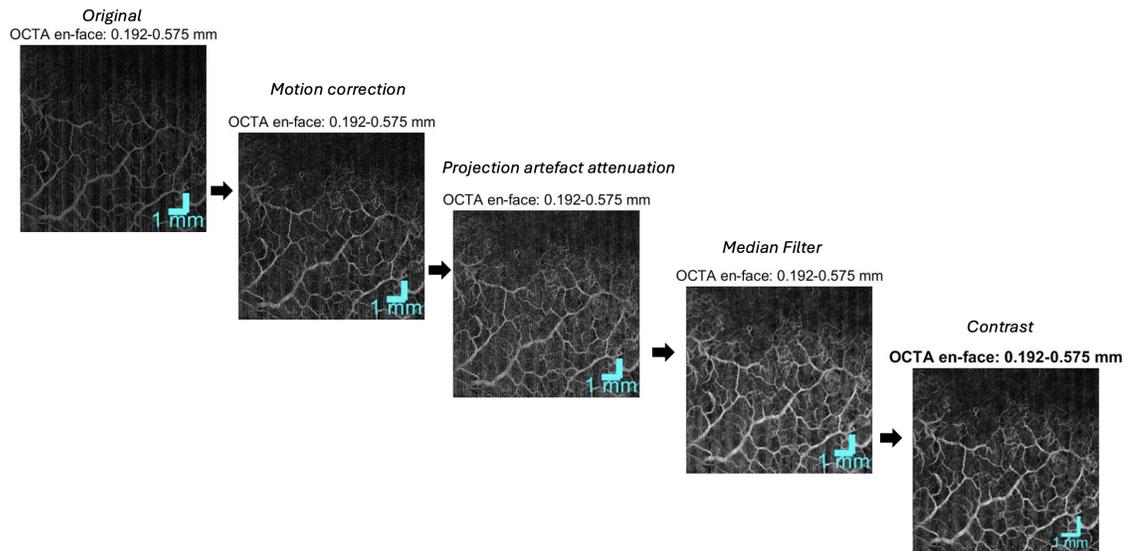


Figure 5.11: Left-to-right presentation of all generally performed editing steps. It can be seen how the final image obtained highlights the vascular structures more than the background.

5.3.4 Depth Color Coding

Following the adjustments presented above, it was decided to implement a colour visualisation in MATLAB for the OCTA volume: the aim is to map the vessels corresponding to different depths with different colours.

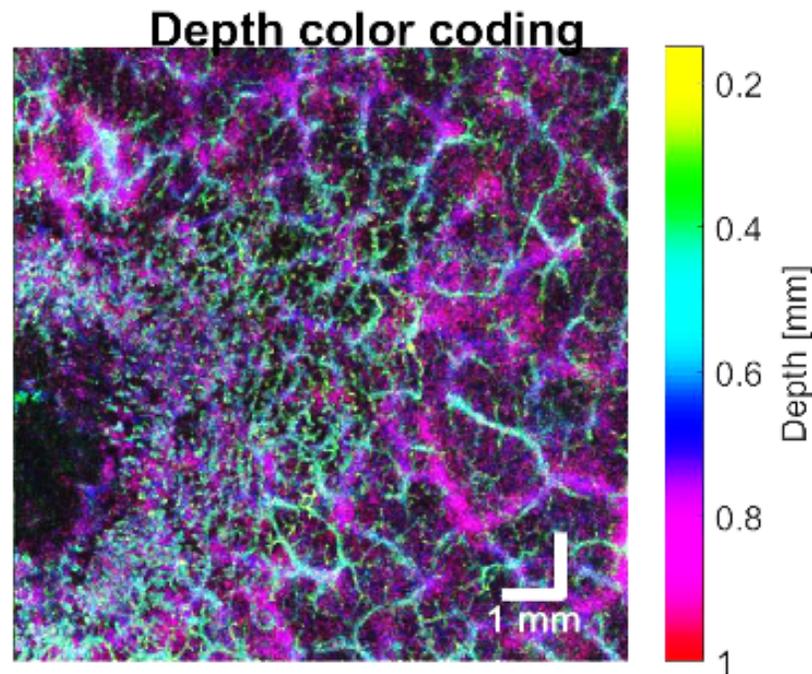


Figure 5.12: Example of representation with color map for different depths.

This visualisation allows the changes in the vasculature on the different layers to be evidenced, taking advantage of the HSV colour map. The code begins by loading the volume *CONTRAST*, the result of a cascaded application of all the changes; the analysis is limited to the first 100 slices to reduce the impact of noise associated with the deepest projections of the volume. Each slice of the volume is normalised on a scale of 0-255 and, using an RGB map, a distinctive colour is assigned to each slice. The resulting images are then combined into a three-dimensional matrices for the final visualisation, reported in Figure 6.5.

The colour-coding presented allowed us to show how the vascular architecture changes with skin depth, an aspect that will be considered and exploited to characterise and discriminate tumour from non-tumour conditions in the following chapters. A first qualitative analysis shows how the superficial vasculature is characterised by smaller vessels, with a more pronounced branching; on the contrary, the deep vasculature has a much more chaotic network, due to the need of the tumour to support growth in deeper areas, with vessels of much greater calibre.

Given this obvious difference between the two layers, it was decided not to analyse the volume as a whole, thus avoiding mixing two pieces of information relating to different vascular structures. Therefore, **the two-dimensional analysis**, presented later, **was conducted by separating the volume**, obtained with a **Field Of View (FOV) of 1 cm²** , into two sub-volumes:

- one containing the **surface layers**, characterised by a depth of *0.192-0.575 mm*, where the vascular network is more branched and ordered;
- one containing the **deep layers**, to a depth of *0.575-0.958 mm*, where vascularisation is more chaotic and dominated by larger vessels.

This subdivision allows a more detailed and targeted analysis, respecting the characteristics of the different skin depths.

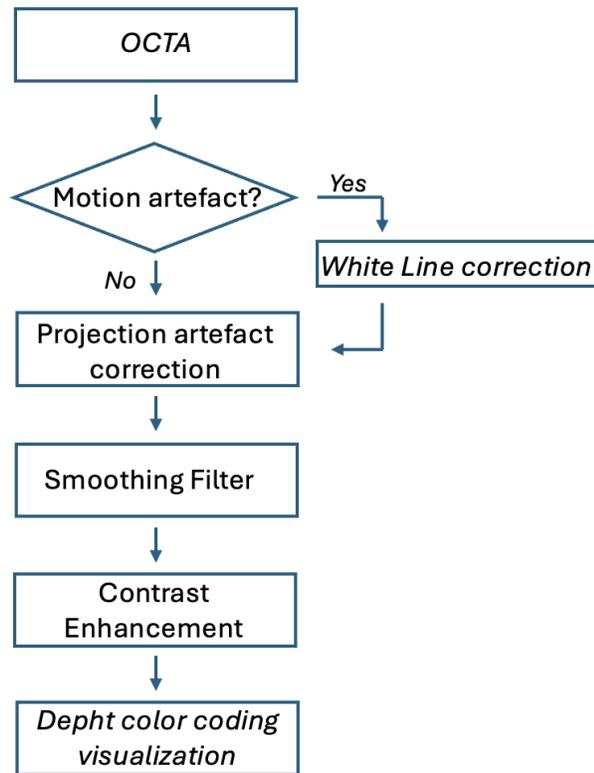


Figure 5.13: Flow chart of the OCTA data improvement process.

5.4 Segmentation and skeletonization

Starting from the volume *CONTRAST* obtained after processing each acquisition, the segmentation and skeleton extraction was carried out using the programmes **Amira** and **MATLAB**, a necessary step for the subsequent calculation of the parameters.

The steps for the two approaches adopted will be described:

- **3D volume:** the volume was considered as a whole, extracting a binary mask and a skeleton, also three-dimensional, in order to calculate 3D parameters.
- **2D Stratification:** Wanting to separate the different information on surface and deep layers, and with the aim of calculating 2D parameters, segmentation was carried out by layer, and then 2D masks and skeletons were extracted.

5.4.1 Three-dimensional whole-volume processing

The volume *CONTRAST* was uploaded to Amira, and immediately went through three editing operations, before proceeding with the segmentation:

- **Crop Editor:** the size of the voxels and the number of slices (limited to 100) was changed, making the OCTA volume more realistic and eliminating noise in the deeper layers.
- **Convert Image Type:** volume was converted to 8-bit unsigned, bringing the dynamics between 0(black) and 255(white). This saves memory and improves visualisation by highlighting relevant details.
- **Brightness-Contrast:** brightness and volume contrast was adjusted, improving the visibility of structures of interest.

Segmentation was achieved using a method involving the setting of two thresholds. The function *Threshold* allows us to select a range of pixels, defining an upper and lower threshold: pixels within this range are set to 1, those outside the range to 0. In this way, a binary mask is obtained, isolating the regions of interest to us, i.e. the vessels. Three other tools were then used to refine the segmentation:

- *Smooth Label*: function used to smooth the edges of a segmentation, which may appear irregular. By applying a smoothing algorithm, the edges appear more regular, improving the look of the segmentation. When necessary, the function was applied to the overall volume, using a 3x3x3 kernel size.
- *Remove Islands*: function removes isolated regions that may be related to noise or imperfections in segmentation. It can be applied either to the overall volume or to individual slices. It has often removed islands of 100-200 pixels in size that do not represent vascular information but are exclusively related to acquisition noise.
- *Brush*: The function allows the segmentation to be modified manually using a brush. It has been applied on individual slices, mainly to eliminate artefacts in the surface areas related to water, air bubbles or glass.

The resulting binary mask is saved as a *3D.tif* file and subsequently imported into MATLAB. For the three-dimensional case, skeleton extraction was chosen in both Amira and MATLAB. The parameters were calculated using both software in order to compare the results and check if the values were comparable.

The **skeletonisation on Amira** is simply obtained with the function *Centerline Tree*, which, from the segmentation, extracts the centre line from each vessel, representing the vascular network as a tree of lines.

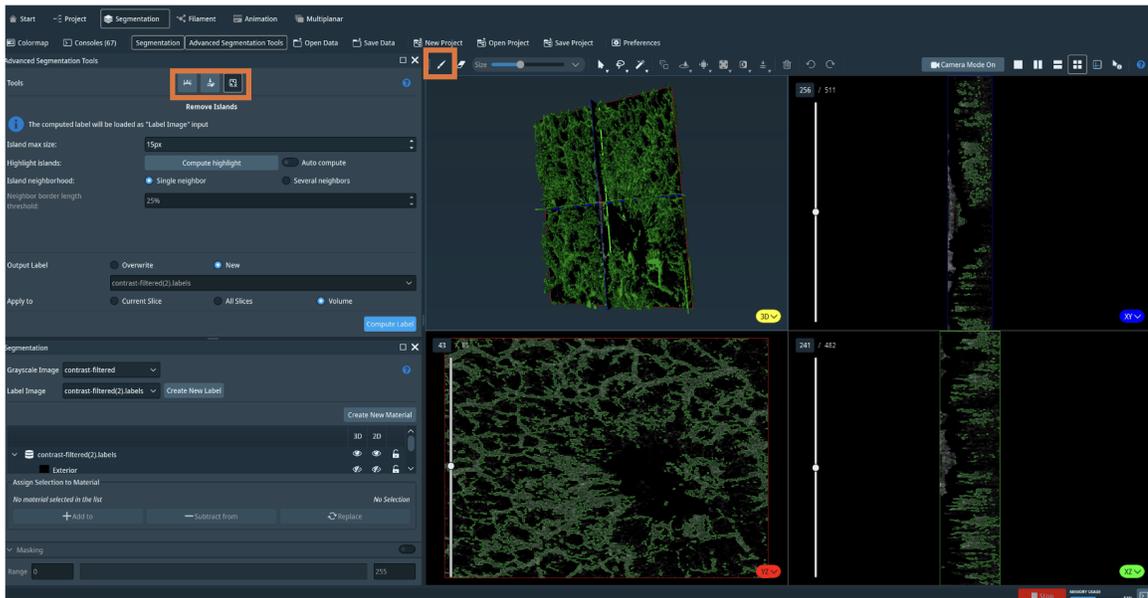


Figure 5.14: Amira interface in the volume segmentation section. The main tools used to carry out segmentation have been squared off in orange. The parts in green are the segmented parts of the volume.

On the other hand, the file *3D.tif* of the segmentation is uploaded to **MATLAB**, converted into a three-dimensional matrix, and given as input to the *bwskel* function: this extracts the skeleton, allowing a parameter to be set that imposes a minimum length of 4 voxels for each branch, thus eliminating the shorter ones. A subsequent cleaning is then performed, removing errors or small unwanted structures.

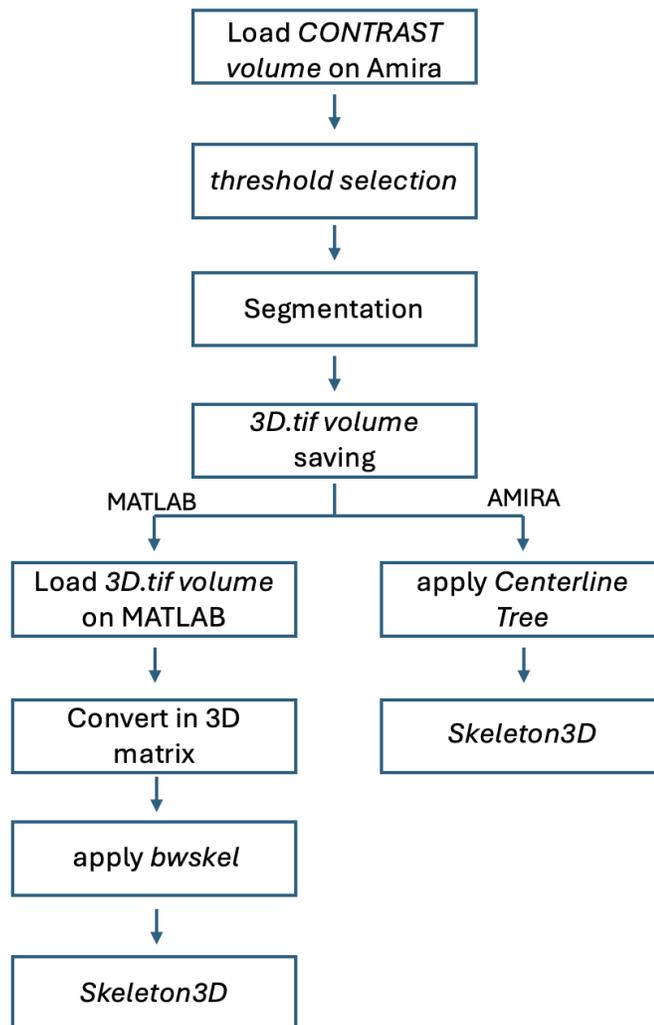


Figure 5.15: Flow chart of the three-dimensional skeleton segmentation and extraction process

5.4.2 Superficial and deep layering for 2D analysis

The segmentation was subsequently repeated, distinguishing between surface and deep layers. Using a MATLAB visualisation, of the 100 slices considered, the slice range containing the surface and deep information was identified for each acquisition. The group of surface and deep slices were then loaded separately into Amira proceeding, as described in the previous section, with the extraction of the segmentation for each layer. Once the two segmentations were obtained and saved as *3D.tif* files, they were loaded into MATLAB, and for each layer, from the respective segmentation, they were calculated:

- **Mean Intensity Projection (MIP):** represents the 2D projection of the volume along its depth.
- **2D Binary Mask:** The MIP calculated above is converted into a binary mask, where values above a certain threshold are set to 1, those below to 0. The threshold is customised for each acquisition.
- **Skeleton 2D:** With the function *bwskel*, the skeleton is calculated from the binary mask. As in the three-dimensional case, the parameter imposing the minimum branch length to be considered is set to 4 voxels, for both the surface and the deep layer.

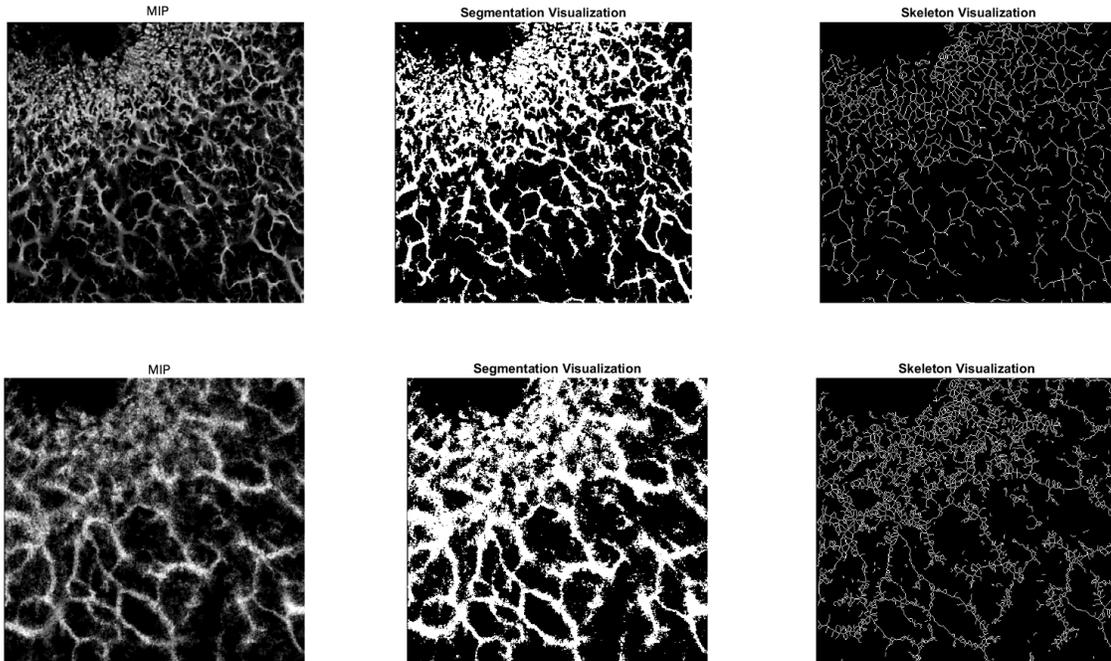


Figure 5.16: From left to right: MIP- segmentation (binary mask with white-vessel, black-background)- skeleton. Top: images related to the surface laser. Bottom: Images referring to the deep laser.

The choice to customise the threshold is based on the desire to obtain the best possible results for each acquisition. Although this compromises the automatism of the method, it guarantees greater accuracy and efficiency in data analysis. The thresholds adopted are in the range $0.20-0.40$ for *surface layers*, $0.50-0.80$ for *deep layers*. The values are generally very similar to each other, especially when considering the two layers separately. This suggests that, in the future, an approach could be developed that identifies an optimal standard threshold for each layer, thus making the method automatable without sacrificing the quality of the analysis.

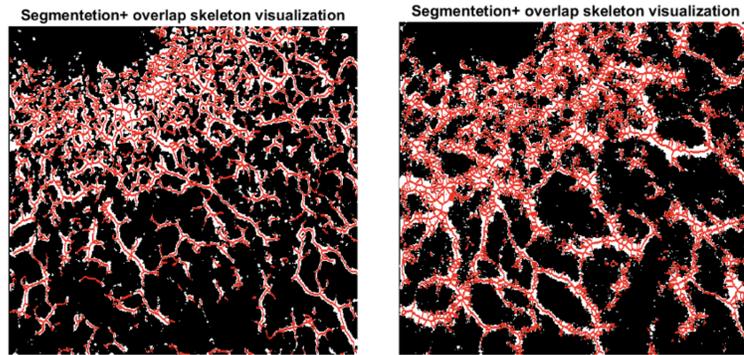


Figure 5.17: Visualization of overlap of binary mask- in white- and skeleton- in red. Left: surface layer. Right: deep layer.

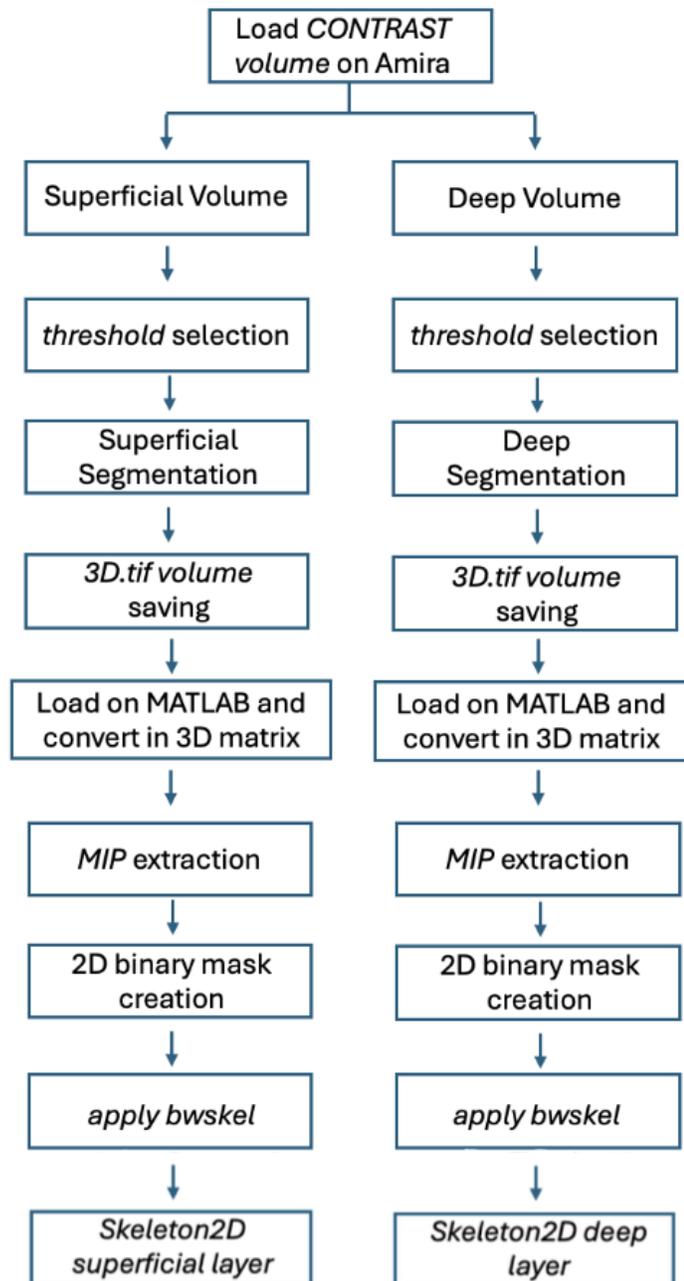


Figure 5.18: Flow chart of the two-dimensional skeleton segmentation and extraction process for superficial and deep layer.

Chapter 6

VASCULAR PARAMETERS EXTRACTION

As reported in the first chapter, the work of the thesis aims at exploiting OCT and OCTA to first of all distinguish between healthy and diseased patients, but above all to find a method to show a difference between actinic keratosis and its two successive tumour evolutions: BCC, a more moderate form of tumour, and SCC, an aggressive and invasive cancer.

The **first road** it was decided to take is that of **OCTA analysis**, going to extract **vascular parameters** that can describe the vascular network in detail, revealing the significant differences between tumour and non-tumour architecture. To justify the extracted parameters, it is described below how the vascular network tends to change due to the presence of tumour, highlighting how the more advanced the tumour stage is, the more pronounced the changes are.

6.1 Characteristics of the Tumor Vascular Network

Angiogenesis is the biological process by which new blood vessels are formed from a pre-existing vascular network. This phenomenon is common because it allows tissue development and regeneration; unfortunately, however, it is also often associated with tumour progression. In fact, in the presence of a tumour, the control of angiogenesis is altered: tumour cells release angiogenic growth factors, such as VEGF (vascular epithelial growth factor), which stimulate the proliferation and displacement of endothelial cells, inducing the blood vessels to grow towards the tumour and thus resulting in its complete vascularisation.

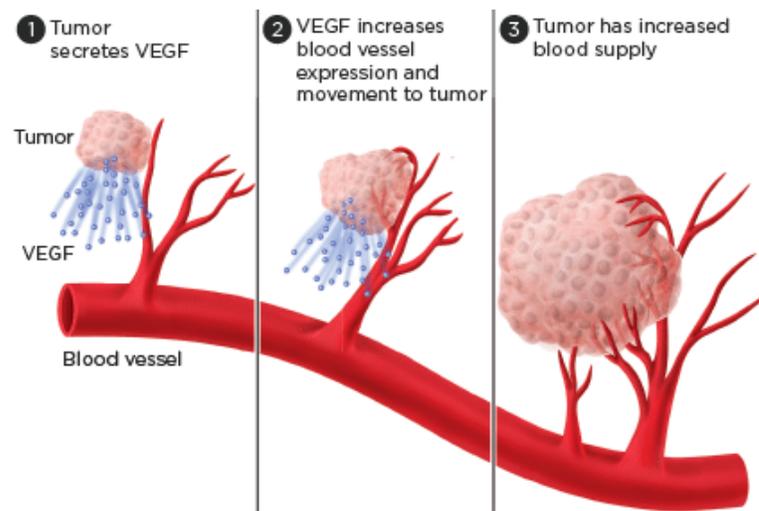


Figure 6.1: Tumor angiogenesis process.

The new vascular network that forms supports tumour growth by providing oxygen and nutrients, and facilitates tumour metastasis to other tissues [9].

In a tumour context, the vasculature tends to show significant changes compared to that of healthy tissue. Angiogenesis in tumours leads to vascular networks characterised by both structural and functional abnormalities[9]. New tumour vessels are typically disorganised, chaotic, with pronounced branching and tortuosity. They miss the normal hierarchy of arterioles, capillaries and venules, and often have direct connections between vessels, making the network irregular. There is an increase in diameters, which contribute to non-uniform perfusion of tumour cells, and a very heterogeneous vascular density, with a much more chaotic arrangement of vessels in the centre of the tumour than at its edges[10].

Carcinomas of the skin, such as basal cell carcinoma (BCC) and squamous cell carcinoma (SCC), show distinctive vascular features that can be used to identify and distinguish between the two tumour types.

BCC and Bowen’s disease are some of the most common skin tumours. Although they have a low tendency to metastasise, they require a well-developed vascular network to support their growth, which tends to infiltrate deep structures destroying

underlying tissues. Their capacity for local invasion may be related to their microvasculature: histopathology shows that the *density of microvessels* in these tumour types is significantly higher in the peritumoral stroma than in normal skin or other benign tumours. Moreover, the vasculature is shown to be much more *tortuous*[11].

The most common vascular pattern for BCC is the sparse pattern, characterised by vessels with an irregular and diffuse *distribution*, and of larger *diameter* than other tumour types. Instead, the most common vascular structure is shown to be arborizing vessels[12].

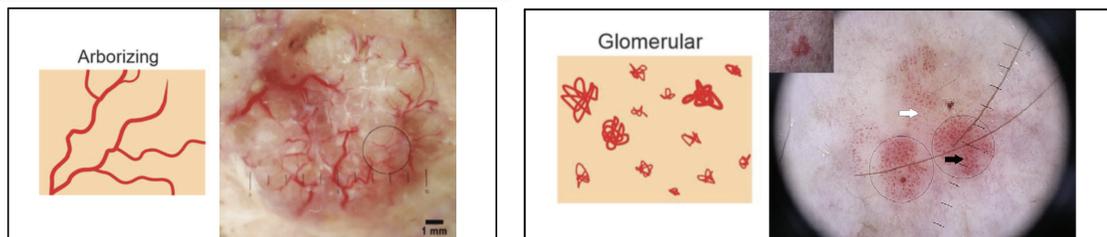


Figure 6.2: Left: graphic representation and dermoscopic image of a vascular structure with arborizing vessels. Right: graphic representation and dermoscopic image of glomerular vessels.

The milky-pink background, often described as a red-white, translucent area, has been found in much superficial BCC; however, it correlates with the vasculature of the lesion and is in fact also present in actinic keratoses, SCC and Bowen’s disease, as well as punctuated vessels.

Glomerular vessels, defined as clustered, tortuous capillaries, are shown to be typical in Bowen’s disease[11, 12].

The **SCC** has a more aggressive behaviour than the other tumour types described above, and is in fact characterised by a more complex and diffuse vascularisation. It has a high vascular density, with dilated vessels, *multiple branches* and high tortuosity. This tumour explores different mechanisms of vascularisation that allow more rapid transport of oxygen and nutrients, unlike BCC which depends mainly on the

proliferation of vascular endothelium[13].

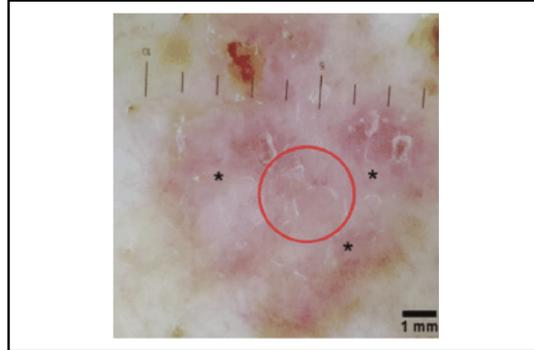


Figure 6.3: Dermoscopic image of a lesion characterized by Milky-pink background.

Actinic keratosis, on the other hand, is a precancerous lesion and therefore has a less complex vascular structure than BCC and SCC: the vessels tend to be less numerous, with a lower degree of tortuosity.

The analysis of vascular features, with the recognition of certain structures, can be fundamental for the differential diagnosis of the above-mentioned pathologies. Extracting the significant parameters, which characterise a vascular network, can allow early recognition of a tumour condition: density, tortuosity, number of branches and degree of disorganisation of the network can be indicators to distinguish a healthy skin from a pathological one, and the different degrees of aggressiveness of tumour lesions.

6.2 Parameter List

Given the vascular characteristics described above, it was decided to go out and extract **12 parameters** that could describe the OCTA vascular networks, in order to discriminate between the four groups presented. The parameters were carefully se-

lected on the basis of their potential to provide useful information for differentiation.

Maximum Radius

Tumors tend to stimulate the formation of new blood vessels through the process of angiogenesis, resulting in vessels that may be abnormal in size. As previously described, the pathological conditions of interest, representing different tumor stages, are also characterized by vessels of varying calibers. Therefore, the maximum radius parameter was selected to help identify differences across these conditions. The maximum radius was calculated by following the steps outlined below:

- the contour of each vessel in the binary OCTA mask was identified;
- The function *bwstgeodesic* then calculates the geodesic distance from each point on the vessel skeleton to the vessel contour.
- The distance values are filtered to preserve only positive and finite values, which represent the vessel radius at each point on the skeleton.
- By applying the *max* function to the vessel radius, the maximum value is obtained.

Number of Vascular Trees (NT) and Number of Branches (NB)

In healthy tissues, vascularization is highly organized, with a moderate number of branches and a few main vascular trees that spread out in a regular pattern. In contrast, the presence of numerous vascular trees and a high number of branches may indicate structural disorganization, often associated with pathological tissues. Greater is the network complexity, worse is the condition.

The *Number of Vascular Trees* is calculated from the skeleton using the function *bwconncomp*: this identifies all the groups of white pixels (value 1) connected to each other. Each connected component is considered a vascular tree, so by counting the

total elements found, the total number of NTs is obtained.

To calculate the *Number of Branches*, the *searchbp-function* function is used instead: this allows to identify the branch points in the main vessels. It is then checked if there are small isolated branches, or structures that do not have end points, to add to the previously calculated branch points. Once all the branch points have been found, the *bwconncomp* function is used to detect the connected components, representing each branch point as an object. The total number of these objects is assigned to NB.

Fractal Dimension (FD)

Calculating the Fractal Dimension (FD) is useful because it provides quantitative information about the complexity and structure of blood vessels.[9]. A high FD may indicate a greater complexity or disorganization of the vascular system, helping to identify structural differences between healthy vascular networks or different pathologies at different stages. The parameter was calculated using the **Box Counting Method**[15]:

- the image is divided into a grid of boxes of side r : initially r is large, but at each iteration it is halved becoming smaller and smaller;
- for each dimension r , we count the number of boxes that contain at least one pixel of the vascular structure. This count is denoted as $N(r)$;
- a log-log graph of the $N(r)$ values versus r is constructed;

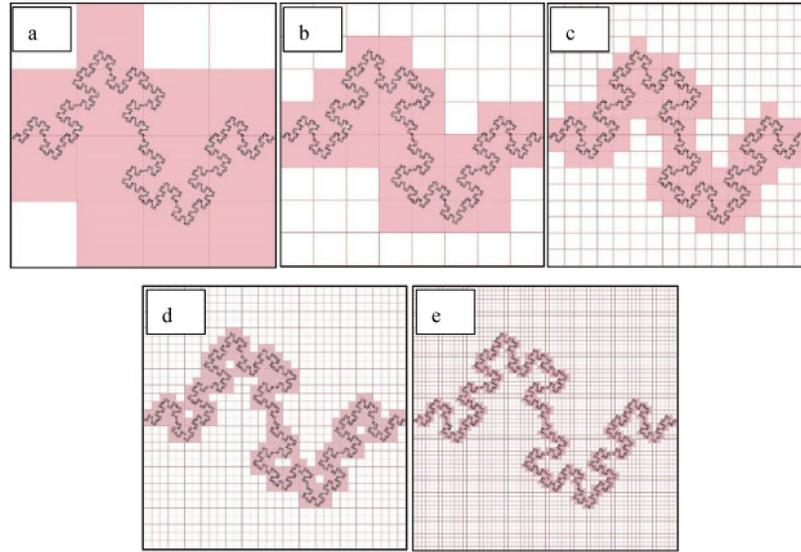


Figure 6.4: Working diagram of the Box Counting Method: notice how from (a) to (e) the boxes into which the image is divided become smaller and smaller. Also, boxes containing useful pixels of the image are shown in pink: these are the boxes that are counted.

The FD is obtained as the negative slope of the straight line that fits the log-log curve in the linear section:

$$FD = - \lim_{r \rightarrow 0} \frac{\log(N(r))}{\log(r)} \quad (6.1)$$

The log-log plot was calculated on 10 points, but not all acquisitions showed the same linearity. In some acquisitions, the curve was linear for a larger number of points, suggesting a higher FD and therefore a greater complexity of the structure. In other acquisitions, however, linearity was present only for a small number of points.

To account for this variation, it was decided to normalize the fractal dimension. The FD is calculated as the slope of the line that approximates the linear portion of the log-log curve, but is then adjusted based on the actual number of points that show a linear behavior. To do this, the FD is multiplied by a normalization index, obtained by dividing the number of linear points by the total number of points in the plot.

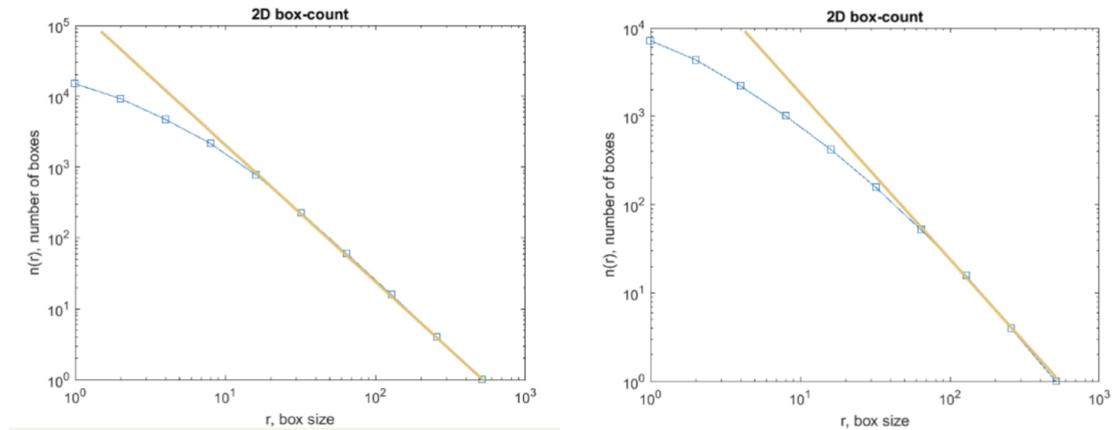


Figure 6.5: An example of the log-log graph described above is illustrated in the figure, showing on the *x-axis* the box size r , on the *y-axis* the number of boxes containing useful pixels $N(r)$. It can be seen that in the graph on the left, the linear behavior is satisfied by 6 out of 10 points, indicating a higher complexity of the structure studied. In the right graph, only 4 out of 10 points are linear, which means that the structure will have a lower fractal dimension.

Vascular density (VD)

In many pathological cases, the tissue develops a denser vascular network to support the abnormal growth of cells. Analyzing the vascular density therefore allows us to detect areas with a greater or lesser concentration of blood vessels, helping to differentiate between healthy tissue and the various tumor stages. A high vascular density may indicate greater pathological activity.

The vascular density is calculated both from the skeleton (VD-skel), and from the segmentation mask (VD-mask), as a ratio between the number of pixels containing vessels (value equal to 1), and the total number of pixels in the image.

Entropy

Entropy is a statistical measure that quantifies the degree of disorder or uncertainty in a structure or distribution. In the case of vasculature, entropy can reflect the geometric complexity and branching of vessels: greater is the entropy, more chaotic and disordered the vascular network will be. It is particularly useful for discriminating

between healthy and pathological tissues and for monitoring changes in vasculature in the event of tumor progression.

Entropy is calculated using the following formula:

$$H = - \sum p(x) \cdot \log_2(p(x)) \quad (6.2)$$

where $p(x)$ represents the probability of belonging to each class.

It should be noted that the parameter was calculated in two ways:

- Entropy-skel: starting from the skeleton. $p(x)$ indicates the probability of belonging to class 0 (background) or 1 (vessel);
- Entropy-mip: starting from the MIP. In this case $p(x)$ represents the probability that each pixel belongs to a certain intensity level (or gray tone).

Avascular Area(AV)

The introduction of the avascular area as a parameter is related to the fact that actinic keratosis, basal cell carcinoma, Bowen's disease and squamous cell carcinoma lesions are often characterized by the presence of a crust. This crust, due to its opaque nature, hinders the ability of OCT to detect the underlying vascularization. As a result, black areas corresponding to regions devoid of vascularization are observed in OCTA images (Figure 6.6). The quantification of these avascular areas therefore becomes a useful parameter, especially in the context of distinguishing between healthy and diseased patients, since it allows to identify the presence of lesions.

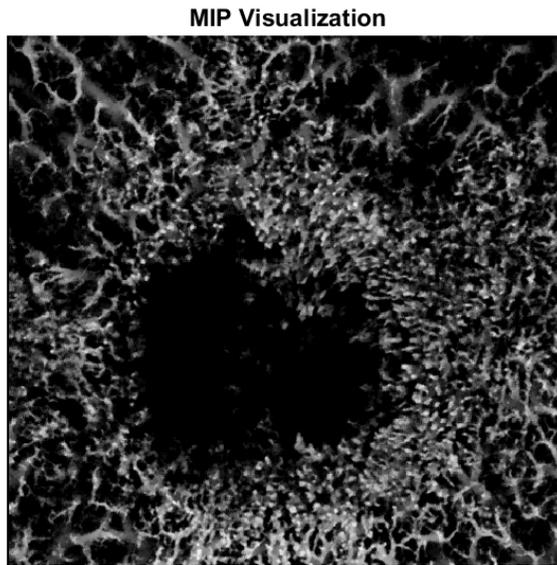


Figure 6.6: Example of acquisition in which, due to the crust of the lesion not allowing visualization of the vasculature, there is a black area in the image.

The percentage of avascular area is calculated from the segmentation: the complement of this was calculated, obtaining an image in which the vessels (originally represented by values 1) become black, while the background (originally represented by values 0) becomes white. Subsequently, using the MATLAB function *regionprops*, the areas of the connected regions in the complementary image, which correspond to the avascular zones, were calculated. In this way, it was possible to quantify the area without blood vessels.

Tortuosity Parameters

Early lesions and tumors often cause changes in the vascular network. In fact, through angiogenesis, tumors tend to produce new vessels: these are characterized by an increase in tortuosity, that is, a greater irregularity in the path they follow.

Quantification of tortuosity is therefore shown to be a useful parameter for our purposes. In particular, we chose to calculate:

- **Distance Metric (DM)**: quantifies the length of a path with respect to the direct distance between the starting point and the arrival point. If DM is greater than 1, the path is tortuous. If it is equal to 1, it is instead straight. The DM is therefore calculated as:

$$DM = \frac{\text{Euclidean distance between the beginning and the end of the path}}{\text{path length}} \quad (6.3)$$

- **Inflection Count Metric (ICM)**: measures the complexity of a curve along a path. In particular, it quantifies the inflection points, or points where the direction of a path changes significantly. The higher the ICM, the greater the curvature and complexity of a branch.

For each point of the path, the vectors A and V (direction of movement and direction between two consecutive points) are calculated, and their variation is used to determine the presence of an inflection. The ICM is obtained with the following formula:

$$ICM = \text{number of inflections} \times DM \quad (6.4)$$

where the number of inflections is obtained by adding the most significant inflections.

- **Sum of Angles Metric (SOAM)**: it is a measure that takes into account the angles formed between consecutive segments of the path. The higher the angles between the segments, the more irregular and complex the path.

To calculate the SOAM three vectors are defined:

- T1: This vector is oriented along the path, i.e. it connects the current point with the previous one;

- T2: represents the direction between the current voxel and the next one along the path, so it indicates the direction of the next segment respect to the considered point.
- T3: is the direction between the considered voxel and the second subsequent voxel.

These three vectors are used to calculate:

- the in-plane (IP) angles between the directional vectors T1 and T2 for each path segment;
- the torsional angles (TP) between the vectors T1, T2 and T3 using the cross product between the vectors.

The calculation of SOAM is done using the following formula:

$$\text{SOAM} = \frac{\sum |\text{CP}|}{\sum \text{T1}} \quad (6.5)$$

where CP is the sum of the in-plane and torsional angles of each segment; the denominator is the sum of the lengths of the segments.

6.3 3D Parameter Extraction

Once the parameters were chosen, we proceeded with their extraction starting from the three-dimensional volume. As previously mentioned, only in this case the parameters were calculated both with the MATLAB software and with Amira. In particular, in the Amira program, starting from the calculated 3D skeleton, a function, *Spatial Graph Statistics*, was applied, which allowed the extraction of a series of morphological parameters. Among these, the most significant ones for our analysis were selected:

- Number of Segments;
- Mean Radius;
- Branching Nodes;
- Mean Tortuosity;
- Mean Volume;
- Mean Chord Length.

Moving on to MATLAB, after loading the segmentation obtained from Amira in *3D.tif* format, and after calculating the 3D skeleton using the *bwskel* function, the following parameters were extracted:

- Max Radius;
- VD-skel;
- VD-mask;
- NT;
- NB;
- DM mean;
- ICM mean;
- SOAM mean;

The first three parameters were calculated considering the segmentation and the three-dimensional skeleton as a whole. On the contrary, for the others listed, it was decided to divide the total volume into ROI, small volumes of size 30x120x128, and

calculate the parameters on these subvolumes, increasing the accuracy of the result. The overall value of NT and NB was obtained by adding the various NT and NB calculated on the individual ROIs; for the tortuosity parameters instead, the average of the values calculated on the individual volumes was considered as the final value. The two groups of parameters were calculated on the overall Dataset, therefore on each acquisition, and saved in two Excel files, one for the MATLAB parameters and one for those of Amira.

The use of both software for the extraction of the parameters, even if with overlapping information, was a deliberate choice to obtain a comparison of the measures resulting from the two different approaches. The main motivation lies in the uncertainty of the calculation of parameters on three-dimensional volumes, which can lead to non-superimposable results. **Comparing the values obtained in fact, it was noted that the measures were not always comparable, highlighting the limits in the accuracy of the 3D analysis.**

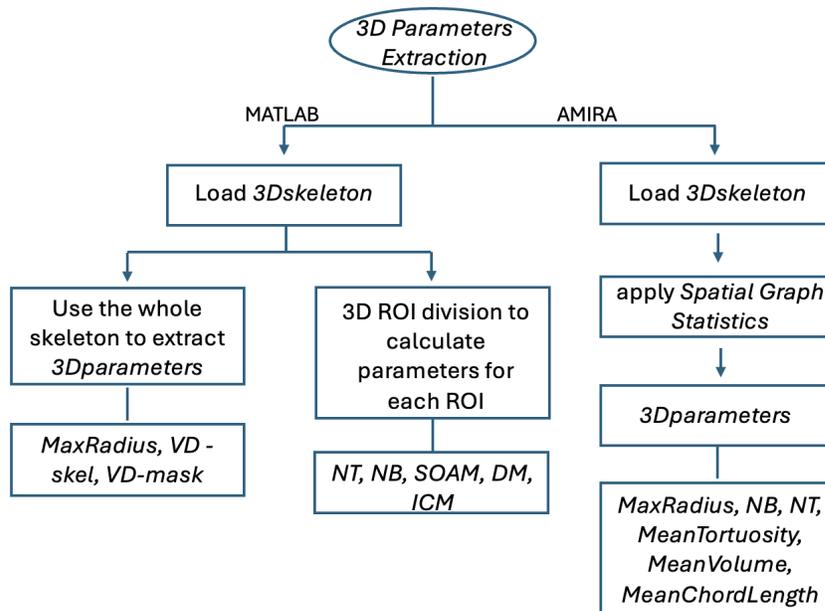


Figure 6.7: Flow chart of 3D parameter extraction with Amira and MATLAB.

For this reason, it was decided to **discard the 3D analysis and proceed with a two-dimensional analysis**, in which the segmentation and skeleton are more defined, allowing for greater accuracy in the calculation. This approach offers a more reliable result, reducing the influence of the uncertainty that is characteristic of the three-dimensional analysis.

6.4 2D Parameters Extraction

As justified in *Chapter 5*, for the two-dimensional case it was decided to split the analysis considering separately the **Surface and Deep Layers**.

After extracting the segmentations with Amira, the two *3D.tif* files were loaded into MATLAB, where MIP, binary mask and skeleton were extracted, respectively for the two layers.

The parameters:

- Fractal Dimension;
- Max Radius;
- Percentage of Avascular Area;

were extracted considering the 2D image as a whole. In particular, the FD and the maximum radius starting from the skeleton, while the avascular area starting from the segmentation.

The remaining parameters were instead calculated becoming the overall image in ROI, sub-portions of size 120x128: this allows to understand and highlight how the parameter varies within the image, obtaining a more accurate extraction.

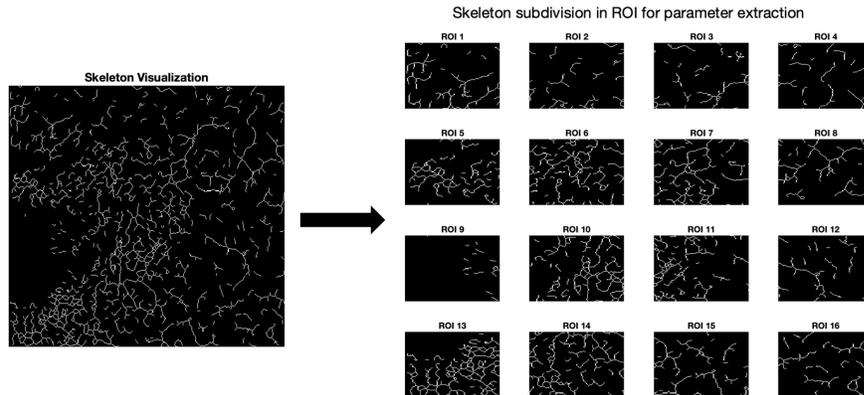


Figure 6.8: Left: representation of an entire skeleton from the constructed dataset. Right: subdivision of the skeleton into 16 2D ROIs for parameter calculation.

In particular, the parameters:

- NT;
- NB;
- DM;
- ICM;
- SOAM;
- VD-skel;
- entropy-skel;

were calculated from the skeleton,

- VD-mask;

starting from the segmentation mask,

- entropy-MIP;

starting from the MIP.

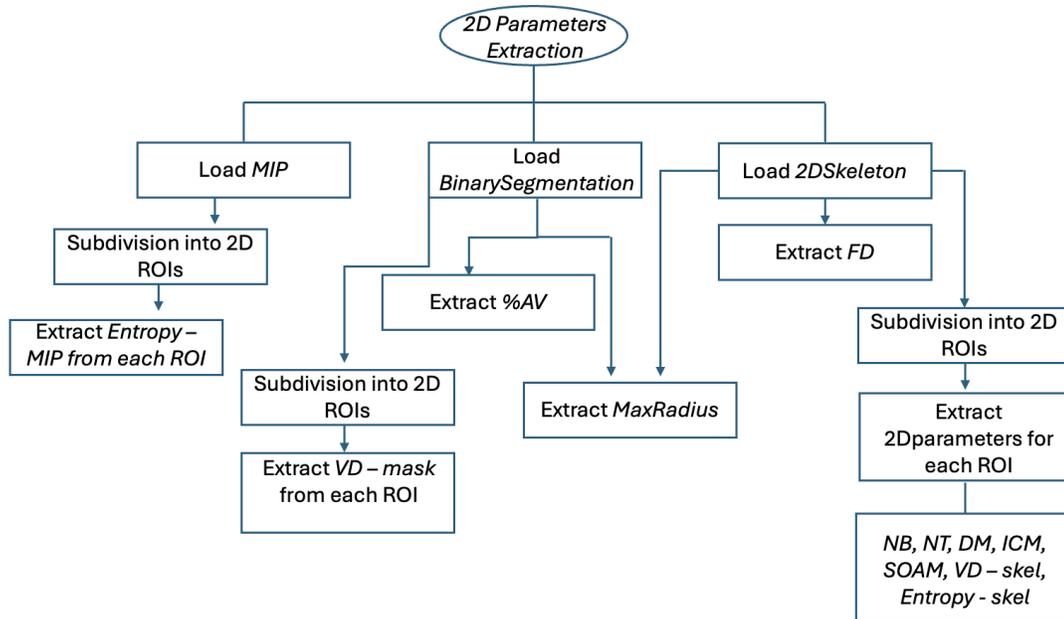


Figure 6.9: Flow chart of 2D parameter extraction with MATLAB. The described procedure was repeated for surface and deep volumes respectively.

Since the overall image may include areas characterized by varying degrees of vascular disorganization – with more chaotic and disorganized regions alongside more structurally regular ones – it was decided to further investigate by creating **Heat Maps** (Figure 8.2) for each parameter on the two layers.

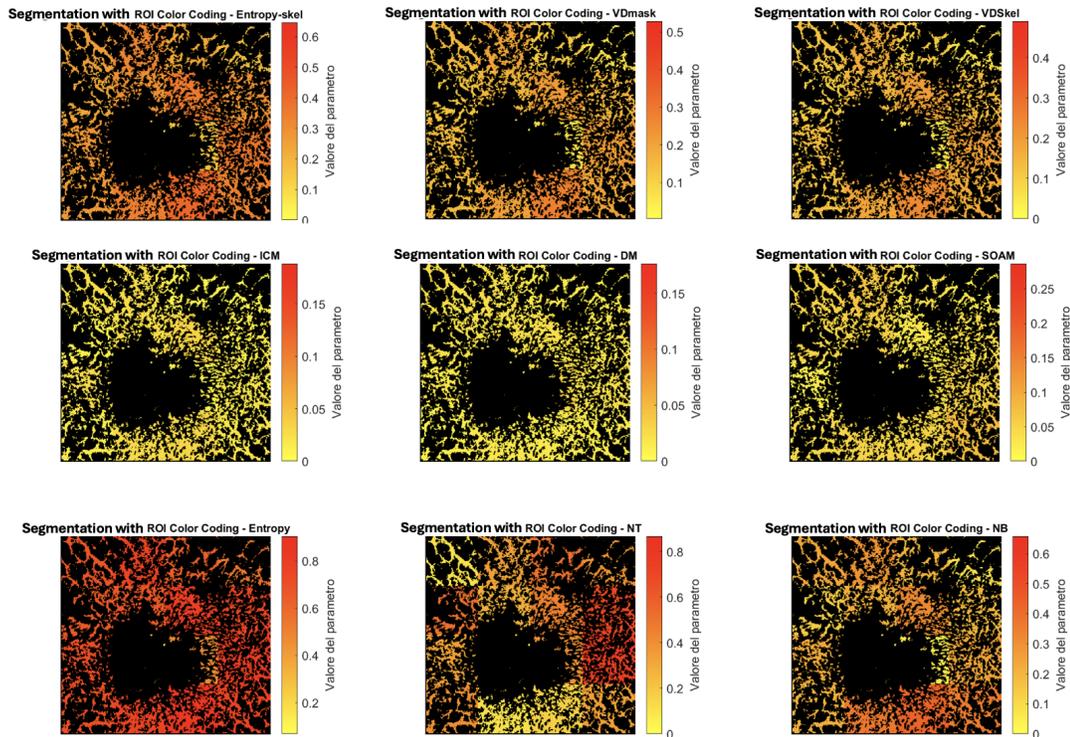


Figure 6.10: Visualization of heat maps constructed for each parameter. In this case the maps are referred to a superficial layer.

These maps allow to visualize the local trend of the vascular parameters within the image, highlighting the differences between more and less altered areas. Each ROI was colored according to the calculated parameter value, using a chromatic scale that goes from red for the highest values to yellow for the lowest values.

To build the heat maps, a **global normalization** process was adopted, which considered the entire data set for both layers (superficial and deep). In particular, for each parameter the maximum and minimum value of the entire dataset was calculated, then the *min-max scaling* method was applied to normalize each value. This choice allows, by displaying the heat maps, a qualitative comparison of the data between the different acquisitions, allowing to search and highlight vascular variations between actinic keratosis and its tumor progressions.

While the first three parameters can be considered almost constant, since for each acquisition only one value is obtained for image, the other parameters present a spatial variability. In fact, since the parameters are calculated on different ROIs, 16 distinct values are obtained per image, one for each ROI. For this reason, it was decided to collect and analyze these parameters according to three different criteria: the **average value, maximum value, and range width** of the 16 values of each image. The goal is to analyze the parameters in these three different perspectives in the hope that they highlight significant differences between the four groups of interest, where the average alone may not be sufficient to capture the variations present in the different conditions.

The parameters were calculated 4 times, where at each step a data improvement modification was introduced.

6.4.1 Step1: Raw Dataset

The above parameters, as previously described, were extracted considering the Raw Dataset, i.e. all acquisitions (51) of all 14 patients.

Thanks to the results of the biopsies performed by the doctors, it was possible to divide the dataset into 4 groups of interest: Healthy, Pre-Cancer, In situ Cancer, Metastatic Cancer. In Table 6.1 the total number of data available for each group is reported.

Table 6.1: DATASET

	NUMBER OF PATIENTS	TOTAL NUMBER OF ACQUISITIONS	NUMBER OF MIP SUP	NUMBER OF MIP DEEP
METASTATIC CANCER	2	6	6	6
IN SITU CANCER	5	10	10	10
PRE CANCER	7	23	23	23
HEALTHY	2	12	12	12

6.4.2 Step2: Dataset and Data Deep Cleaning

Observing the processed data, it was noticed that the extraction of parameters was negatively affected by the presence of very noisy data in the dataset. It was therefore decided to perform:

- **Dataset Cleaning:** acquisitions that were too noisy or rich in motion artefacts, air bubbles or surface water, which covered the useful vascular information of the OCTA, were eliminated from the dataset;

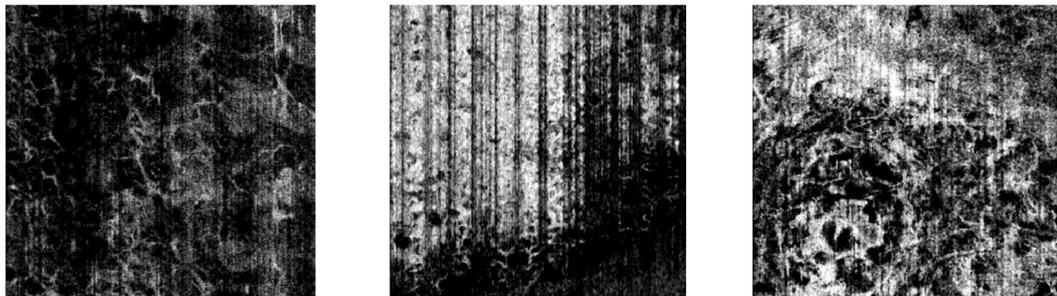


Figure 6.11: Example of noisy and artefact-rich data removed from the Dataset.

- **Deep Data Improvement:** as discussed in the previous chapters, the analysis of each volume was initially limited to the first 100 slices. However, looking at the extracted MIPs, it was found that the result was not very accurate due to the presence of noise in the deeper slices, often devoid of significant information. To improve the quality of the segmentation, it was therefore decided to reprocess all the deep data with Amira, including for the segmentation extraction only the deep slices that contained useful information and excluding those dominated by noise. This approach allowed to obtain significantly more accurate segmentations and skeletons.

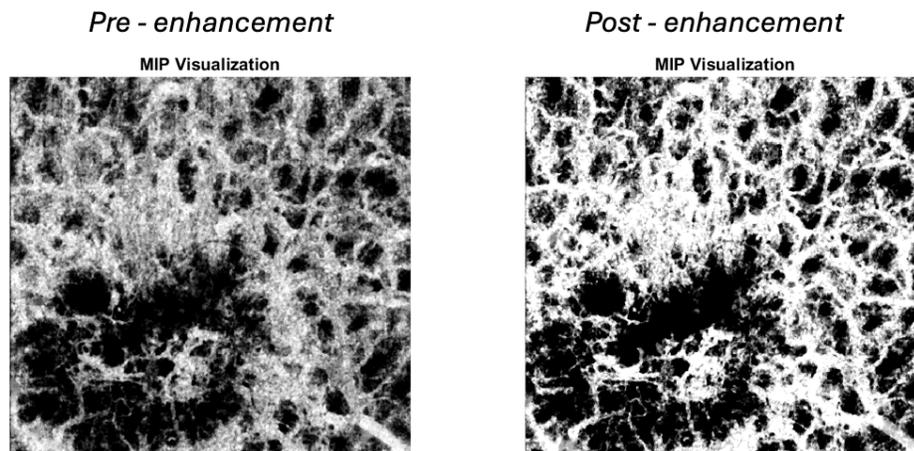


Figure 6.12: Example MIP pre and post elimination of the noisiest deep slice in the volume. It can be seen from the image that the vessels are better defined.

In Table 6.2 the new dataset is reported, with the number of useful acquisitions considered for each group.

Table 6.2: CLEAN DATASET

	NUMBER OF PATIENTS	TOTAL NUMBER OF ACQUISITIONS	NUMBER OF MIP SUP	NUMBER OF MIP DEEP
METASTATIC CANCER	2	6	5	5
IN SITU CANCER	5	10	9	6
PRE CANCER	7	23	23	19
HEALTHY	2	12	12	9

6.4.3 Step3: Normalization with BodyIndex

Vascular parameter extraction is used to describe vascular networks, highlighting the differences between healthy tissues and lesions of different nature. However, a significant limitation has been found: parameter values do not depend solely on the presence or stage of the lesion, but are also influenced by the anatomical position in which the lesion is located. Vascularization varies significantly in different areas of the body: the overall structure of the vascular network adapts according to the functional and metabolic needs of the specific region. For example, the vasculature of the head and neck differs in density and morphology from that of the leg, which has larger but less dense vessels.

Not considering this anatomical diversity, there is a risk of attributing variations in parameters to the stage of the pathology, when in reality they could be explained by intrinsic differences in the vascularization of the different areas. For this reason, it was decided to introduce a **normalization index-Body index** that allows to adapt the parameter values according to the anatomical position. This index aims to make the parameters comparable to each other, reducing the variations due to the acquisition

area and thus allowing a more objective comparison of the lesion status.

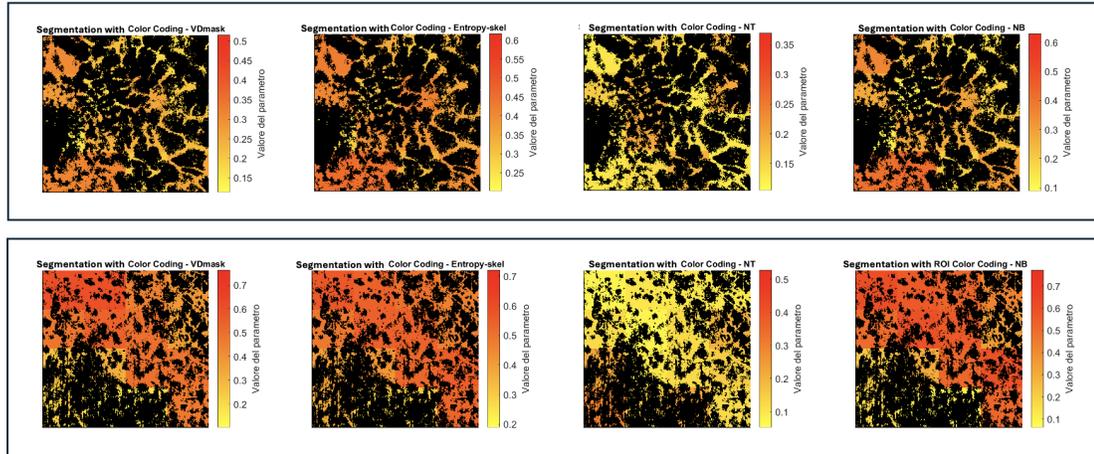


Figure 6.13: *Top: heat map corresponding to a lesion on the leg. Bottom: heat map of a head lesion. The two lesions both belong to the class ‘metastatic-cancer’. From the figure it can be seen that, especially for vascular density, NB, entropy, the values are higher in the lesion on the head, where a denser and richer vasculature is expected due to its anatomical position.*

The main acquisition areas of the studied lesions include head, forehead, face, ear/-neck, sternum, shoulder and leg. The vascular parameters most influenced by the anatomical position were found to be: vascular density, entropy, maximum radius, fractal dimension, number of branches and vascular trees. On the contrary, parameters such as tortuosity and avascular areas were excluded from the normalization, since they reflect more directly an intrinsic characteristic of the tumor lesion and less the basic vascular structure.

The normalization index was designed to balance the parameters based on the characteristics of the vascularization in the different anatomical areas:

- *Forehead-Head-Face:* The vascularization in these areas[16] is typically characterized by a very dense network of small capillaries. Here, the values of vascular

density, entropy, number of vascular trees and branches, as well as the fractal dimension, tend to be naturally high. Therefore, the normalization index reduces these parameters, bringing them back to levels comparable with areas where the vascularization is more rarefied, such as the leg.

- *Leg*: In this area, the blood vessels are less dense, but of a larger caliber to support adequate blood flow. In this case, the maximum radius parameter has been reduced by the index, to allow a fair comparison with other areas.
- *Neck-ear*: These areas have an intermediate vascularization, where the density and number of branches are greater than the leg, but less than the head and forehead. For these areas, the index applies an intermediate normalization that rebalances the parameter values.
- *Sternum-shoulder*: The vasculature in these areas is considered fairly standard, for this reason the Body-index was not applied, believing that the values could already be considered comparable with the others.

The Table 6.3 shows the index values adopted for each parameter based on the anatomical area.

Table 6.3: BODY INDEX

	Rmax	FD	VD	NB	NT	Entropy	%AV
Leg	0.65	1	1	1	1	1	0.65
Forehead	1	0.8	0.65	0.65	0.8	0.8	1
Head	1	0.8	0.65	0.65	0.8	0.8	1
Ear	1	0.8	0.75	0.75	0.8	0.8	1
Face	1	0.8	0.8	0.8	0.8	0.8	1
Sternum	0.65	1	1	1	1	1	0.65
Shoulder	0.65	1	1	1	1	1	0.65

6.4.4 Step4: Skeleton Improvement

It was observed that the skeleton of deep layers was less accurate. Ideally, the skeleton should represent the centerline of each vessel; however, in larger vessels, multiple branches were often generated within the same vessel, leading to an overrepresentation of the vascular structure. To solve this problem it was first acted on the *segmentation mask*, applying a *majority operation* on the pixels considering the neighbours (usually in the 3x3 window):

- if most neighbours are 1 (white), the pixel becomes 1;
- if most neighbours are 0 (black), the pixel becomes 0.

This allows both to eliminate noise, but also to close small holes in the image. It was therefore decided to adapt the *minimum branch length parameter* of the function that calculates the skeleton on an individual basis:

- Deep layer: it was decided to increase the value of the parameter, going from eliminating branches with less than 4 pixels to eliminating those with less than **24 pixels**, obtaining a more precise skeleton.
- Surface layer: For the most accurate skeletons, the parameter remained set at **4 pixels**; where the skeleton over-represented vessels, the parameter was increased to **24 pixels**;

In addition, further cleaning of the skeleton was carried out by removing isolated pixels, most likely due to noise.

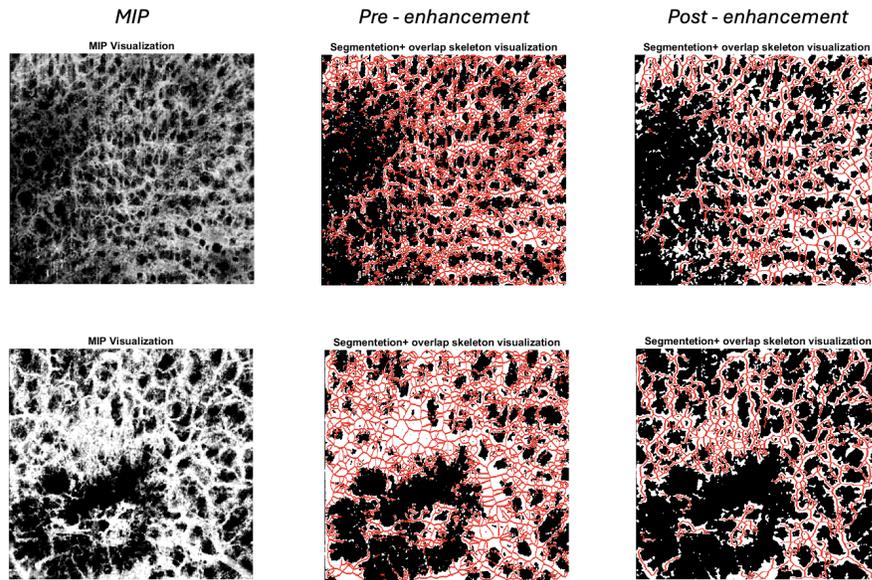


Figure 6.14: Example of the result of skeleton improvement. It can be seen from the image that previously the skeleton tended to overrepresent vessels. By adjusting the *bwskel* parameter, a more accurate result can be obtained.

Chapter 7

PARAMETER ANALYSIS

For each group of extracted parameters (Range, Mean, Maximum) and for each step, an Excel file was created. The rows represent the acquisitions for each patient, while the columns contain the extracted parameters. In particular, the first 12 columns contain the parameters relating to the superficial layer, while the last 12 contain those of the deep layer.

These Excel files, structured in this way, were used in MATLAB to perform three different statistical and classification analyses, with the aim of evaluating the validity and potential of the extracted data. In particular, the t-test showed significant results on both the superficial and deep layers. However, to investigate further, it was decided to conduct further analyses, such as the MANOVA test and classifiers, to assess whether the combination of the different parameters could improve the ability to discriminate between groups. For the purpose of completeness, however, the analyses were also applied to the surface layer data.

7.1 T-test and boxplot representation

As a first step, all possible pairs that could be realised from the four groups under analysis were considered, as shown below:

- METASTATIC CANCER-HEALTHY;
- IN SITU CANCER-HEALTHY;
- PRE CANCER-HEALTHY;

- METASTATIC CANCER-IN SITU CANCER;
- METASTATIC CANCER-PRE CANCER;
- IN SITU CANCER-PRE CANCER;

For each of these pairs, and for each of the 12 calculated parameters, **boxplots** were created in MATLAB, which are useful for visualising the distribution of the data and making comparisons. Boxplots are important as they summarise the main characteristics of a data distribution. They are characterised by:

- A central box, where the upper and lower bounds represent the first and third quartiles, respectively. The length of the box is the interquartile range, containing the middle 50 per cent of the data;
- Line inside the box, representing the median, i.e. the central value of the data;
- Whiskers, which extend the box to the maximum and minimum data values;
- Outliers, isolated points represented with asterisks.

Within each pair, a **T-test** was performed for the superficial and deep data, respectively, to test whether actually the observed difference between the parameters was statistically significant or dictated by random chance.

The independent t-test used, applying the MATLAB function *ttest2*, is a statistical test that determines whether a statistically significant difference exists between the averages in two independent groups, and consists of the following steps[17]:

- Null and alternative hypothesis formulation: with the null hypothesis (H0) it is assumed that there is no difference between the averages of the two groups; the alternative hypothesis assumes that there is a statistically significant difference between the averages.

- Based on mean, variance and sample size, a t-value is calculated using the t-test;
- The t-value is translated into a **p-value**, which indicates the probability of obtaining the observed data if the null hypothesis is true. Based on the p-value, a decision is made whether or not to reject H_0 :
 - $p < 0.05$: rejects H_0 concluding that the difference between the two groups is statistically significant;
 - $p > 0.05$: there is insufficient evidence to reject H_0 , and therefore nothing can be said about the significance of the data.

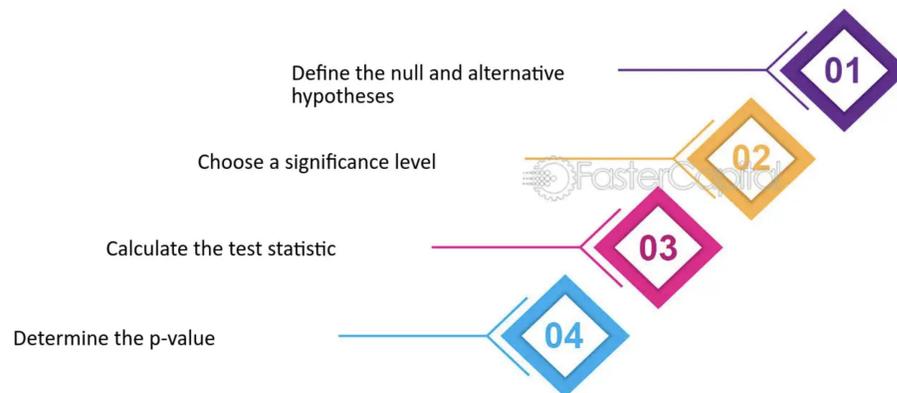


Figure 7.1: Schematisation of T-test steps.

Boxplot visualisation and the application of the T-test are extremely useful tools because they allow us to understand how parameters vary between the different groups, and above all with which parameters we can discriminate healthy conditions, actinic keratosis and its subsequent progressions.

7.2 MANOVA test

The previous analysis makes it possible to assess the usefulness of each parameter in distinguishing between the different pairs of groups. However, it was considered useful to go further and test whether a combination of parameters, thus considering the information in an integrated manner always on the two layers, could offer a greater contribution in discriminating between different pathological and non-pathological conditions. The **Multivariate Analysis of Variance (MANOVA)** considers several dependent variables together, combining them into a single variable obtained from a weighted linear combination of the individuals. This allows us to assess how the dependent variables (for us, the parameters), taken as a whole, change according to the levels of the independent variable (the groups to be discriminated), providing a more complete and detailed analysis of the relationships between variables.

The MANOVA test was conducted through the following steps, now described[18]:

- Hypothesis testing: In order to perform the MANOVA test, the input data must satisfy certain hypotheses. The most problematic condition of the data extracted in the study is that the data must be *uncorrelated* in order to perform the test. For this reason, by applying MATLAB's *corr* function and visualising the 12-parameter matrix, only those parameters were selected where the correlation with the others was low or moderate (correlation coefficient < 0.7 [19]), not considering those with a high correlation. This resulted in a subgroup with respect to the initial parameters.

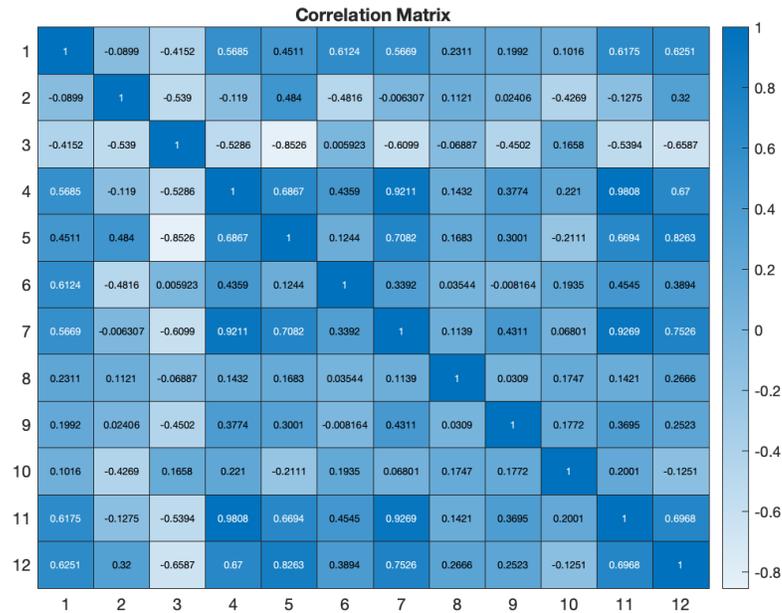


Figure 7.2: Example of the correlation matrix with colour coding. Intense blue or very light colours, towards white, indicate a high correlation between the parameters.

- MANOVA test: the test was performed with MATLAB's *manova1* function.

The function requires as input:

- X: matrix of independent variables (parameters to be analysed). Each column represents a parameter, while each row represents an acquisition;
- Y: vector indicating to which group to discriminate each acquisition belongs;

and provides as output:

- p-value, which indicates the statistical significance of the analysis;
- stats: structure containing test-related statistical variables such as Wilks' Lambda and canonical vectors.

- Interpretation of the output: Wilks' lambda is an indicator of the "force" of the separation between the groups based on the dependent variables. Low values of lambda (close to 0), together with a p-value < 0.05 , indicate that the linear combination achieved by associating a specific weight with each parameter significantly differentiates the input groups. In contrast, higher values of lambda (close to 1), with a p-value > 0.05 , show that the combination does not statistically discriminate the input groups. In the case of statistical difference, it was also chosen to analyse the canonical loadings, which represent the contribution of each parameter in the linear combination separating the groups: higher absolute values of the loadings indicate a greater contribution of the parameter in identifying the significant difference of the input data.

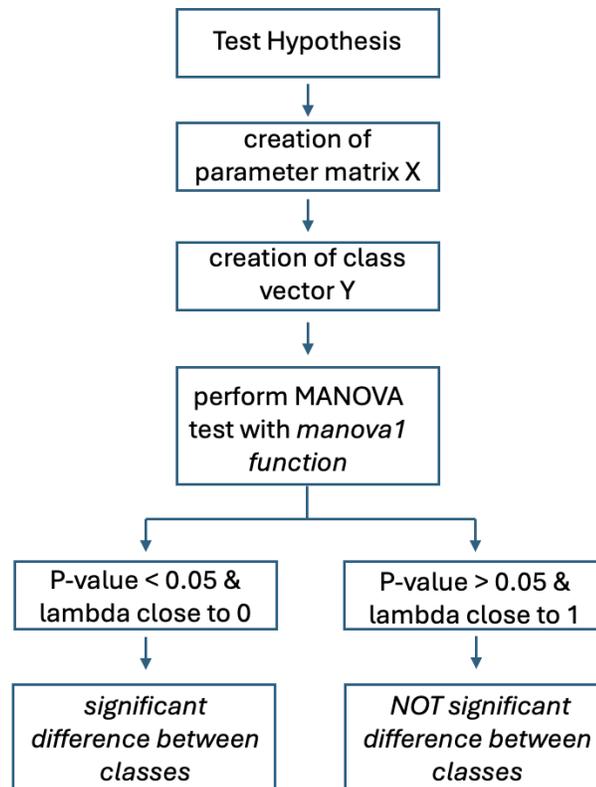


Figure 7.3: Flow chart of MANOVA test.

The MANOVA test can also be carried out by giving as input not only the individual parameters, but also their interaction (e.g. product between pairs of parameters), in order to test whether or not such interactions can further support discrimination.

The test was repeated for the parameter groups (Range-Mean-Max):

- with and without parameter interaction;
- considering the correlated parameters, and then a subset of these. In fact, observing the trend of the boxplots described above, it was decided, among the correlated parameters, to select only those in which there was a visual difference between the groups, with a behaviour similar to what should be the ideal case (parameters increasing with increasing severity of the pathology);
- for the surface and deep layer data respectively.

7.3 Logistic regression and multiclass ECOC classifier

Always with the idea of exploiting the overall information of all parameters, it was decided to implement:

- **A binary classifier**, which distinguished between healthy and pathological group, by means of logistic regression;
- **A multiclass classifier**, which discriminated between the four groups analysed, also differentiating the three pathological conditions, using the ECOC technique with the *k-fold cross validation* approach.

The following steps were followed for both:

- Hypothesis testing: again, in order to realise the classifier, it is necessary for the input data to satisfy certain conditions. The most problematic proved to be, as with the MANOVA test, the correlation between the parameters; the subgroups of parameters previously selected for the analysis presented above were therefore used, in which the non-correlation was verified;
- Input data: both models required input:
 - Data matrix: characterised by parameters selected above by columns, and acquisitions of the various patients by rows;
 - Group vector: vector that assigns a label (numerical value) to each acquisition according to the group to which it belongs.
- Model realisation and prediction:
 - *Binary classifier*: the function *fitglm* is used, which builds a logistic regression model with a binomial distribution. Based on the input data and the true classes, the function estimates the coefficients of the model with the aim of producing an output as close as possible to the true class. The model is trained on a *training set*, consisting of 80% of the total dataset, providing a balance between classes. Prediction is instead performed by applying MATLAB's *predict* function on the *test set*, consisting of 20% of the remaining data. The division between *training set* and *test set* is done randomly, but ensuring that all acquisitions of the same patient belong exclusively to one of the two sets, thus avoiding that data of the same patient are present in both sets.
 - *K-Fold cross validation classifier*: The model in this case is created using the MATLAB function *fitcecoc*, which applies the ECOC technique. The ECOC method divides the multiclass classification problem into a series of

binary classifications: each pair of classes is treated separately by applying a *SVM base classifier*. A combination of the results is used to determine the final class. Given the limited number of data for some classes, especially metastatic-cancer and in situ-cancer, it was decided to validate the model with the k-fold cross validation method [20]. The technique consists of dividing the dataset into k folds (in the case under analysis 5) of similar size, and using each fold once as a test set and the remaining k-1 times as a training set. At each iteration, the model then trains on the k-1 folds, and tests on the remaining fold; this is repeated k times. Again, the creation of the two sets is done randomly, providing the precautions specified earlier. The great advantage is that with this method, all data is used for both training and validation, which is a plus point having so limited data available.

- Finally, a confusion matrix (CM) is created to evaluate the performance of the model on the test set. In the case of the k-fold, the final CM obtained is the average of the CMs calculated at each iteration.

The classifiers were developed separately for each of the three parameter groups (Range, Mean and Max), considering the parameters of the surface layer and the deep layer, respectively.

Chapter 8

OCT PROCESSING

OCTA analysis may be useful in achieving the aim of this thesis work: the extraction of parameters describing the morphology of the vascular structure may indeed highlight differences between healthy and pathological groups, and especially between pre-tumour, tumour in situ and metastatic tumour vasculature, such as those analysed. However, the results can be greatly influenced by the quality of the acquisitions, making the calculation of the parameters very sensitive to noise.

Looking at the B-scans of each acquisition, it was noted that another important piece of information can be extracted from the OCT volume, especially for discriminating between different pathological conditions. As mentioned above, OCT is an imaging technique that exploits light interference to obtain cross-sections of tissue, where grey tones represent the intensity of the backscattering signal. This intensity is closely related to the structure of the tissue being examined: a healthy tissue tends to show a homogeneous intensity distribution, while a pathological tissue shows variations in intensity right at the lesion.

It is also known that the more advanced a lesion is in its tumour stage, the more it tends to involve tissue in the deeper layers.

It was therefore thought that these variations in intensity may be particularly significant when correlated with the depth of the lesion.

On the basis of these observations, a complementary approach was then proposed to discriminate between the three pathological conditions analysed: to calculate, for each lesion, the depth from the OCT volume, exploiting the variations in intensity of the grey tones that characterise pathological tissues.

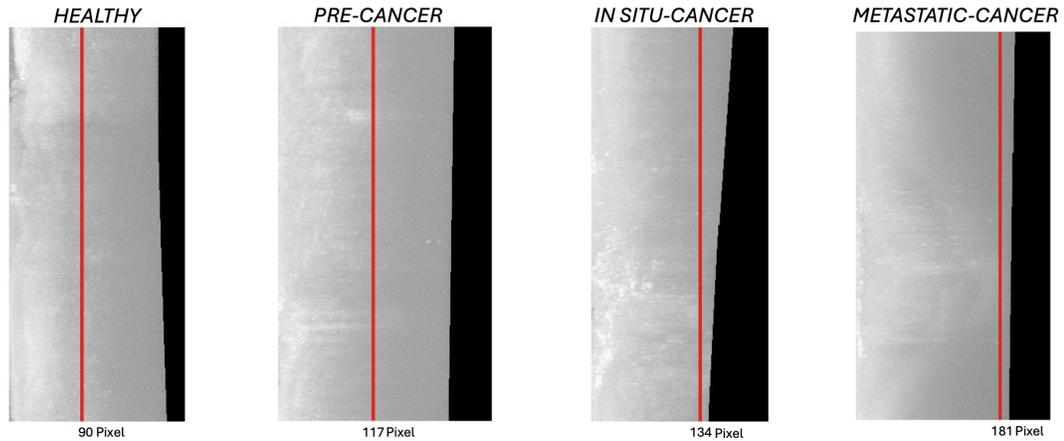


Figure 8.1: Example of the depths for lesions in the four groups in the study. Depth propagates along the horizontal direction; the maximum depth reached by the lesion is indicated in red. It can be seen that the more invasive the tumour, the more the lesion tends to involve the deep layers of the skin.

This parameter could offer an additional element to support differentiation, distinguishing between the different stages of tumour progression.

8.1 Lesion Depth Calculation

The calculation of the lesion depth was carried out by implementing a MATLAB code.

First, for each patient's lesion, an acquisition was identified that was sufficiently centred on the lesion itself. Volume v1, one of the 4 OCT volumes acquired, is subsequently loaded into the Amira software, which allows visualisation and exploration of the volume on different layers along the three dimensions. In particular, after contrast and brightness changes to highlight the intensity variations at the lesion, the B-scan range containing the lesion itself was identified, with the aim of isolating it from the rest of the acquisition not useful for this analysis.

At this point, the v1 volume limited in the selected range is loaded into MATLAB, for then:

- extract a **MIP** along the depth axis (third dimension) of the volume. This process projects the maximum intensity values along the slices, resulting in a representative 2D image of the volume. The MIP is then transposed so that the depth is represented along the horizontal direction (Figure);
- The *imbinarize* function transforms the MIP into a **binary image**, where pixels with intensities above a threshold, customised for each lesion, are set to 1 (white), while those below are set to 0 (black). This mask identifies the lesion with respect to the background.
- By setting a threshold due to noise, it can happen that pixels not belonging to the lesion are selected; for this reason, after a display of the binary mask and MIP, pixels of value 1 in columns outside the lesion are set to zero;
- using the binary mask, the code identifies the pixels belonging to the lesion. The coordinates of these pixels are extracted using the *find* function. In particular:
 - **rows** contains the coordinates of the rows, which represent the depth values of the lesion;
 - **cols** contains the coordinates of the columns, indicating the lateral position.

The maximum column (*col-max*), which represents the deepest point of the lesion, is calculated.

The depth data of each lesion were then collected in a table, which can be found in the *RESULTS* section.

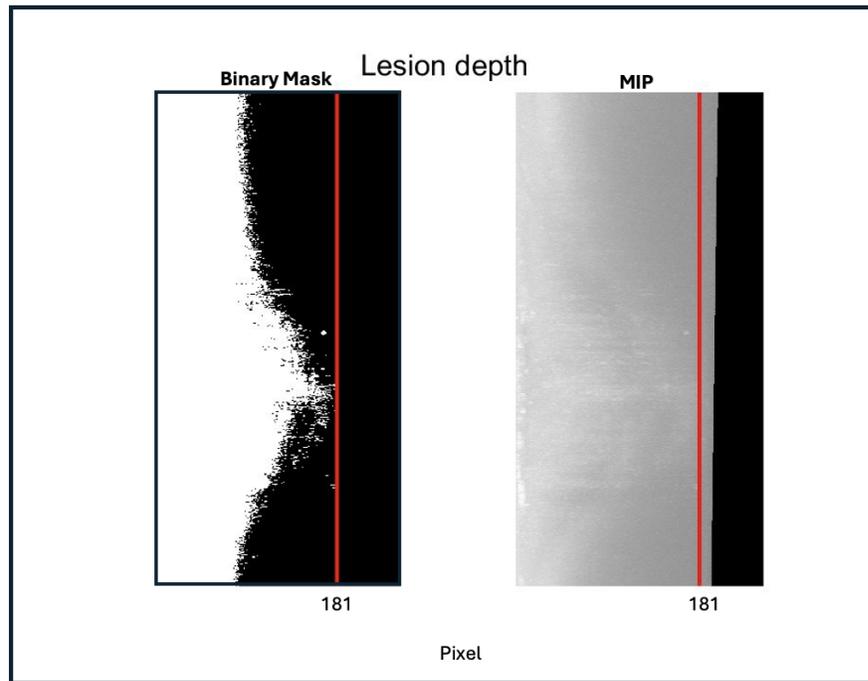


Figure 8.2: Right: the MIP obtained by isolating the B-scans containing the lesion. Left: the binary segmentation mask obtained by setting a threshold on the intensity of the MIP. The red line indicates the maximum depth reached by the lesion, shown along the horizontal direction.

PART III

RESULTS

Chapter 9

RESULTS PRESENTATION

9.1 T-test

As explained in *Chapter 6*, a statistical analysis was conducted on the data collected via t-test.

For each parameter, box plots were constructed in order to represent the overall trend of the data in the four conditions studied. In particular:

- The box plots for the surface parameters are shown on the left- side of each graph;
- the deep parameters are shown on the right, separated from the previous ones by a vertical dotted line.

To improve the visual clarity of the representation, colour coding was used:

- Green- HEALTHY;
- Yellow- PRE-CANCER;
- Orange- IN SITU CANCER;
- Red- METASTATIC CANCER.

In addition, to distinguish the depth layers, lighter shades were used for the deep data and full colours for the surface layer.

For each pair of conditions, a t-test was performed to compare the surface and deep

parameters separately. Where the difference was found to be statistically significant, the box plots of the two groups were connected by a black line with an asterisk, indicating significance. Only plots in which at least one statistical difference is present are reported and analysed.

The resulting plots are shown below, analysing the constant parameters and then Range, Mean and Maximum Value of each parameter separately, with the aim of highlighting the main information obtained from this analysis.

9.1.1 Constant Parameters

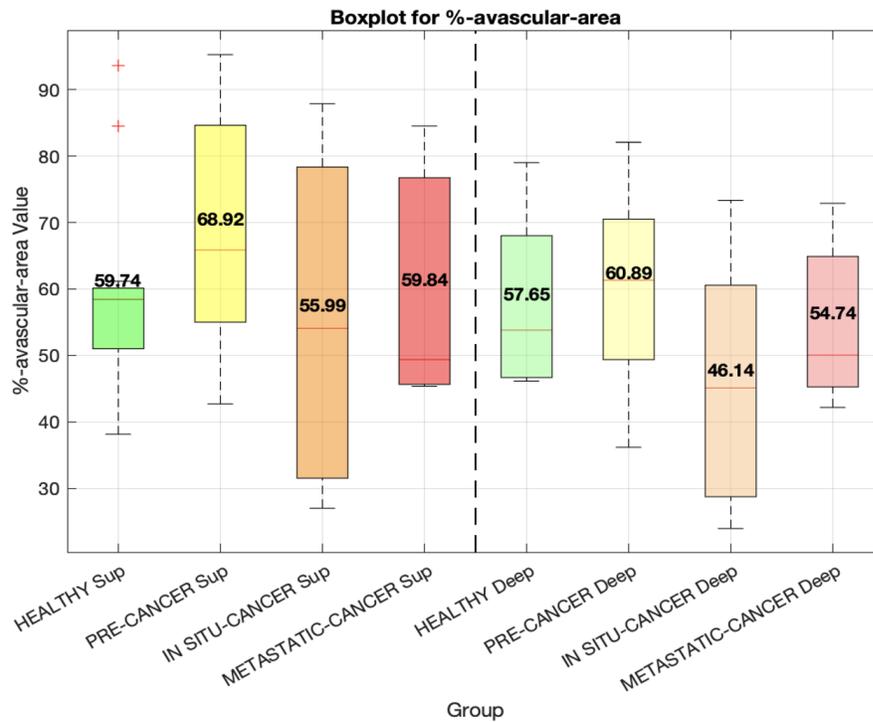


Figure 9.1: Boxplot Range of % Avascular Area in the 4 groups for superficial and deep layers respectively.

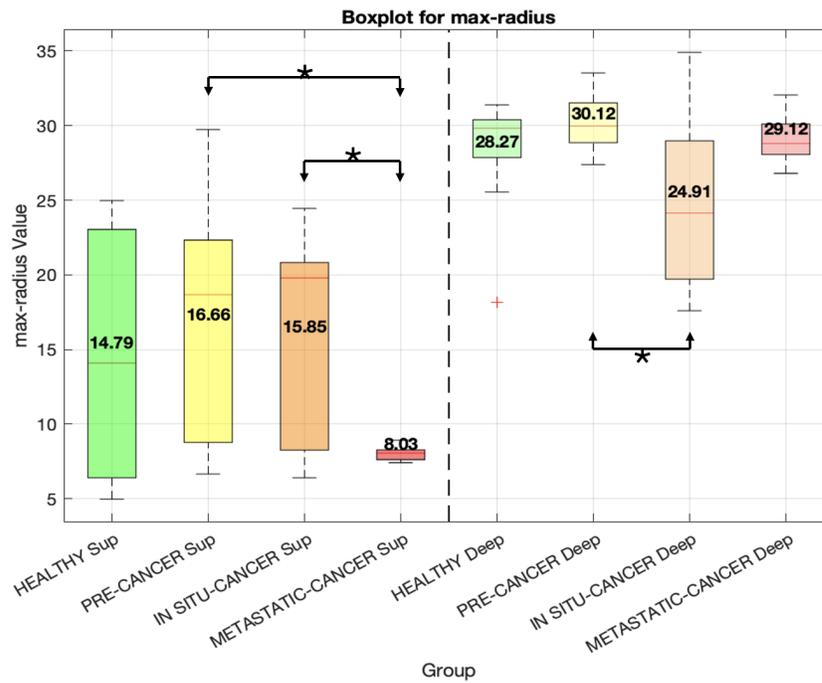


Figure 9.2: Boxplot of Maximum Radius in the 4 groups for superficial and deep layers respectively.

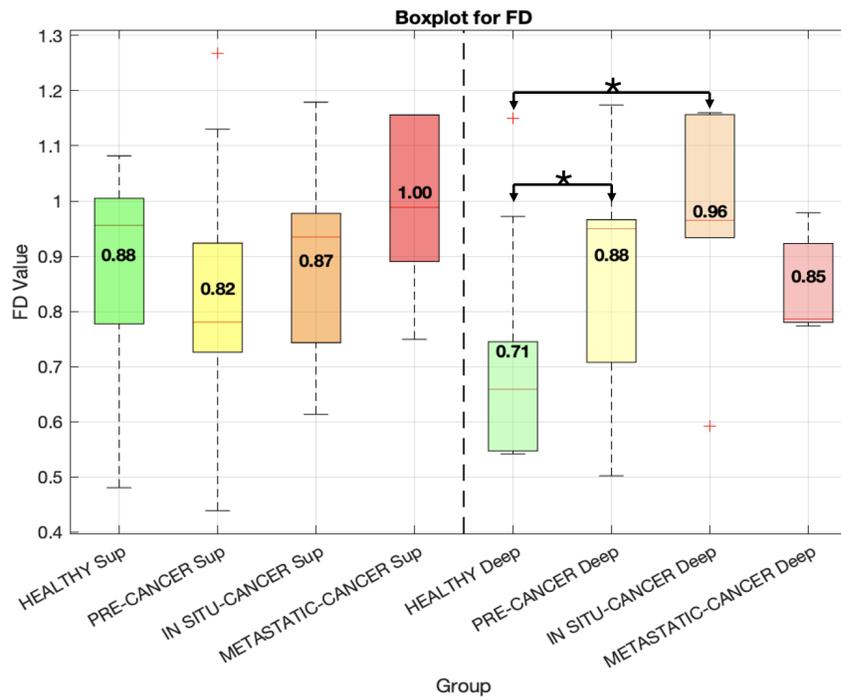


Figure 9.3: Boxplot of Fractal Dimension in the 4 groups for superficial and deep layers respectively.

The boxplots above represent the trends between the 4 groups under analysis of the 3 parameters that were calculated on the overall image: %Avascular Area, Maximum Radius and FD.

Concerning the *Avascular Area*, not only was no statistical difference found to discriminate between the 4 classes analysed, but in addition the trend is not as predicted. Indeed, it was expected that, as the lesions are characterised by a crust, which manifests itself as a black hole in the image, the percentage of avascular area would be greater in pathological acquisitions than in healthy ones. It must be considered, however, that this parameter is strongly influenced by how the acquisition is performed and the chosen ROI: depending on this, the lesion may or may not be central in the image. The trend obtained in general is therefore quite random.

Even for *Maximum Radius*, the overall trend is not as anticipated: ideally it would have been expected that, due to the process of angiogenesis and the tumour's need for continuous blood and nutrient support to grow, the radius would increase as the tumour condition worsens. However, it is noticeable that, in the superficial layer, compared to the healthy, pre-cancer and in situ cancer groups, which show a similar trend, the metastatic cancer group presents a strong decrease in the parameter. However, it is thought that the results obtained were strongly influenced by the quality of the data, which may have caused an inaccurate calculation of this parameter.

Finally, the *Fractal Dimension* shows a trend quite in agreement with expectations: ideally, as the condition worsens, the complexity of the vascular network should increase due to the chaotic vascularisation that the tumour tends to create. We see this ascending trend on both the superficial and deep layers, except in the latter case for the metastatic-cancer group, again due to poor data quality. We also see how the fractal dimension allows for statistical differences especially in discriminating between healthy and pre-cancer/in situ cancer conditions.

9.1.2 Range

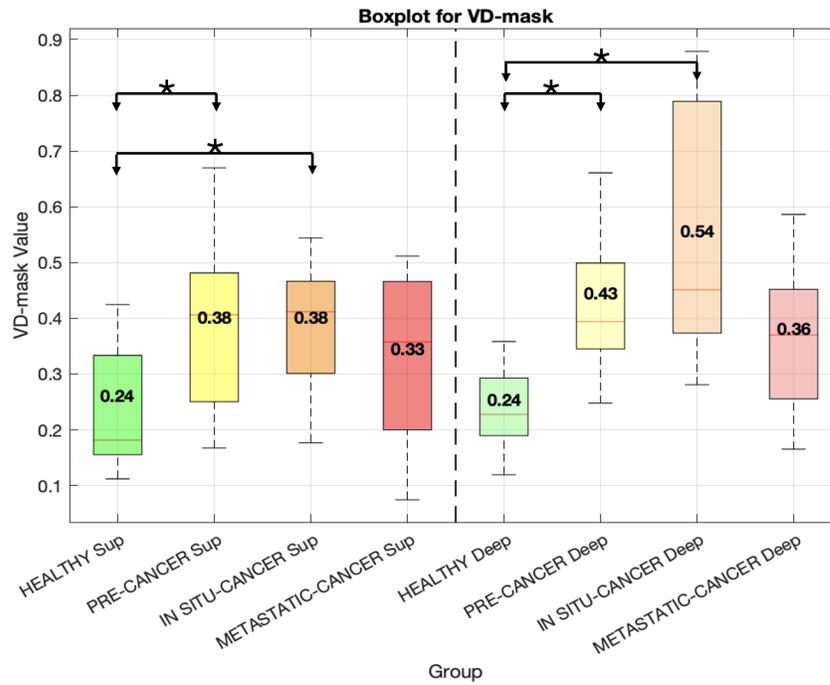


Figure 9.4: Boxplot Range of Vascular Density calculated on the segmentation mask in the 4 groups for superficial and deep layers respectively.

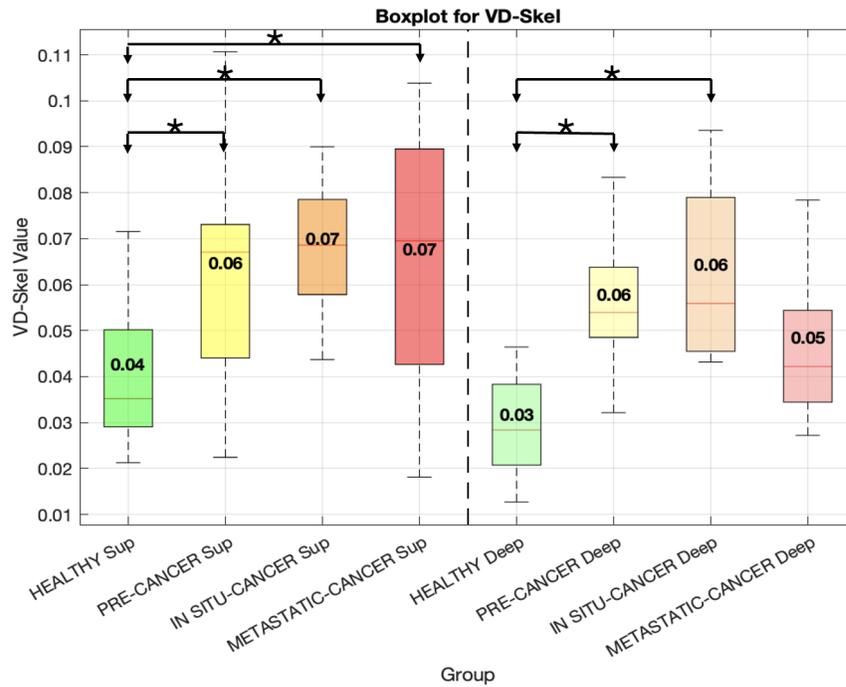


Figure 9.5: Boxplot Range of Vascular Density calculated on the skeleton in the 4 groups for superficial and deep layers respectively.

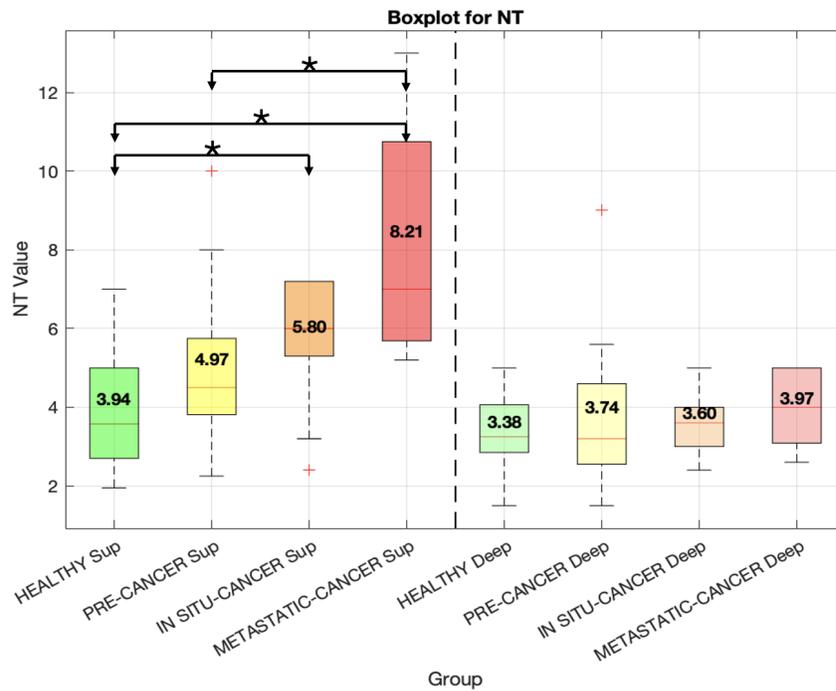


Figure 9.6: Boxplot Range of Number of Trees in the 4 groups for superficial and deep layers respectively.

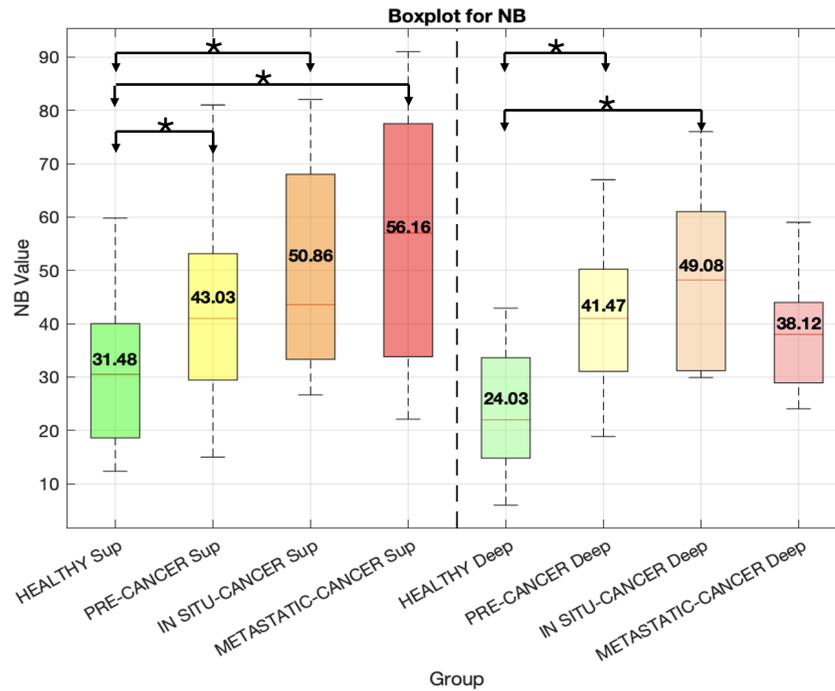


Figure 9.7: Boxplot Range of Number of Branches in the 4 groups for superficial and deep layers respectively.

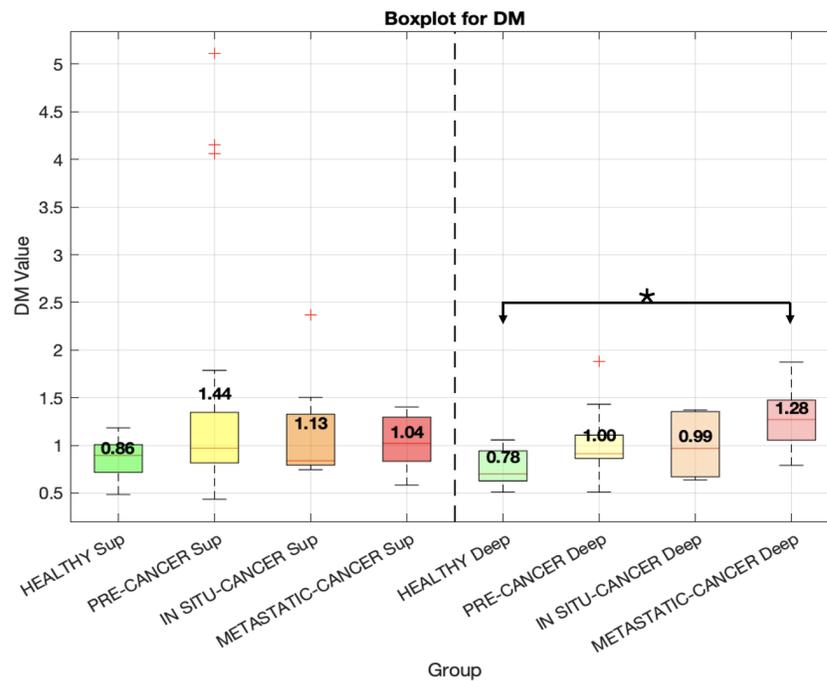


Figure 9.8: Boxplot Range of Distance Metric in the 4 groups for superficial and deep layers respectively.

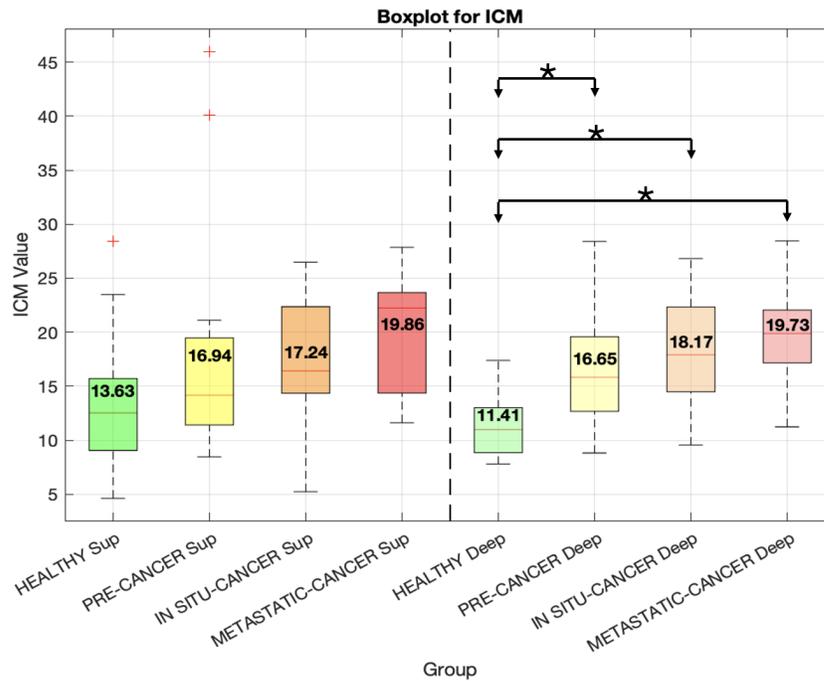


Figure 9.9: Boxplot Range of Inflection Count Metric in the 4 groups for superficial and deep layers respectively.

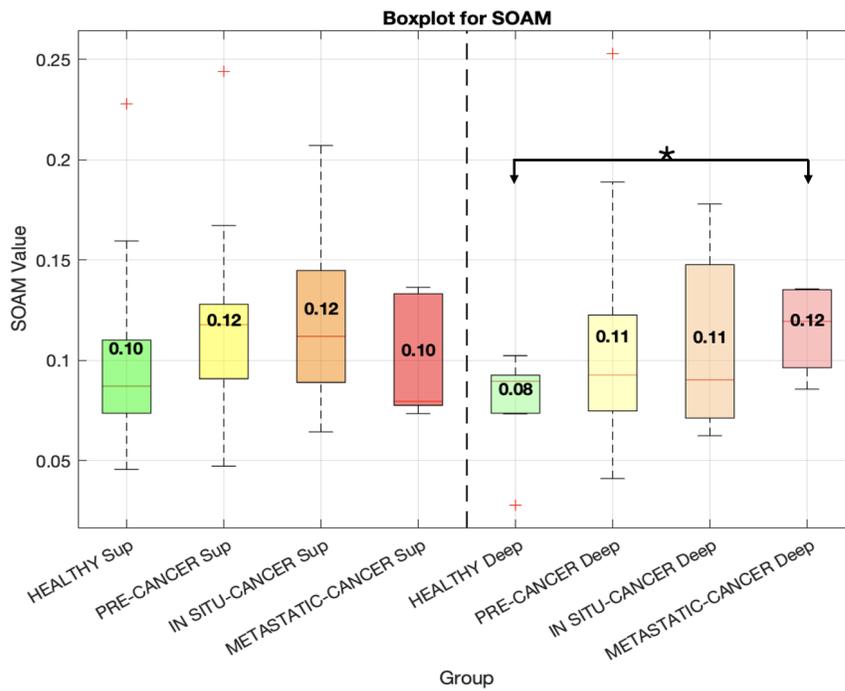


Figure 9.10: Boxplot Range of Sum of Angles Metric in the 4 groups for superficial and deep layers respectively.

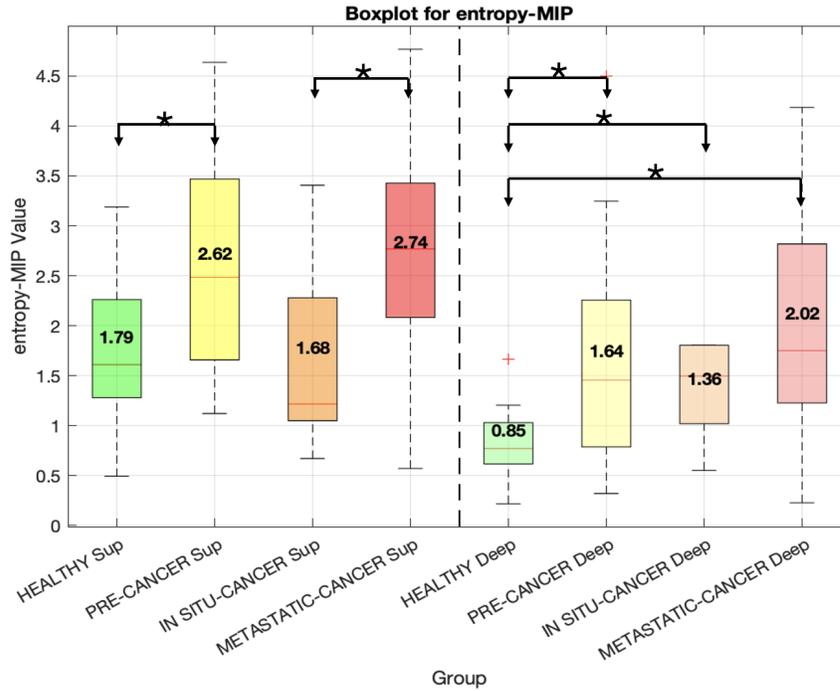


Figure 9.11: Boxplot Range of Entropy calculated on the MIP in the 4 groups for superficial and deep layers respectively.

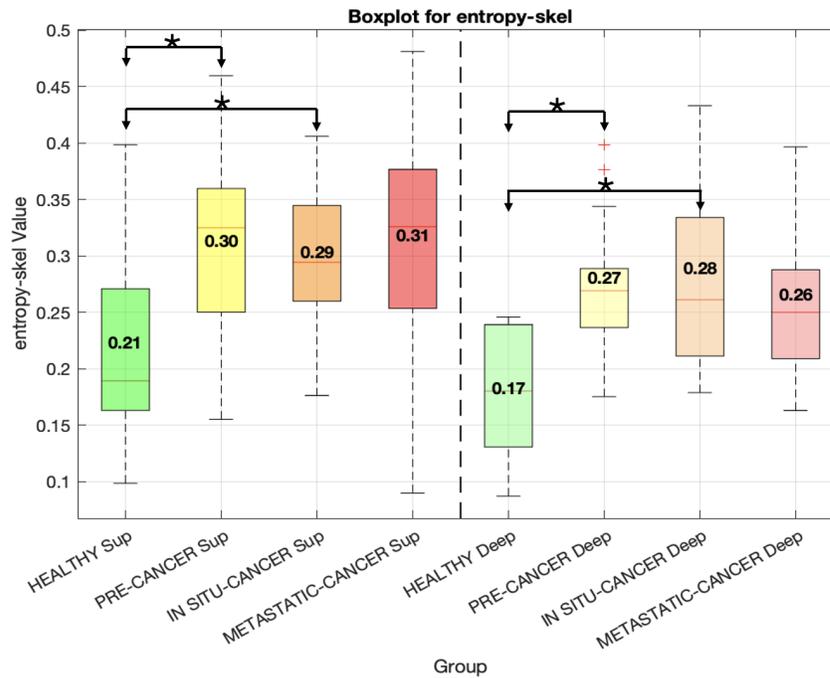


Figure 9.12: Boxplot Range of Entropy calculated on the skeleton in the 4 groups for superficial and deep layers respectively.

From the analysis of the boxplots, it can first be seen that, considering the Range, all calculated parameters result in at least one statistical difference on a group.

Ideally, we would expect that, as the vasculature is more organised and regular, the parameter ranges for less severe conditions such as healthy or pre-cancerous would be lower, compared to cancerous conditions in which, due to the chaotic network, the parameters should be more variable, and thus characterised by higher ranges. This trend is quite evident in the boxplots of VD-mask, NT, NB and ICM, respectively in Figure 9.4, Figure 9.6, Figure 9.7, Figure 9.9. However, for some parameters such as vascular density, NB, SOAM, especially in the deep layer, there is always a decrease in the metastatic-cancer condition, instead of, as would be expected, an increase. This could be related either to a characteristic of the tumour itself, but, perhaps more likely, to the low quality of the data, which may have adversely affected the calculation.

Finally, we note that the number of statistical differences on the surface layers is comparable to the differences on the deep layers. It must also be considered, however, that most of the statistical differences found allow discrimination between the healthy group and one of the other three conditions; on the other hand, as far as the differences between the three pathological groups are concerned, only a few parameters lead to this discrimination (NT superficial, Entropy-MIP - Figure 9.6, Figure 9.11), particularly on the surface layers.

9.1.3 Mean

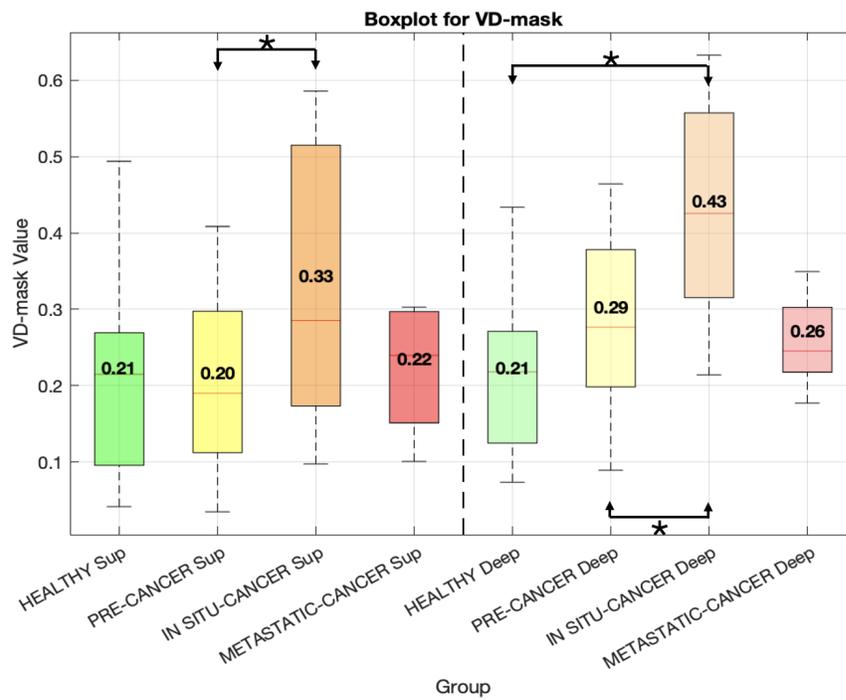


Figure 9.13: Boxplot of the Mean of the Vascular Density calculated on the segmentation mask in the 4 groups for superficial and deep layers respectively.

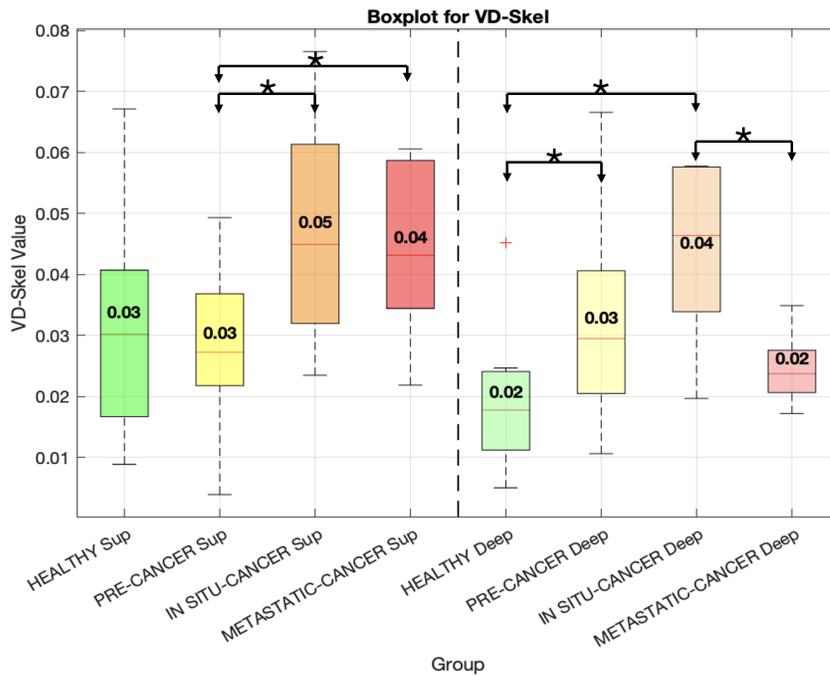


Figure 9.14: Boxplot of the Mean of the Vascular Density calculated on the skeleton in the 4 groups for superficial and deep layers respectively.

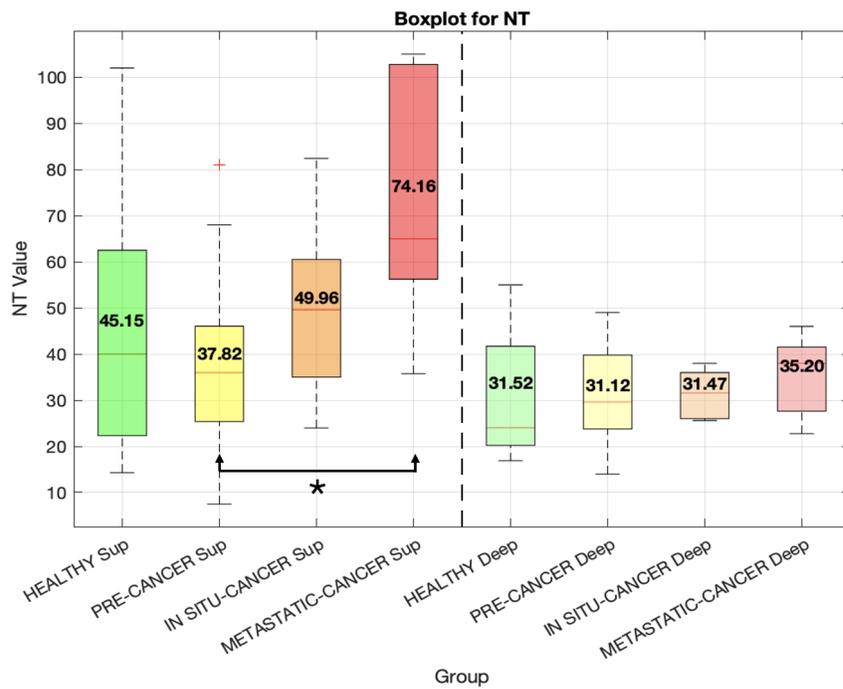


Figure 9.15: Boxplot of the Mean of the Number of Trees in the 4 groups for superficial and deep layers respectively.

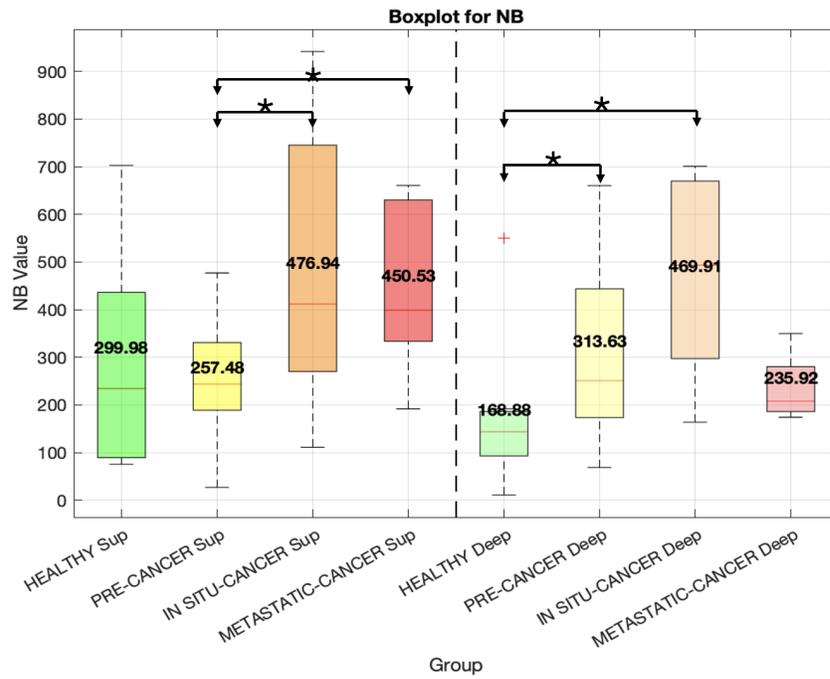


Figure 9.16: Boxplot of the Mean of the Number of Branches in the 4 groups for superficial and deep layers respectively.

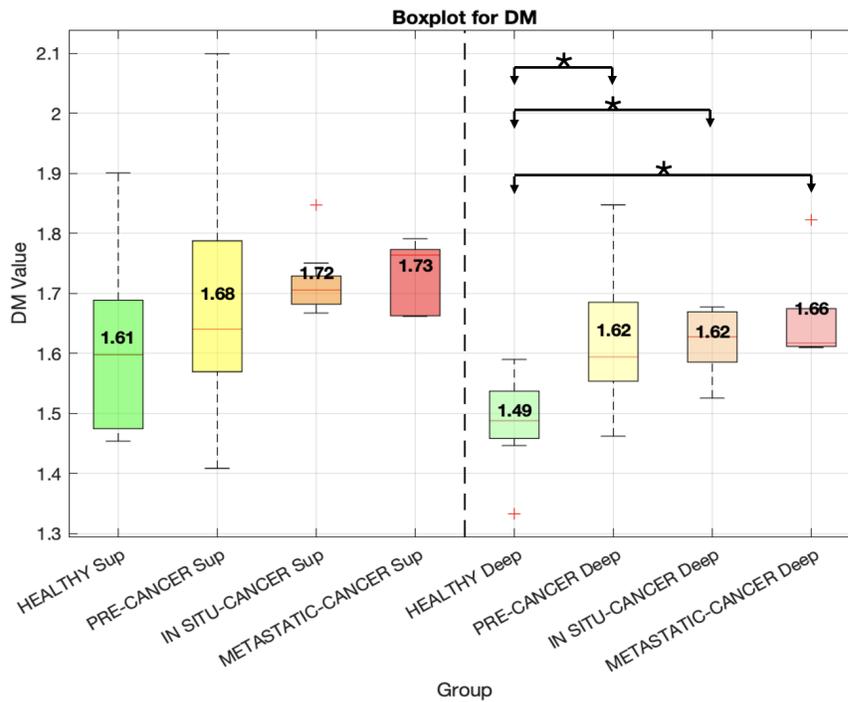


Figure 9.17: Boxplot of the Mean of the Distance Metric in the 4 groups for superficial and deep layers respectively.

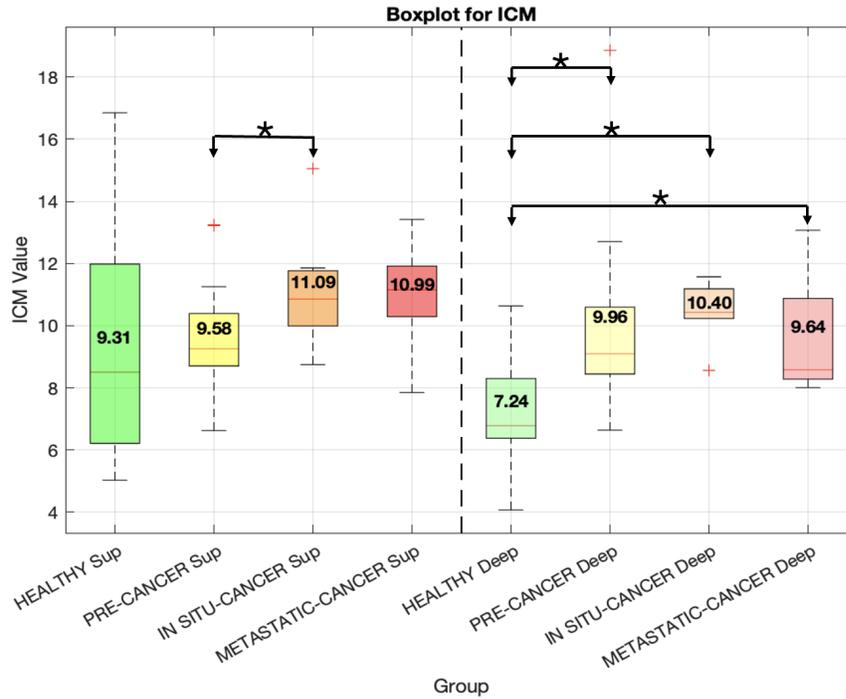


Figure 9.18: Boxplot of the Mean of the Inflection Count Metric in the 4 groups for superficial and deep layers respectively.

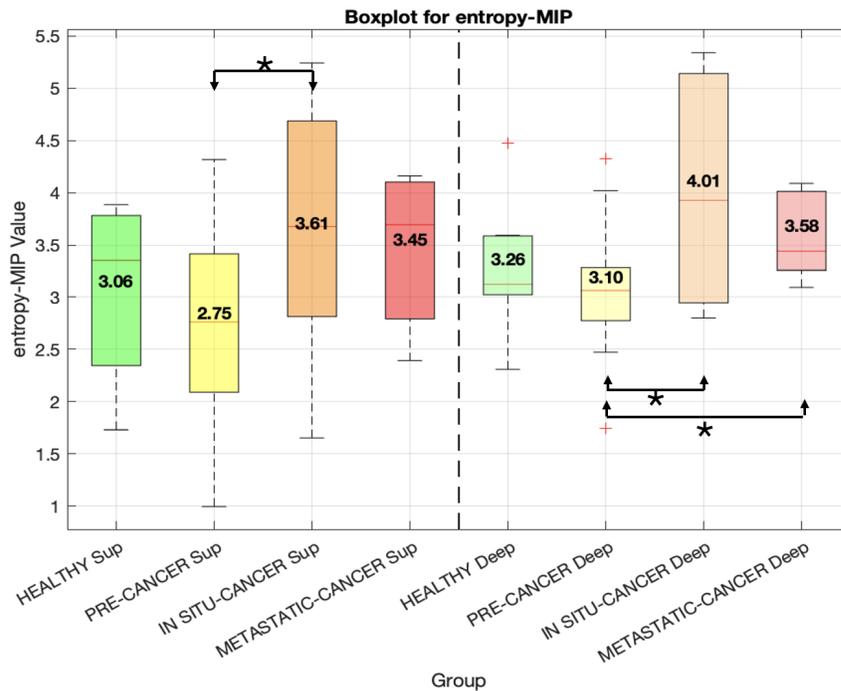


Figure 9.19: Boxplot of the Mean of the Entropy calculated on the MIP in the 4 groups for superficial and deep layers respectively.

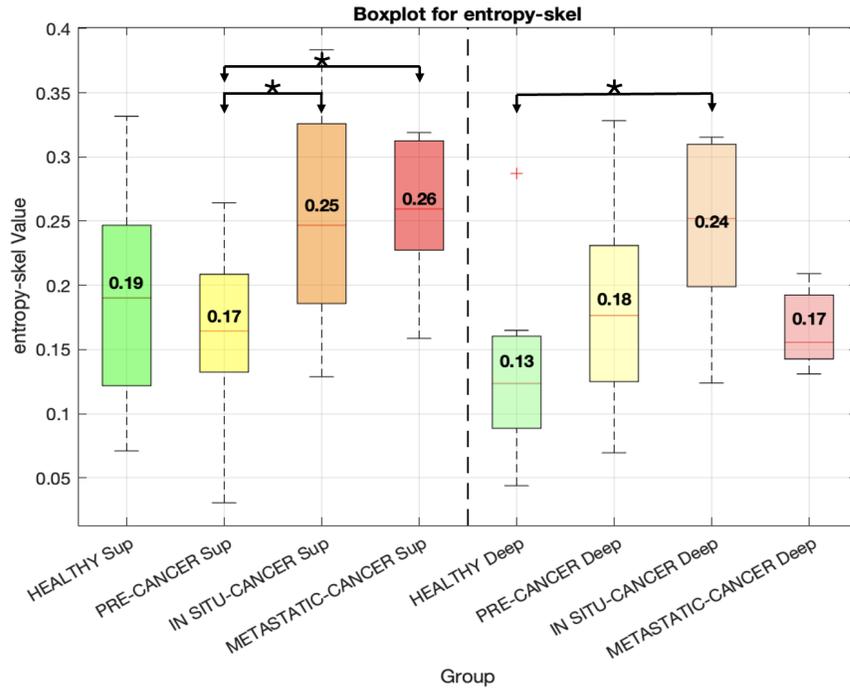


Figure 9.20: Boxplot of the Mean of the Entropy calculated on the skeleton in the 4 groups for superficial and deep layers respectively.

The analysis of the mean leads, unlike the previous case, to results on all calculated parameters except for SOAM. From the Boxplots, it can be seen that the analysis shows statistical differences especially on the deep layers: there are 18 pairs in the t-test success in the deep layers, against 11 pairs in the surface layers.

The expected trend is an increase in the average of the parameters passing from pre-cancerous conditions to more invasive ones: most of the parameters show this behaviour, with some exceptions such as the vascular density calculated on the skeleton (Figure 9.14) where on the deep layer the behaviour between the Metastatic /in situ cancer pair is opposite.

It should also be noted that this analysis is the one that leads to the greatest discrimination on the three pathological conditions, with parameters such as entropy, vascular density, NB and NT, as shown in Figure 9.19, Figure 9.14, Figure 9.15, Figure 9.16.

9.1.4 Max

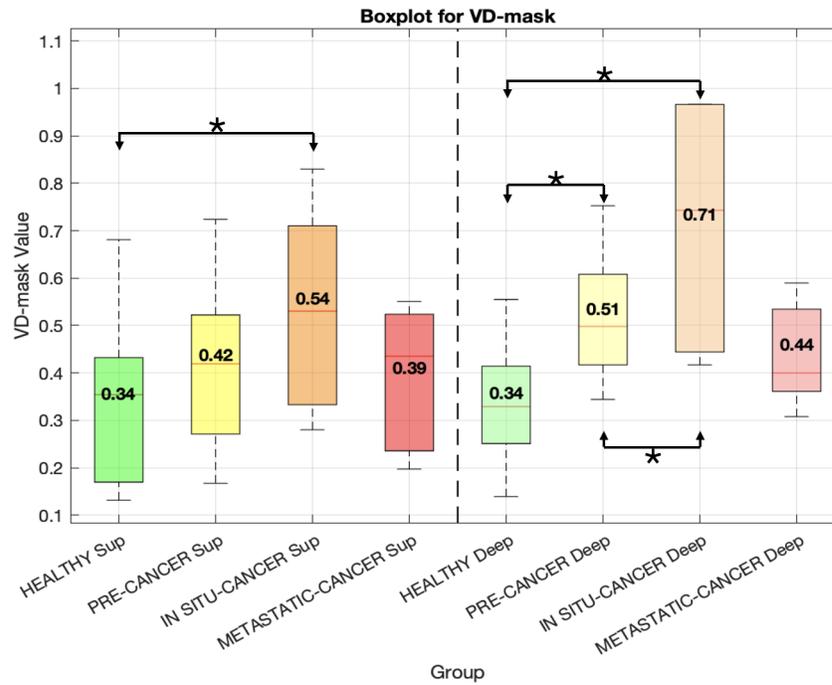


Figure 9.21: Boxplot of the Maximum values of the Vascular Density calculated on the segmentation mask in the 4 groups for superficial and deep layers respectively.

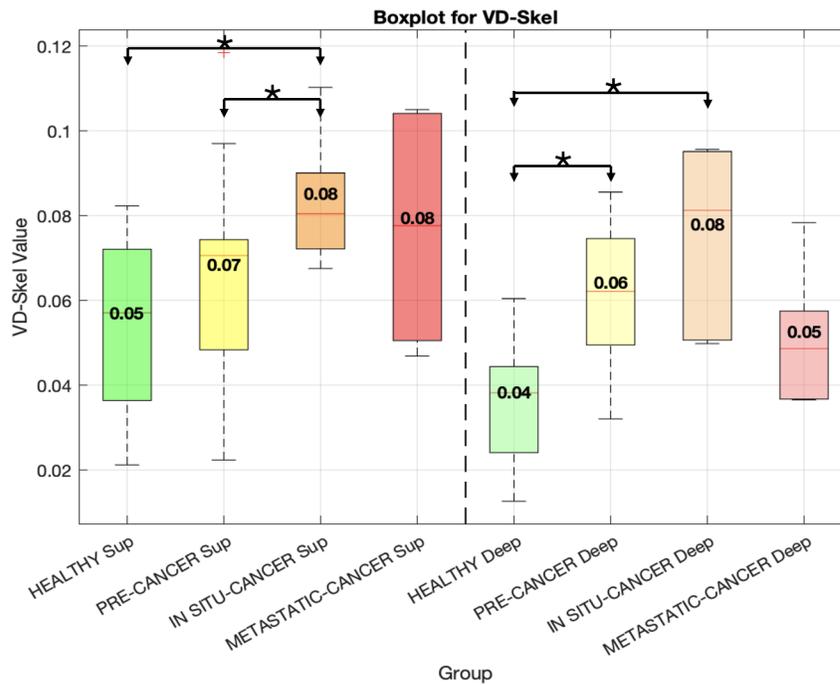


Figure 9.22: Boxplot of the Maximum values of the Vascular Density calculated on the skeleton in the 4 groups for superficial and deep layers respectively.

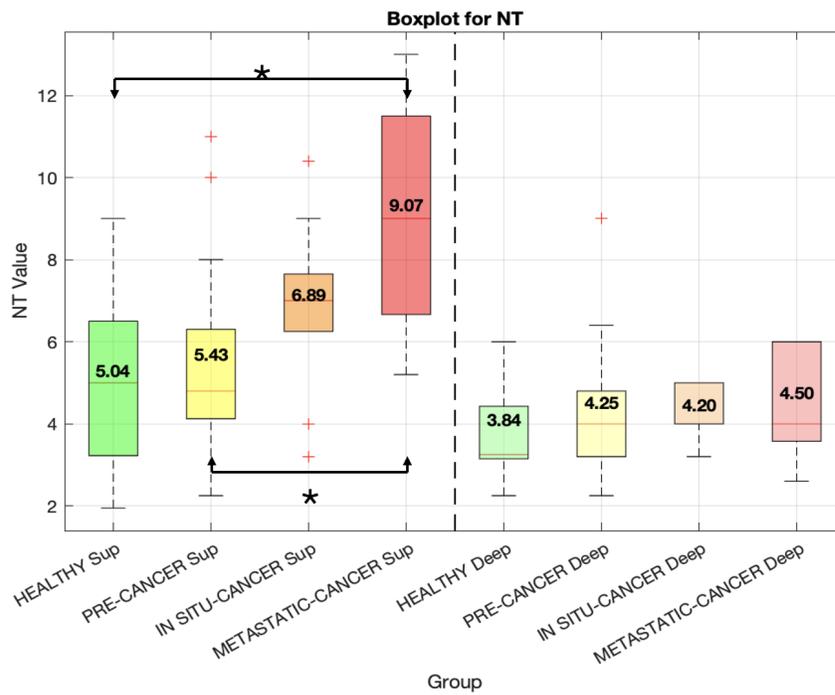


Figure 9.23: Boxplot of the Maximum values of the Number of Trees in the 4 groups for superficial and deep layers respectively.

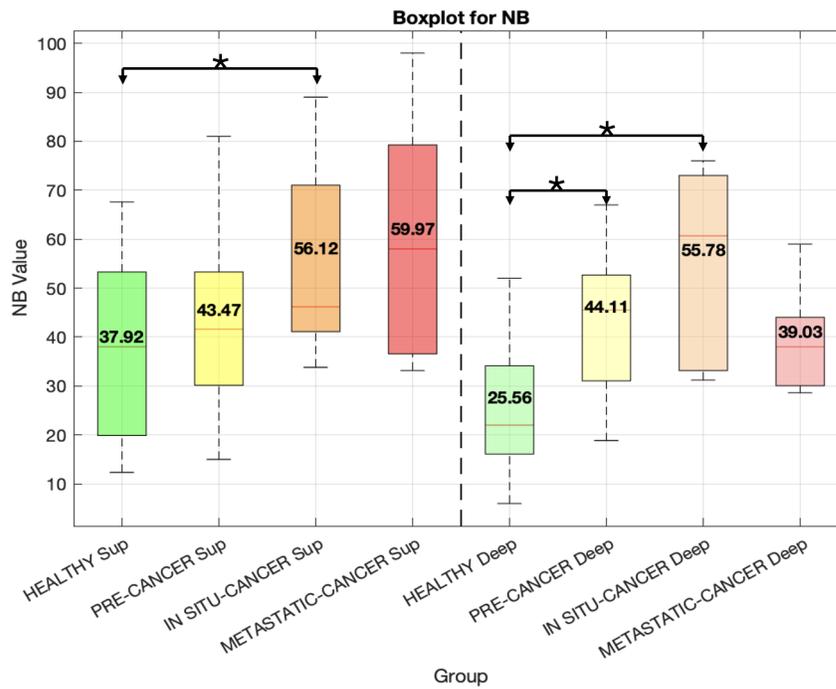


Figure 9.24: Boxplot of the Maximum values of the Number of Branches in the 4 groups for superficial and deep layers respectively.

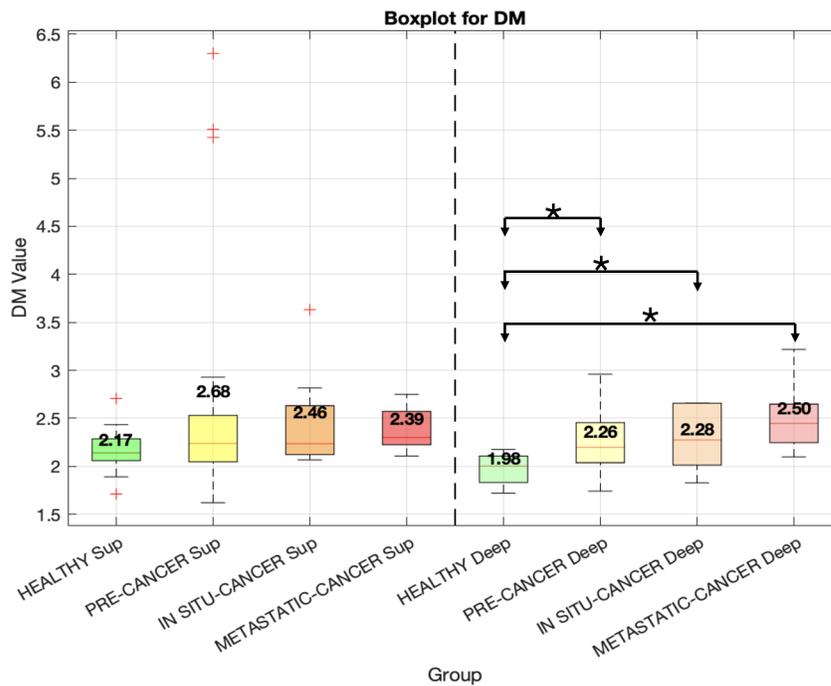


Figure 9.25: Boxplot of the Maximum values of the Distance metric in the 4 groups for superficial and deep layers respectively.

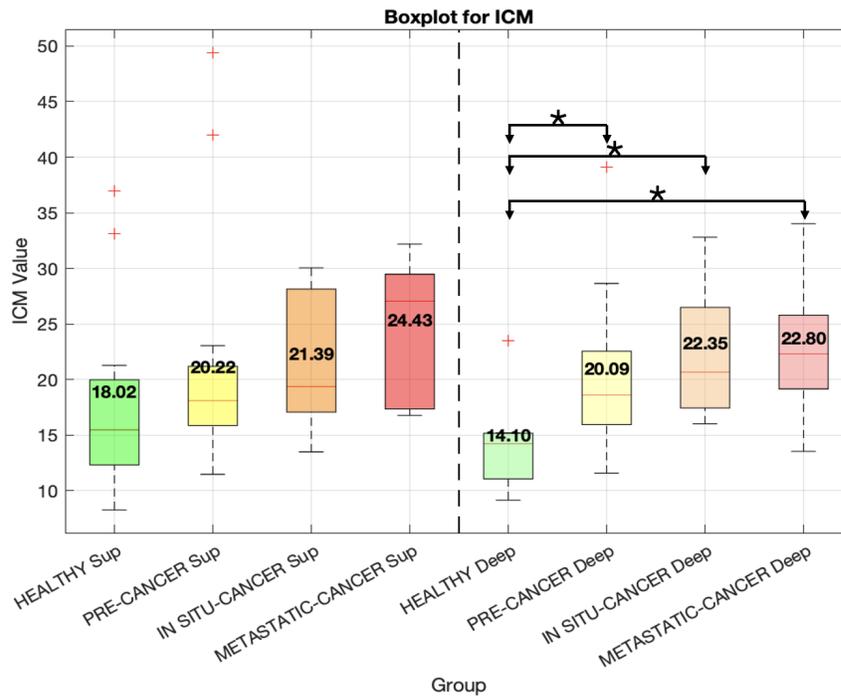


Figure 9.26: Boxplot of the Maximum values of the Inflection Count Metric in the 4 groups for superficial and deep layers respectively.

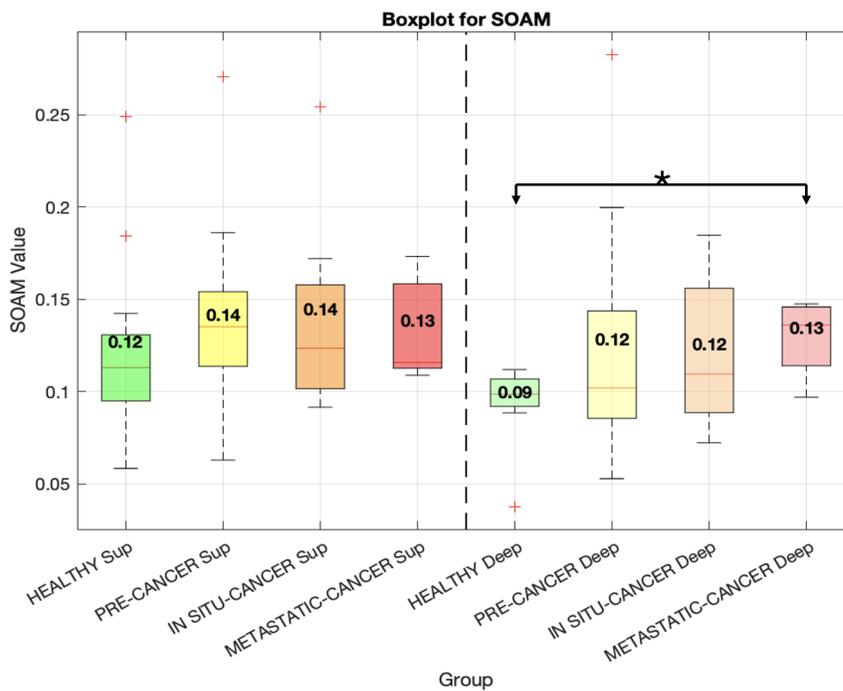


Figure 9.27: Boxplot of the Maximum values of the Sum of Angles Metric in the 4 groups for superficial and deep layers respectively.

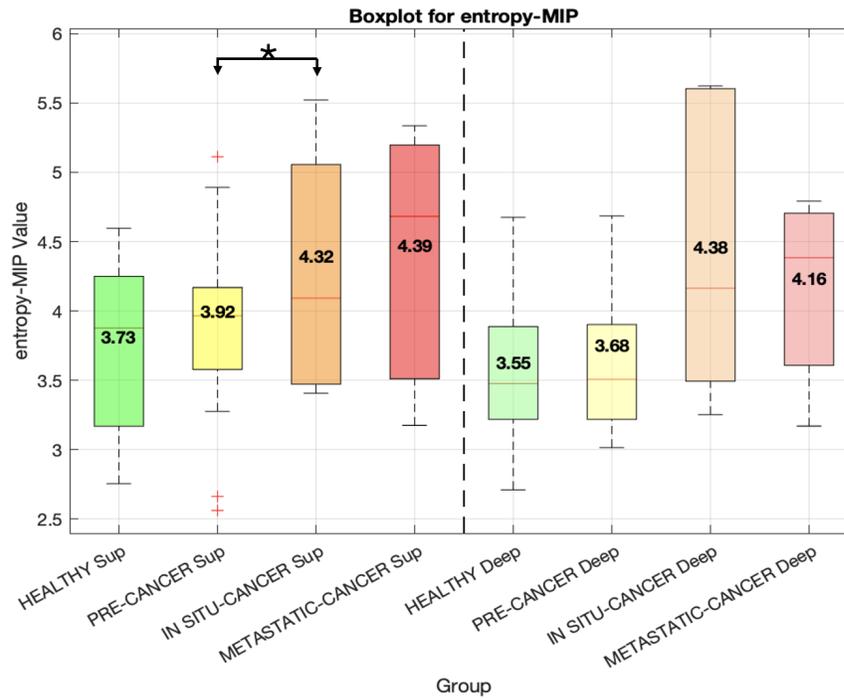


Figure 9.28: Boxplot of the Maximum values of the Entropy calculated on the MIP in the 4 groups for superficial and deep layers respectively.

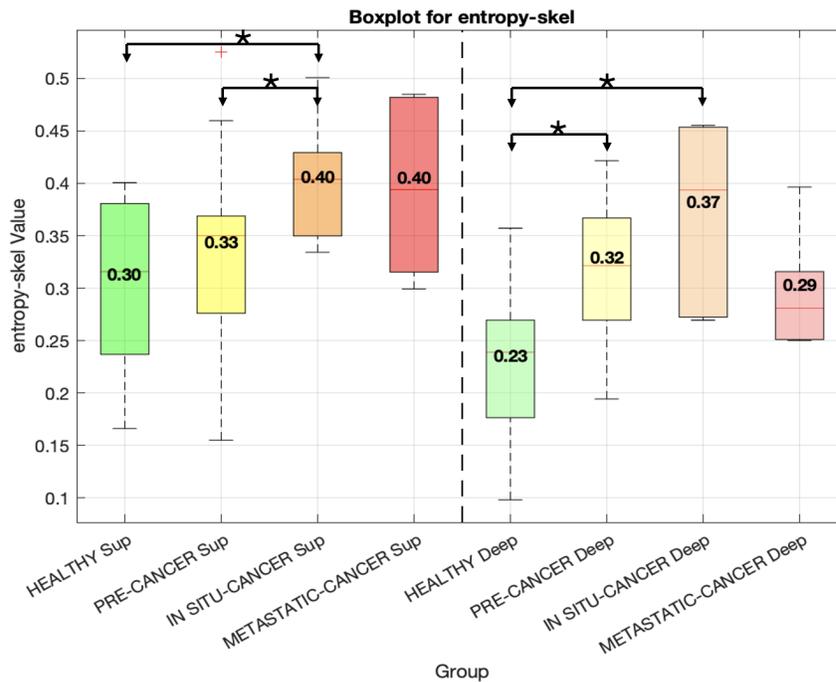


Figure 9.29: Boxplot of the Maximum values of the Entropy calculated on the skeleton in the 4 groups for superficial and deep layers respectively.

As with the Range, analysing the Maximum values also produces at least one statistical difference for each parameter.

From the boxplots, it can also be observed that almost twice as many results are shown on the Deep layers, compared to the superficial layers.

Talking about maximum values, the expected trend is always an increasing trend going towards high tumour conditions: this is a characteristic found in the results of this analysis, even if, for some parameters such as NB deep and VD-mask, reported in Figure 9.24 and Figure 9.21, the values associated with the metastatic tumour condition always tend to be lower.

It is also evident that most of the statistical differences discriminate between healthy and the other 3 conditions, with lower results in differentiating between the 3 disease groups.

9.2 MANOVA-test

The MANOVA-test described in *Chapter6*, usually outputs 3 possible linear combinations of the input parameters, where the respective p-value and Wilks Lambda are calculated on each. In this section, we analyse the results of this test, where for each test performed we report:

- **A table:** The MANOVA test in MATLAB usually provides 3 p-values and lambda as output, referring to three different linear parameter combinations. For the analysis conducted below, the results for only the first linear combination have been reported, as this combination is shown to be the best in maximising the difference between the classes analysed. Each test was conducted with and without interaction between the parameters; in the case of a significant result, an asterisk is shown in the table to highlight this significance.

- If the MANOVA-test is successful, leading to combinations with p-value < 0.05 , a **Bar-plot** is also reported. This bar-plot shows on the x-axis the parameters and any interactions, where for each of these a bar is represented relative to the canonical parameter load in each combination. On the y-axis is the value of the canonical load itself. This visualisation is useful to understand how much each parameter weighs in the linear combination that allowed discrimination between the groups (for us always the *Canonical1*, shown in Blue).

The analysis is conducted by commenting on the results obtained for Range, Mean Value and Maximum Parameter, respectively. The test was carried out for both deep and superficial data, considering only uncorrelated parameters. Only in the case of superficial parameters the analysis was repeated by selecting, among the uncorrelated parameters, a subset of parameters: looking again at the boxplots obtained above, it was decided to try the MANOVA test only on those parameters that visually presented an ideal trend for most of the groups, i.e. increasing values as the condition worsened. This was done with the idea of understanding whether, by excluding parameters with more random trends, and keeping those with an expected trend, the results could improve.

9.2.1 Range

	SUPERFICIAL LAYER		
	P-value		Wilks Lambda
Without interaction	0.0019	★	0.2470
With interaction	0.0171	★	0.1543

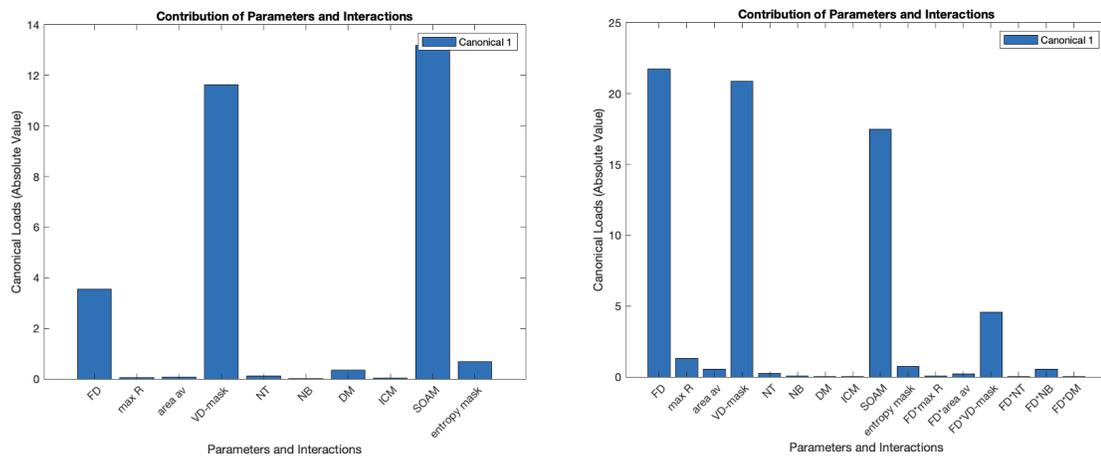


Figure 9.30: Results MANOVA test performed on uncorrelated parameter range values on the surface layer. Left: canonical loads of analysis without interaction. Right: canonical loads of the analysis with interaction.

The MANOVA test on the surface uncorrelated parameters (Figure 9.30) leads to discriminating the 4 conditions both considering the interaction between parameters and not. It can be seen from the table that the p-value, although associated with a higher Lambda value, tends to be significantly lower in the analysis without parameter interactions. Looking at the bar-plots of the canonical loads, it is evident that a large part of the contribution to the differentiation is related to the parameters *FD*, *VD-mask* and *SOAM*, with a small contribution from the combination *FD*VD-mask* in the case of the analysis with interaction.

	SUPERFICIAL LAYER		
	P-value		Wilks Lambda
Without interaction	0.0060	*	0.4767
With interaction	0.1716		0.3674

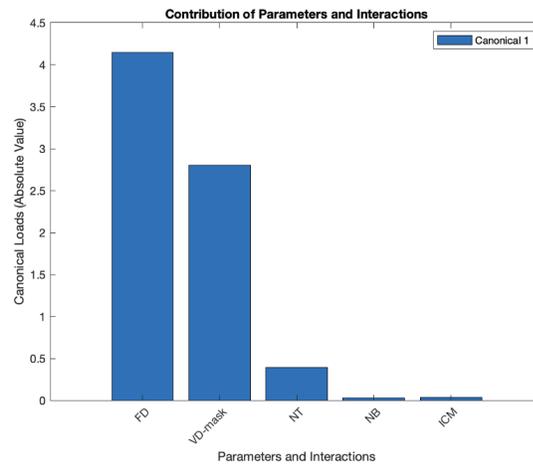


Figure 9.31: Results MANOVA test performed on the values for the range of uncorrelated parameters, with an increasing visual trend in the boxplots from healthy to metastatic tumour condition, on the superficial layer.

The analysis considering the subgroup of parameters with a specific trend in the Boxplots (Figure 9.31) allows us to make considerations on the role of *SOAM*: although this parameter does not show a great difference between the 4 groups analysed, in terms of range of values, from this analysis we note that not considering it we lose the statistical difference in the analysis with interaction, and have a greater p-value in the MANOVA-test without interaction. Discrimination between the 4 groups in this second case takes place thanks to a strong contribution of the Fractal Dimension. It is therefore interesting to note how the *SOAM* parameter, although it does not show a great value taken individually, combined with other parameters plays a fundamental role in discrimination.

	DEEP LAYER		
	P-value		Wilks Lambda
Without interaction	0.0070	*	0.4479
With interaction	0.0244	*	0.2189

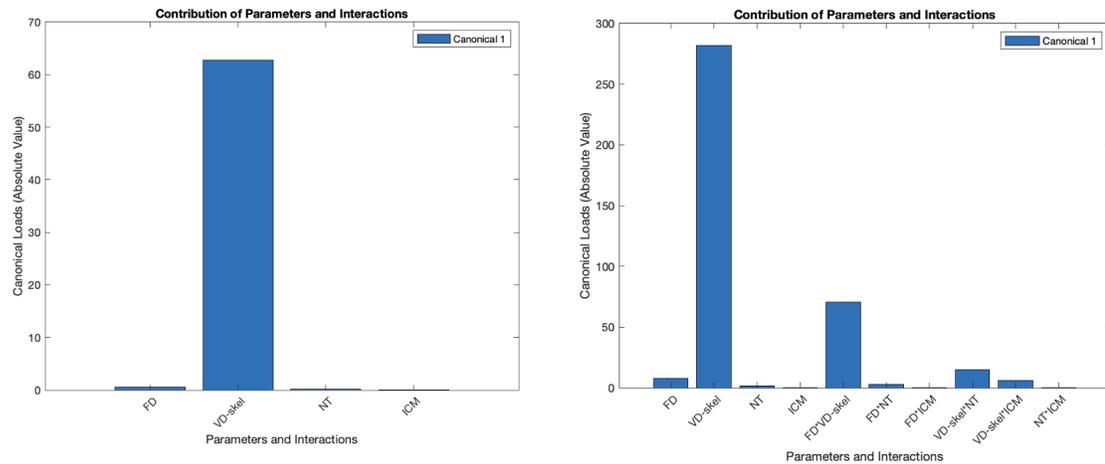


Figure 9.32: Results MANOVA test performed on uncorrelated parameter range values on the deep layer. Left: canonical loads of analysis without interaction. Right: canonical loads of the analysis with interaction.

Considering finally the parameter range on the deep layer, it can be seen that in both analyses the parameter *VD-skel* plays a predominant role in discrimination, even combined with other parameters in the analysis with interaction. However, if one looks at the boxplots in Figure 9.5, one can see that the trend of the parameter in the pathological conditions is not what one would expect with regard to metastatic cancer, as the value tends to decrease. The analysis in this case may therefore be biased due to the low quality of the analysed data.

9.2.2 Mean

	SUPERFICIAL LAYER		
	P-value		Wilks Lambda
Without interaction	0.0053	★	0.3792
With interaction	0.0395	★	0.2434

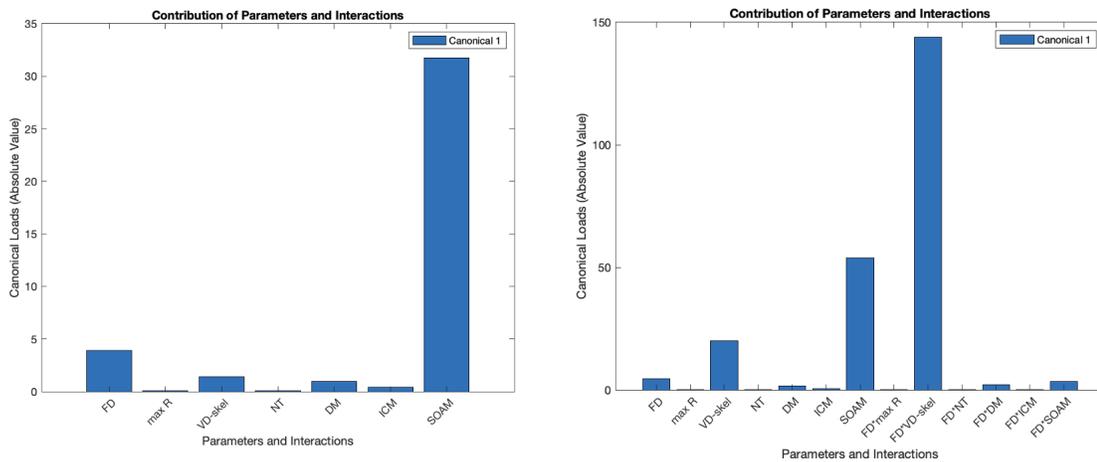


Figure 9.33: Results MANOVA test performed on uncorrelated parameter mean values on the surface layer. Left: canonical loads of analysis without interaction. Right: canonical loads of the analysis with interaction.

The analysis of the mean values of the surface uncorrelated parameters leads to significant results in both cases, but with a lower p-value for the analysis without interaction, where the main contribution to the linear combination that discriminates between the 4 groups is related to the parameter *SOAM*. Considering the interaction, it can be seen that *FD* and *VD-skel*, taken individually, do not make a large contribution; considering instead their combination, they weigh more in maximising the difference between the 4 classes.

	SUPERFICIAL LAYER		
	P-value		Wilks Lambda
Without interaction	0.0204	★	0.5797
With interaction	0.0280	★	0.3220

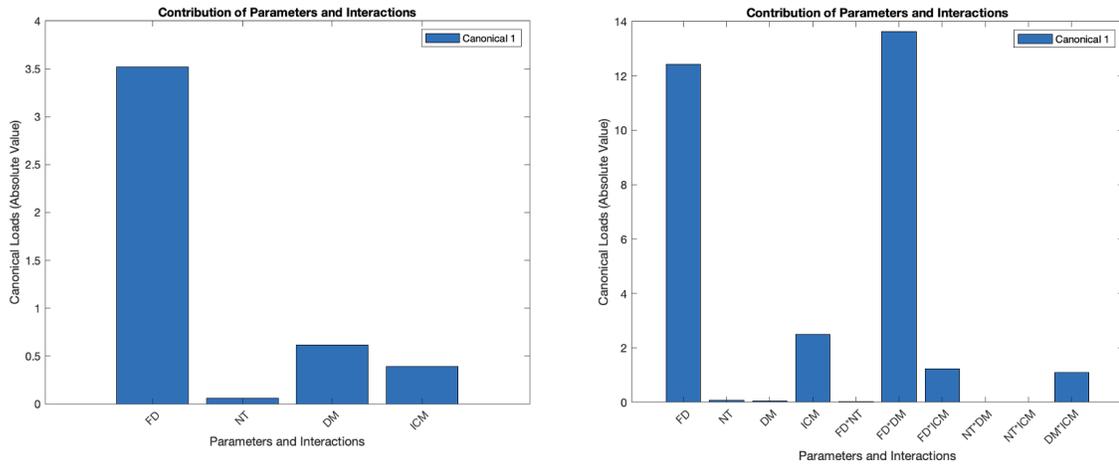


Figure 9.34: Results MANOVA test performed on the values for the mean of uncorrelated parameters, with an increasing visual trend in the boxplots from healthy to metastatic tumour condition, on the superficial layer. Left: canonical loads of analysis without interaction. Right: canonical loads of the analysis with interaction.

Selecting the subgroup of parameters with visual difference in the Boxplots, it can be seen that both analyses always lead to significant results. The p-value in the analysis without interaction goes down, with a strong contribution in the discrimination between the groups due to *Fractal Dimension*. In this case, the exclusion of *SOAM*, which weighed more strongly before, leads to an improvement in the results. The analysis without interaction results in a higher p-value than in the case presented above; however, the predominant role is still played by *Fractal Dimension* and its combinations with the other parameters.

	DEEP LAYER		
	P-value		Wilks Lambda
Without interaction	0.0129	★	0.3107
With interaction	0.0446	★	0.1540

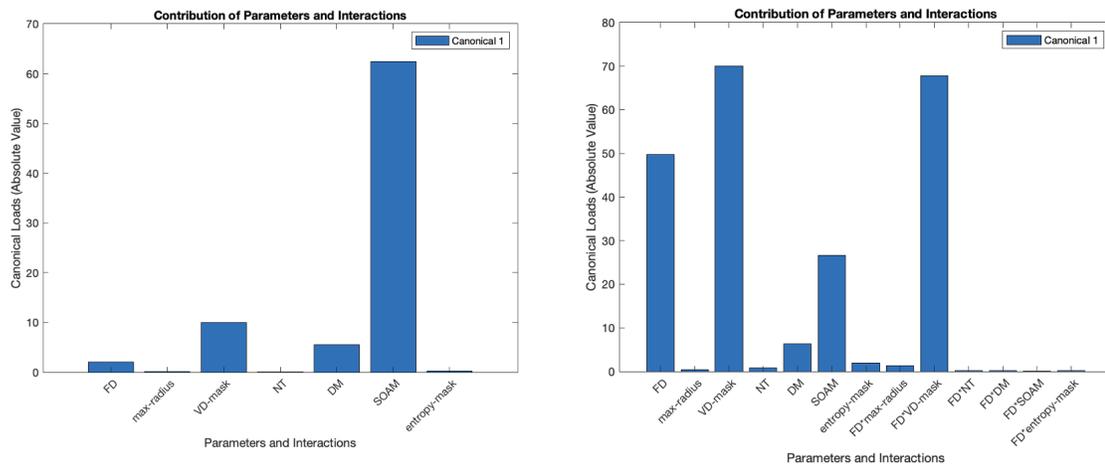


Figure 9.35: Results MANOVA test performed on uncorrelated parameter mean values on the deep layer. Left: canonical loads of analysis without interaction. Right: canonical loads of the analysis with interaction.

Even with deep data, the MANOVA-test is successful both considering parameter interactions and without. The greatest contribution in the combination of parameters discriminating the 4 groups, in the analysis without interaction, is due to *SOAM*; the second test shows how *FD*, *VD-mask*, *SOAM* and the combination *FD*VD-mask* play a decisive role in maximising the difference between the conditions analysed.

Again, it is interesting to point out that although the parameter *SOAM* did not lead to statistical differences in the T-test, combined with other parameters its role becomes crucial.

9.2.3 Max

	SUPERFICIAL LAYER		
	P-value		Wilks Lambda
Without interaction	0.0082	★	0.3933
With interaction	0.0815		0.2690

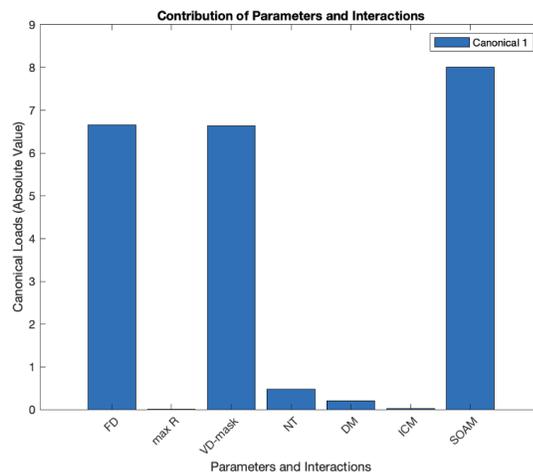


Figure 9.36: Results MANOVA test performed on uncorrelated parameter maximum values on the surface layer.

The result of the MANOVA test on the maximum parameters calculated on the surface layers shows that the product between parameters does not lead to a statistical difference between the 4 groups. In the non-interaction analysis, it can be seen that the main contribution to the combination is made by *SOAM* and *FD*, parameters that in the t-test (Figure 9.27, Figure 9.3), taken individually, showed no statistical difference.

	SUPERFICIAL LAYER		
	P-value		Wilks Lambda
Without interaction	0.0014	★	0.4832
With interaction	0.0035	★	0.2614

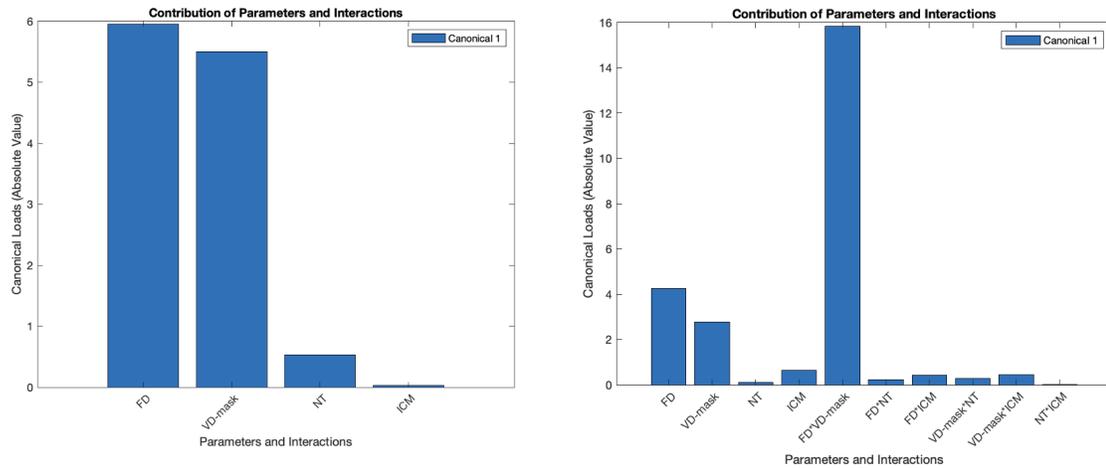


Figure 9.37: Results MANOVA test performed on the values for the maximum of uncorrelated parameters, with an increasing visual trend in the boxplots from healthy to metastatic tumour condition, on the superficial layer. Left: canonical loads of analysis without interaction. Right: canonical loads of the analysis with interaction.

Going to select the parameters not correlated with a specific trend in the boxplots, it can be seen from the figures above that both tests show a significant difference in discriminating between the groups, thus gaining one more result than in the previous case.

	DEEP LAYER	
	P-value	Wilks Lambda
Without interaction	0.0916	0.3971
With interaction	0.2100	0.2115

Figure 9.38: Results MANOVA test performed on uncorrelated parameter max values on the deep layer.

The MANOVA test on the deep layers showed no statistical differences, so no combination of input parameters can discriminate between the 4 conditions analysed.

9.3 Binary and Multiclass classifiers

As explained in *Chapter6*, the models trained on our surface and deep parameters, respectively, were used to generate two types of classifiers:

- A multiclass classifier: to discriminate the 4 groups analysed;
- A binary classifier: to differentiate the healthy/pathological groups.

To evaluate the validity of these classifiers, Confusion Matrixes were constructed, characterised by:

- Rows: where the predicted classes are given;
- Columns: where the actual classes are given;
- In orange: the false classified;
- In green: the correct classified.

Also wanting to quantify how well the implemented classifiers perform, the **Balanced Accuracy** was calculated:

$$\text{BA} = \frac{1}{k} \sum_{i=1}^k \text{TPR}_i \quad (9.1)$$

Where:

$$\text{TPR}_i = \frac{\text{TP}_i}{\text{TP}_i + \text{FN}_i} \quad (9.2)$$

is the class sensitivity i , calculated by considering the class i as positive and all others as negative.

We chose this metric, and not the standard *Accuracy*, because it takes into account the balance of classes, penalising the model if some classes are neglected. Since our DATASET is unbalanced, it was felt that this metric would lead to a more accurate result.

The analysis is always carried out by commenting on the results obtained on Range, Mean Value and Maximum Value of the parameters respectively.

9.3.1 Range

		SUPERFICIAL	REAL			
			METASTATIC CANCER	IN SITU CANCER	PRE-CANCER	HEALTHY
PREDICTED	METASTATIC CANCER	0	2	0	1	
	IN SITU CANCER	1	2	5	1	
	PRE-CANCER	3	4	15	4	
	HEALTHY	1	1	3	6	

		DEEP	REAL			
			METASTATIC CANCER	IN SITU CANCER	PRE-CANCER	HEALTHY
PREDICTED	METASTATIC CANCER	0	0	0	0	
	IN SITU CANCER	0	0	0	0	
	PRE-CANCER	4	5	16	6	
	HEALTHY	1	1	3	3	

Figure 9.39: Confusion Matrix representing the results of the multiclass classifier, taking parameter ranges as input. Above are the classifier outputs considering only the surface parameters; below is the classifier obtained using the deep parameters.

		SUPERFICIAL	REAL	
			PATHOLOGIC	HEALTHY
PREDICTED	PATHOLOGIC	6	0	
	HEALTHY	1	2	

		DEEP	REAL	
			PATHOLOGIC	HEALTHY
PREDICTED	PATHOLOGIC	6	0	
	HEALTHY	0	2	

Figure 9.40: Confusion Matrix representing the results of the binary classifier, taking parameter ranges as input. Right: output of the classifier considering only surface parameters; left: results of the classifier with only deep data as input.

From the figures above, one can immediately see that the *multiclassifier* works much better on the surface parameters, as in the deep parameters the two classes representing the tumour conditions are not recognised at all. This can certainly be justified by the limited number of data available for the conditions *cancer in situ* and *metastatic cancer*. At the same time, it should be noted that both classifiers give good results

on the pre-cancer and healthy conditions.

The *binary* classifier provides very good results on both layers.

9.3.2 Mean

		SUPERFICIAL	REAL			
			METASTATIC CANCER	IN SITU CANCER	PRE-CANCER	HEALTHY
PREDICTED	METASTATIC CANCER	1	1	2	1	
	IN SITU CANCER	2	0	1	2	
	PRE-CANCER	2	8	18	8	
	HEALTHY	0	0	2	1	

		DEEP	REAL			
			METASTATIC CANCER	IN SITU CANCER	PRE-CANCER	HEALTHY
PREDICTED	METASTATIC CANCER	0	0	0	1	
	IN SITU CANCER	0	3	0	1	
	PRE-CANCER	5	3	18	7	
	HEALTHY	0	0	1	0	

Figure 9.41: Confusion Matrix representing the results of the multiclass classifier, taking mean parameters as input. Above are the classifier outputs considering only the surface parameters; below is the classifier obtained using the deep parameters.

		SUPERFICIAL	REAL	
			PATHOLOGIC	HEALTHY
PREDICTED	PATHOLOGIC	5	2	
	HEALTHY	2	0	

		DEEP	REAL	
			PATHOLOGIC	HEALTHY
PREDICTED	PATHOLOGIC	5	1	
	HEALTHY	1	1	

Figure 9.42: Confusion Matrix representing the results of the binary classifier, taking mean parameters as input. Right: output of the classifier considering only surface parameters; left: results of the classifier with only deep data as input.

Even when taking the average of the parameters, it can be seen how the *multiclassifiers* on both layers have difficulty in classifying tumour classes. In this case, however,

the ability to differentiate the healthy group well from the other groups in the study is also lost: indeed, it can be observed from the CMs above how very often healthy acquisitions are classified as pre-cancer. On the other hand, the classifier’s ability to discriminate *pre-cancer* acquisitions is always maintained. The results with the *binary* classifier worsen; looking at the confusion matrix above, it can be seen that the ability to correctly classify the healthy class is mostly lost.

9.3.3 Max

PREDICTED	SUPERFICIAL	REAL			
		METASTATIC CANCER	IN SITU CANCER	PRE-CANCER	HEALTHY
METASTATIC CANCER		0	1	2	0
IN SITU CANCER		2	1	2	1
PRE-CANCER		3	7	18	10
HEALTHY		0	0	1	1

PREDICTED	DEEP	REAL			
		METASTATIC CANCER	IN SITU CANCER	PRE-CANCER	HEALTHY
METASTATIC CANCER		0	0	0	0
IN SITU CANCER		1	3	1	1
PRE-CANCER		3	2	15	4
HEALTHY		1	1	3	4

Figure 9.43: Confusion Matrix representing the results of the multiclass classifier, taking maximum parameters as input. Above are the classifier outputs considering only the surface parameters; below is the classifier obtained using the deep parameters.

PREDICTED	SUPERFICIAL	REAL	
		PATHOLOGIC	HEALTHY
PATHOLOGIC		6	1
HEALTHY		1	1

PREDICTED	DEEP	REAL	
		PATHOLOGIC	HEALTHY
PATHOLOGIC		4	1
HEALTHY		2	1

Figure 9.44: Confusion Matrix representing the results of the binary classifier, taking maximum parameters as input. Right: output of the classifier considering only surface parameters; left: results of the classifier with only deep data as input.

Even for maximum parameters, the *multiclassifier* cannot distinguish the metastatic-cancer group. However, on both layers, good results are obtained for the healthy and pre-cancer group, and a discrete ability, more visible in the deep layer, in classifying the in situ-cancer group.

The results of the *binary classifier* are always good, especially on the superficial layer.

Table 9.1: *Balanced Accuracy for each built classifier*

		BALANCED ACCURACY	
		MULTICLASS CLASSIFIER	BINARY CLASSIFIER
SUPERFICIAL LAYER	RANGE	34,36%	92,85%
	MEAN	21,65%	35,71%
	MAX	24,43%	67,85%
DEEP LAYER	RANGE	30,56%	100%
	MEAN	36,18%	66,67%
	MAX	43,35%	66,67%

The Balanced Accuracy calculated on each classifier is summarised in Table 9.1. It can be seen that on the whole, as expected, the best results are obtained on the *Binary Classifier*, since, by combining the acquisitions, two classes with a larger number of data are obtained. It must also be considered that the binary classifiers were validated on a very small test set, making classification easier. The *Multiclassifier*, on the other hand, is built on 4 very unbalanced groups, where the *healthy* and *pre-cancer* classes are clearly more represented. The multiclass classifier built on the parameter range shows the best classification ability on the surface data; for the deep data, however, it is by exploiting the information of the maximum of the parameters that fewer errors are made.

9.4 Depth of lesions

In the Table 9.2, the depth in pixels is reported for each analysed lesion. The table was constructed by highlighting the four conditions under analysis with different colours, to facilitate visualisation. Where the result is not reported, it is because lesion detection in the B-scans of the OCT volume was very complex and insecure. Therefore, to avoid affecting the overall results, it was decided not to include these depths.

Interestingly, the table shows a clear difference between the depths each lesion reaches according to its group. In fact, analysing the results, it is possible to identify ranges:

- depth < 100 pixel: lesion of the group *healthy*;
- $100 < \textit{depth} < 120$ pixel: lesion belonging to the group *pre-cancer*;
- $120 < \textit{depth} < 140$ pixel: lesion of the group *in situ-cancer*;
- depth > 140 pixel: *metastatic-cancer* lesion.

Table 9.2: Lesion depth in pixels

	LESION DEPTH	GROUP
PZ001	181	METASTATIC CANCER
PZ002	102	PRE-CANCER
PZ003	117	PRE-CANCER
PZ004-lesion1	125	IN SITU-CANCER
PZ004-lesion2	100	PRE-CANCER
PZ005	102	PRE-CANCER
PZ006	129	IN SITU-CANCER
PZ007	100	PRE-CANCER
PZ008	134	IN SITU-CANCER
PZ009	/	PRE-CANCER
PZ010	140	METASTATIC CANCER
PZ011	90	HEALTHY
PZ012-lesion1	/	IN SITU-CANCER
PZ012-lesion2	/	PRE-CANCER
PZ013	95	HEALTHY
PZ014	130	IN SITU-CANCER

9.5 Combined Analysis

Given the promising results obtained by extracting the lesion depth, which allows a clear distinction between the groups, it was decided to consider this lesion characteristic as a new parameter. To see if this could contribute, along with the other parameters, to differentiating the 4 study groups, a PCA (Principal Component Analysis) was conducted pre- and post-inclusion of lesion-depth. PCA is a statistical technique that reduces the dimensionality of a dataset by transforming the original variables into new variables, called principal components, ordered by variance. The visualisation of principal components makes it possible to understand whether there is a difference between the groups, or a partial/total overlap between them. The PCA was repeated:

- for superficial and deep layers;
- for range, mean and maximum parameters;
- considering the 4 groups metastatic cancer/ in situ cancer/ pre-cancer/healthy.

However, the analysis on the different parameter groups and the two layers produced very similar results, which is why only one of the results is shown in Figure 9.45.

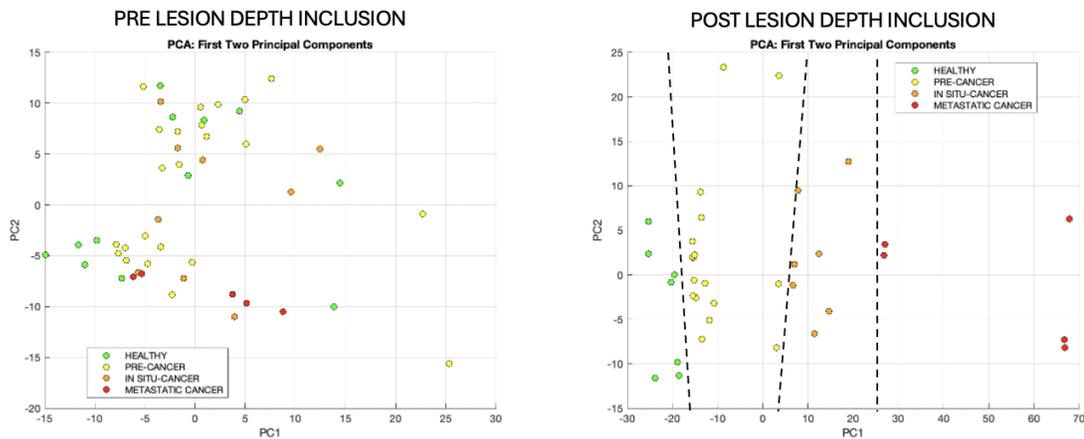


Figure 9.45: Results PCA analysis conducted: left - before insertion of lesion depth as a new parameter; right - after insertion of the new parameter. The dotted lines show how there is a difference in the distribution in space between the 4 groups.

The PCA highlights the strong discriminating power that the introduction of lesion depth as a parameter has on the 4 groups under analysis.

For this reason, having obtained with the multiclass classifiers a not too good performance in differentiating the groups, the analysis was repeated adding both for the superficial and deep layers the lesion depth, in the hope that such a significant parameter in discriminating between the groups could improve the results.

Below are the results for Range, Mean and Maximum value of the parameters for surface and deep layer, respectively.

9.5.1 Range

		SUPERFICIAL	REAL			
			METASTATIC CANCER	IN SITU CANCER	PRE-CANCER	HEALTHY
PREDICTED	METASTATIC CANCER		3	0	0	0
	IN SITU CANCER		2	7	3	0
	PRE-CANCER		0	2	20	5
	HEALTHY		0	0	0	7

		DEEP	REAL			
			METASTATIC CANCER	IN SITU CANCER	PRE-CANCER	HEALTHY
PREDICTED	METASTATIC CANCER		5	0	0	0
	IN SITU CANCER		0	4	2	0
	PRE-CANCER		0	2	17	7
	HEALTHY		0	0	0	2

Figure 9.46: Confusion Matrix representing the results of the multiclass classifier, taking range parameters as input and the depth lesion. Above are the classifier outputs considering only the surface parameters; below is the classifier obtained using the deep parameters.

9.5.2 Mean

		SUPERFICIAL	REAL			
			METASTATIC CANCER	IN SITU CANCER	PRE-CANCER	HEALTHY
PREDICTED	METASTATIC CANCER		3	0	0	0
	IN SITU CANCER		2	7	0	0
	PRE-CANCER		0	2	23	5
	HEALTHY		0	0	0	7

		DEEP	REAL			
			METASTATIC CANCER	IN SITU CANCER	PRE-CANCER	HEALTHY
PREDICTED	METASTATIC CANCER		5	1	0	0
	IN SITU CANCER		0	4	0	0
	PRE-CANCER		0	1	19	7
	HEALTHY		0	0	0	2

Figure 9.47: Confusion Matrix representing the results of the multiclass classifier, taking the mean of the parameters as input and the depth lesion. Above are the classifier outputs considering only the surface parameters; below is the classifier obtained using the deep parameters.

9.5.3 Max

		SUPERFICIAL	REAL			
			METASTATIC CANCER	IN SITU CANCER	PRE-CANCER	HEALTHY
PREDICTED	METASTATIC CANCER	3	0	0	0	
	IN SITU CANCER	2	7	2	0	
	PRE-CANCER	0	2	21	5	
	HEALTHY	0	0	0	7	

		DEEP	REAL			
			METASTATIC CANCER	IN SITU CANCER	PRE-CANCER	HEALTHY
PREDICTED	METASTATIC CANCER	5	1	0	0	
	IN SITU CANCER	0	4	3	0	
	PRE-CANCER	0	1	16	4	
	HEALTHY	0	0	0	5	

Figure 9.48: Confusion Matrix representing the results of the multiclass classifier, taking maximum parameters as input and the depth lesion. Above are the classifier outputs considering only the surface parameters; below is the classifier obtained using the deep parameters.

From the confusion matrix above, it can be seen that, for each group of parameters, the results of the classifiers improve significantly compared to the previous case.

Not only is superior performance obtained in distinguishing the healthy and pre-cancer groups, but, more importantly, **the classifier acquires the ability to separate the two cancerous groups**. In particular, these two groups are classified correctly in most cases on both layers; when errors occur, the classifier still tends to confuse the classes with a similar condition and not a completely different one.

For all groups, overall good performance is obtained; only in the *healthy* group, although most of the data are correctly classified, there is a greater tendency to confuse them with the pre-cancer group, particularly in the surface layer.

Table 9.3: *Balanced Accuracy for new classifiers*

BALANCED ACCURACY		
MULTICLASS CLASSIFIER		
SUPERFICIAL LAYER	RANGE	70,76%
	MEAN	74%
	MAX	71,85%
DEEP LAYER	RANGE	69,59%
	MEAN	73,80%
	MAX	76,61%

In Table 10.1 the Balanced-Accuracy calculated for each classifier is summarised. It is highlighted that the results are clearly better, with performance about twice compared to the previous case. Comparable behaviour of the classifiers on the superficial and deep layers is also observed in the three different analyses. The best result is obtained for the classifier calculated on the maximum deep parameters, but with a minimal difference in performance compared to the others.

PART IV

CONCLUSION

Chapter 10

CONCLUSIONS, LIMITATIONS AND FUTURE DEVELOPMENTS

The aim of the thesis project presented was to verify whether OCT and OCTA, interference-based imaging techniques that are already widely used clinically in the ophthalmological field, could also have potential in dermatology, specifically in the diagnosis of actinic keratosis and its progressions, overcoming the limitations of the standard techniques used today.

To this end, both a pathway focused on the characterisation of the vasculature was explored, with the aim of exploiting the differences in terms of vascular architecture that the process of angiogenesis induces in tumours; and a more morphological analysis, focused on the lesion itself and on how much it invades the deeper layers according to its invasiveness.

From the results presented in *Chapter 9*, we can therefore **conclude** that:

- The extraction of distinct parameters by layers makes it possible to effectively distinguish a healthy condition from any pathological condition analysed (*pre-cancer, in situ-cancer, metastatic-cancer*). This difference is evident both in the parameters on the surface layers and, even more markedly, in the parameters calculated on the deep layers, for which the t-test showed a greater number of significant results. These results were observed on *tortuosity parameters*, such as *DM* and *ICM*, parameters related to the *structure of the vascular network*, such as *NT* and *NB*, and especially parameters describing the *complexity of the vascular network*, such as *entropy, vascular density* and *FD*.

- With regard to the discrimination between the three pathological conditions (*pre-cancer*, *in situ-cancer*, *metastatic-cancer*), the parametric analysis revealed fewer discriminating parameters than in the previous case. However, for each pair of pathological conditions (*pre-cancer/in situ-cancer*, *pre-cancer/metastatic-cancer*, *in situ-cancer/metastatic-cancer*), it was always possible to identify at least one parameter able to differentiate the classes. The most significant results were obtained from the t-tests conducted on the mean values of the parameters, compared to analyses conducted on range or maximum value, showing differences in terms of *vascular density*, *Entropy*, *Avascular Area* and *NT*. Contrary to the hypothesis, however, this discriminating ability was manifested mainly in the superficial layers. Indeed, it was expected that, as tumour forms with different degrees of invasiveness, the differences would be more evident in the deep layers.
- In order to look for a difference on the deep layers, the MANOVA test was implemented, a technique that uses the combination of parameters to differentiate between the groups of interest. For completeness, the analysis was then also repeated on the surface parameters. The tests showed that the uncorrelated mean parameters, considered overall and not individually, respectively with and without interaction, resulted in a statistical difference between the 4 conditions studied. The greatest contribution was always made by parameters describing network complexity, i.e. *FD*, *VD-mask*, and tortuosity parameters, i.e. *SOAM*. Good results were also obtained on the surface layer, both considering all uncorrelated parameters, and subgroups of parameters that showed an increasing trend in the boxplots, going from less severe (*pre-cancer*), to invasive tumour conditions (*metastatic-cancer*). The MANOVA test played a key role in the study, demonstrating how parameters that individually did not lead to

statistical evidence, when combined with others, played a crucial role in differentiation.

- At this point, an investigation was conducted to verify whether the combination of parameters would produce a result in predicting the analysed condition: a binary classifier, trained on a training set and validated on a test set, was implemented to discriminate between healthy-pathological; a multiclass classifier, validated with the k-fold cross validation method, to differentiate between healthy- pre-cancer-cancer in situ-metastatic cancer. The aim was to make global use of the input vascular parameters to associate each acquisition with a specific condition. Although conscious of the intrinsic limitations of the dataset used, characterised by a significant imbalance between classes and a limited overall numerosity, the classifiers nevertheless showed discrete results. A good performance of the binary classifier trained on the parameter intervals emerged; inferior results were obtained on the multiclassifier, where, although a discrete ability to classify the healthy and pre-cancer classes was observed, the other two classes associated with a cancer condition were practically not recognised. The balanced accuracy, used as a validation metric, showed an average performance of 30% for the multiclassifier and 70% for the binary classifier.
- The most significant analysis emerged from the OCT study: using these volumes and based on the variations in intensity, it was possible to identify the lesion and determine its depth. The sum of the depths of each lesion was the key to discriminating between the four conditions in the study. It was found that when moving from a healthy to a metastatic condition, not only the depth of the lesion increased progressively, but it was also possible to identify specific depth ranges characteristic of each group.

- In view of the significant results of lesion depth in discriminating between the 4 study groups, a PCA analysis was performed to see whether, considering this tumour characteristic as a new parameter, combining it with the other study parameters would increase differentiation between the groups. The results of this analysis showed that the addition of this parameter resulted in a clearly visible spatial separation between the 4 classes.

Therefore, given the strong potential of the depth lesion in what is the purpose of the study, the multiclassifier analysis was repeated with this new parameter, obtaining much better results, with a strong ability to discriminate between the 4 groups, including metastatic-cancer and in situ-cancer that were not recognised by the previous classifier. Thus, on both the superficial and deep layers, average Balanced Accuracy of 70% was obtained.

Table 10.1: Summary conclusions

	KEY FINDINGS
Distinction between Healthy and Pathological Conditions	Clear separation achieved using parameters such as tortuosity (DM, ICM), vascular network structure (NT, NB), and complexity (Entropy, VD, FD). Better results observed in deep layers , with significant t-test results.
Discrimination among the Pathological Conditions	Fewer discriminating parameters, but at least one parameter differentiates each pair of pathological conditions. Significant parameters: VD, Entropy, NT especially in surface layers .
Multivariate Analysis (MANOVA)	Combining uncorrelated parameters showed significant differences between groups both surface and deep layers. Key parameters: FD, VD-mask, SOAM .
Binary and Multiclass Classification	Binary classifier (healthy-pathological): 70% balanced accuracy . Multiclassifier (4 groups): poorer performance (30% balanced accuracy), with <i>difficulty distinguishing cancer conditions</i> .
Lesion Depth Analysis	Lesion Depth strongly correlates with tumor severity, showing progressive increase from healthy to metastatic conditions. Depth ranges specific to each group identified.
PCA with Lesion Depth Parameter	PCA with lesion depth added to other parameters showed clear spatial separation between the groups.
Impact on multiclassification with Depth Lesion	Including lesion depth improved multiclass classification results: 70% balanced accuracy for 4-group discrimination, with significant <i>improvement in distinguish metastatic and in situ cancer</i> .

The project did, however, present a number of **limitations**, mainly related to:

- To the *Data Quality*: many data were noisy and characterised by artefacts. This certainly affected the quality of the analysis, especially in terms of parameter extraction. In fact, it was seen that some parameters, such as Maximum Radius and % Avascular Area, exhibited opposite or random behaviour to that expected, most likely linked to a low quality of the data that influenced the calculation.
- To the *Limited number of data*: the starting dataset, already not particularly large with about 51 acquisitions, was further reduced due to the need to perform a data cleaning process. The poor quality of some acquisitions made it necessary to exclude them from the analysis, thus further reducing the amount of data available and consequently affecting the analyses conducted. This limitation had a particularly significant impact on the analysis with classifiers. Starting from an already low number of data, the division between training and test set resulted in a very small test set, especially in the case of the binary classifier. Consequently, the results obtained must be interpreted with caution, as the validity of the test may have been compromised by the limited number of samples available.
- To the *Dataset Unbalanced*: the overall data was divided into the four analysis groups that were to be discriminated. However, a problem of imbalance between the classes emerged: some, such as *healthy and AK*, were characterised by a significantly higher number of acquisitions, while for others, such as *in situ cancer and pre-cancer*, the number of acquisitions was far lower. This imbalance affected the analysis, mainly to the disadvantage of the least represented classes.

In view of these limitations, some **future developments** could be:

- *Collecting more data by increasing the numerosity of the Dataset:* this would certainly lead to a more accurate analysis, characterised by better results also in terms of the statistical tests performed, especially with regard to classifiers, which the higher the number of data they train on, the better they perform.
- Connecting to the previous point, a significantly greater quantity of data could allow the study to be conducted by exploiting the *Deep Learning*: this would allow a network to be trained in such a way that, having taken the data processed as input, it extracts the binary mask and from it the skeleton, to then continue with the calculation of the parameters. Based on the extracted parameters, it could then predict the class to which the input acquisition belongs. A tool like this could be extremely useful to doctors as confirmation or support for diagnosis.

BIBLIOGRAPHY

- [1] C. P. H. Reinehr and R. M. Bakos. “Actinic keratoses: review of clinical, dermoscopic, and therapeutic aspects”. In: *Anais Brasileiros de Dermatologia* 94 (2020), pp. 637–657.
- [2] L. Del Regno, S. Catapano, A. Di Stefani, et al. “A review of existing therapies for actinic keratosis: current status and future directions”. In: *American Journal of Clinical Dermatology* 23.3 (2022), pp. 339–352.
- [3] I. Zalaudek, S. Piana, E. Moscarella, et al. “Morphologic grading and treatment of facial actinic keratosis”. In: *Clinics in dermatology* 32.1 (2014), pp. 80–87.
- [4] M. Fernandez Figueras. “From actinic keratosis to squamous cell carcinoma: pathophysiology revisited”. In: *Journal of the European Academy of Dermatology and Venereology* 31 (2017), pp. 5–7.
- [5] S. Aumann, S. Donner, J. Fischer, and F. Müller. “Optical coherence tomography (OCT): principle and technical realization”. In: *High resolution imaging in microscopy and ophthalmology: new frontiers in biomedical optics* (2019), pp. 59–85.
- [6] R. Fiorucci. *Dall’interferometro di Michelson alla misurazione delle onde gravitazionali*. Italian. 2022. URL: <https://it.emcelettronica.com/dallinterferometro-di-michelson-alla-misurazione-delle-onde-gravitazionali>.
- [7] R. Gniadecki and altri autori. “Approaches to quantify optical coherence tomography angiography metrics”. In: *International Journal of Molecular Sciences* (2020). URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7576021/#main-content>.

- [8] Z. Chen, M. Liu, M. Minneman, et al. “Phase-stable swept source OCT angiography in human skin using an akinetic source”. In: *Biomedical optics express* 7.8 (2016), pp. 3032–3048.
- [9] J. C. Chimal-Eguía, E. Castillo-Montiel, and R. T. Paez-Hernández. “Properties of the vascular networks in malignant tumors”. In: *Entropy* 22.2 (2020), p. 166.
- [10] D. Ribatti, B. Nico, E. Crivellato, and A. Vacca. “The structure of the vascular network of tumors”. In: *Cancer letters* 248.1 (2007), pp. 18–23.
- [11] M. Lupu, C. Caruntu, M. I. Popa, et al. “Vascular patterns in basal cell carcinoma: Dermoscopic, confocal and histopathological perspectives”. In: *Oncology letters* 17.5 (2019), pp. 4112–4125.
- [12] G. Argenziano, I. Zalaudek, R. Corona, et al. “Vascular structures in skin tumors: a dermoscopy study”. In: *Archives of dermatology* 140.12 (2004), pp. 1485–1489.
- [13] I. Pastushenko, T. Gracia-Cazaña, S. Vicente-Arregui, et al. “Squamous cell carcinomas of the skin explore angiogenesis-independent mechanisms of tumour vascularization”. In: *Journal of skin cancer* 2014.1 (2014), p. 651501.
- [14] P. Anvari, M. Ashrafkhorasani, A. Habibi, and K. G. Falavarjani. “Artifacts in optical coherence tomography angiography”. In: *Journal of Ophthalmic & Vision Research* 16.2 (2021), p. 271.
- [15] J. Wu, X. Jin, S. Mi, and J. Tang. “An effective method to compute the box-counting dimension based on the mathematical definition and intervals”. In: *Results in Engineering* 6 (2020), p. 100106.
- [16] T. Taylor. *The Cardiovascular System of the Head and Neck*. 2024. URL: <https://www.innerbody.com/anatomy/cardiovascular/head-neck>.

- [17] P. Mishra, U. Singh, C. M. Pandey, et al. “Application of student’s t-test, analysis of variance, and covariance”. In: *Annals of cardiac anaesthesia* 22.4 (2019), pp. 407–411.
- [18] M. Hassan. *MANOVA (Multivariate Analysis of Variance) – Method Examples*. March 26, 2024. URL: <https://researchmethod.net/manova/>.
- [19] J. Frost. *Interpreting Correlation Coefficients*. URL: <https://statisticsbyjim.com/basics/correlations/>.
- [20] T.-T. Wong. “Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation”. In: *Pattern recognition* 48.9 (2015), pp. 2839–2846.
- [21] L. Milanovic, S. Milenkovic, N. Petrovic, et al. “Optical Coherence Tomography (OCT) Imaging Technology”. In: *2021 IEEE 21st International Conference on Bioinformatics and Bioengineering (BIBE)*. 2021, pp. 1–5. DOI: 10.1109/BIBE52308.2021.9635099.
- [22] G. Fredman, S. R. Wiegell, M. Haedersdal, and G. R. Untracht. “Vascular feature identification in actinic keratosis grades I-III using dynamic optical coherence tomography with automated, quantitative analysis”. In: *Archives of Dermatological Research* 316.7 (2024), p. 391.
- [23] W.-J. Chen, Y.-Y. Chang, S.-C. Shen, et al. “In vivo detection of UV-induced acute skin effects using optical coherence tomography”. In: *Biomed. Opt. Express* 9.9 (Sept. 2018), pp. 4235–4245. DOI: 10.1364/BOE.9.004235. URL: <https://opg.optica.org/boe/abstract.cfm?URI=boe-9-9-4235>.
- [24] B. P. Tabaei and W. H. Herman. “A Multivariate Logistic Regression Equation to Screen for Diabetes : Development and validation”. In: *Diabetes Care* 25.11 (Nov. 2002), pp. 1999–2003. ISSN: 0149-5992. DOI: 10.2337/diacare.25.11.1999.

eprint: <https://diabetesjournals.org/care/article-pdf/25/11/1999/588640/dc1102001999.pdf>. URL: <https://doi.org/10.2337/diacare.25.11.1999>.

- [25] K. M. Meiburger, Z. Chen, C. Sinz, et al. “Automatic skin lesion area determination of basal cell carcinoma using optical coherence tomography angiography and a skeletonization approach: Preliminary results”. In: *Journal of biophotonics* 12.9 (2019), e201900131.
- [26] J. Lee, W. H. Jang, S. Shim, et al. “Characterization of early-stage cutaneous radiation injury by using optical coherence tomography angiography”. In: *Biomedical Optics Express* 11.5 (2020), pp. 2652–2664.
- [27] R. Leitgeb, F. Placzek, E. Rank, et al. “Enhanced medical diagnosis for dOCTors: a perspective of optical coherence tomography”. In: *Journal of Biomedical Optics* 26.10 (2021), pp. 100601–100601.
- [28] L. Veselka, L. Krainz, L. Mindrinos, et al. “A quantitative model for optical coherence tomography”. In: *Sensors* 21.20 (2021), p. 6864.
- [29] L. Milanovic, S. Milenkovic, N. Petrovic, et al. “Optical coherence tomography (OCT) imaging technology”. In: *2021 IEEE 21st International Conference on Bioinformatics and Bioengineering (BIBE)*. IEEE. 2021, pp. 1–5.
- [30] Y. Ji, S. Yang, K. Zhou, et al. “A Machine Learning Based Quantitative Data Analysis for Screening Skin Abnormality Based on Optical Coherence Tomography Angiography (OCTA)”. In: *2021 IEEE International Ultrasonics Symposium (IUS)*. IEEE. 2021, pp. 1–4.