UNIVERSIDADE D COIMBRA
FACULDADE DE CIÊNCIAS E TECNOLOGIA

TADEU FERREIRA

Comparative Analysis and Evolution Assessment of Dermoscopic Images

Thesis submitted to the
University of Coimbra for the degree of
Master in Biomedical Engineering

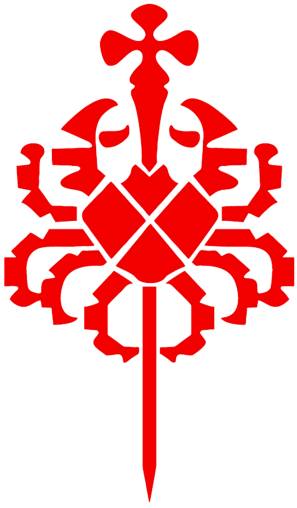
Supervisors:
Prof. Dr. Nuno Gonçalves (Institute of Systems and Robotics)

Coimbra, 2018

This work was developped in collaboration with:

Institute of Systems and Robotics

Portuguese Institute of Oncology Francisco Gentil



IPO COIMBRA

Agradecimentos

Em primeiro lugar, por todo o apoio, carinho, sacrifício e paciência ao longo destes 5 anos, um imenso obrigado aos meus pais, que nunca será suficiente para realmente lhes agradecer. O vosso apoio incondicional foi fundamental e a vocês dedico este trabalho.

Ao Professor Nuno Gonçalves um agradecimento especial por toda a orientação e ajuda inestimável ao longo do desenvolvimento desta dissertação.

Quero agradecer ao Instituto Português de Oncologia Franscisco Gentil pela vontade de colaborarem neste estudo, partilhando o conjunto de dados e, especialmente à Dra. Raquel Cardoso pela incansável disponibilidade na clarificação de qualquer dúvida.

Agradecimento especial ao Francisco Nunes pelas sugestões valiosas na fase final da dissertação da tese.

Ao Ricardo Barata e aos restos membros do laboratório de Visão por Computador pela atenção e tempo que dispensaram para as minhas dúvidas.

Ao Ziggy, amigo de todas as horas, mesmo nas de silêncio mais árduo, por toda a companhia.

Por fim, aos meus amigos de fé, meus irmãos camaradas por toda a companhia, apoio e amizade partilhada durante estes anos.

Resumo

De acordo com as estatísticas, o cancro da pele é um dos cancros mais prevalentes mundialmente com um rápido aumento do número de casos nas últimas décadas. O Melanoma é o cancro da pele mais perigoso, causando um maior número de mortes que qualquer outro cancro. No entanto, através de uma monitorização regular o seu diagnóstico prematuro e consequente irradicação é possível. Através da monitorização da pele, os dermatologistas são capazes de localizar alterações de padrões em lesões cutâneas previamente existentes. No entanto, mesmo para profissionais mais experientes, a quantificação de padrões ao longo do tempo é uma tarefa difícil.

Deste modo, por forma a facilitar a monitorização da pele e a classificar lesões malignas, vários sistemas de diagnóstico assistido têm vindo a ser desenvolvidos ao longo dos anos. No entanto, é escasso o trabalhado desenvolvido direcionado para a avaliação da evolução de lesões cutâneas. O desenvolvimento de um sistema automatizado que detete uma evolução significativa das lesões pode ser usada como ferramenta adicional, fornecendo um aviso aos dermatologistas e pacientes. O estudo presente nesta tese investiga a avaliação da evolução de lesões cutâneas ao longo do tempo através da análise comparativa de imagens dermoscópicas adquiridas ao longo do tempo. Nomeadamente a influência da aquisição de imagens dermoscópicas na avaliação da evolução das lesões foi estudada.

Esta avaliação foi realizada recorrendo a um algoritmo de duas fases. Inicialmente foi feita uma segmentação automática da área da lesão nas duas imagens a serem comparadas. Posteriormente, um alinhamento da máscara binária foi realizado de modo a extrair características da borda e forma da lesão que pudessem estar relacionadas com a evolução desta. A avaliação foi realizada com os métodos Fuzzy C-Means e Modelo de Mistura de Gaussianas e Classificação de Árvores de Decisão. Por sua vez, a performance foi avaliada usando rótulos extraídos de vários anotadores.

Os resultados obtidos sugerem que a avaliação da evolução com apenas características de borda e forma é insuficiente. Este estudo indica também que uma

aquisição incorreta de imagens dermoscópicas tem impactos negativos na extração de características. Foi ainda realizada uma análise que demonstra que a extração de rótulos de vários anotadores sem experiência médica afeta o desempenho do algoritmo, limitando a avaliação da evolução da lesão cutânea.

Os resultados da avaliação da evolução são aqui apresentados de modo a serem usados como incentivo para atrair mais investigação nesta área de estudo.

Palavras-chave: lesão cutânea, avaliação da evolução, dermoscopia, segmentação de imagem, registro de imagem.

Abstract

According to statistics, Skin Cancer represents one of the most prevalent cancers worldwide with a rapid increase in the last decades. Within skin cancer, Melanoma represents the most dangerous causing more deaths than any other cancer. Fortunately, early diagnosis of Melanoma has proven to be curable making the regular monitoring of the skin essential. When monitoring the skin, dermatologist looks for patterns changes in previously pigmented skin lesions. However, quantification of patterns over time is a difficult task even for the most experienced doctor.

Over the years, many automated diagnostic systems have been developed for classification of malignant lesions. However, there has been very little work directed towards the assessment of the evolution of pigmented skin lesions. Development of an automated system which detects a significant evolution of the skin lesion can be used as an additional tool by providing a warning to dermatologists and patients. This study investigates the assessment of the evolution of a skin lesion through comparative analysis of dermoscopic images acquired over time. Special attention has also been paid off how the acquisition of dermoscopic images may influence the assessment of evolution.

The assessment of evolution was performed in a two-step algorithm. The first step consisted of automated segmentation of the lesion area in the two images being compared. Afterward, alignment of the binary mask was performed in order to extract border and shape features which could relate to an evolution of the lesion. The assessment was computed with the clustering methods Fuzzy C-Means and Gaussian Mixture Models and, Decision Tree Classification. The performance was evaluated using labels extracted from multiple annotators.

The results suggested that assessing the evolution with only border and shape features is insufficient. This study also indicates that an incorrect acquisition of the dermoscopic images would have negative impacts on the correct extraction of features. An analysis has also been conducted that shows that the extraction of labels

from annotators with no medical experience affects the algorithm performance which limited the assessment of the evolution of the skin lesion.

The results of the assessment of evolution are presented such that they can be used as an incentive to attract more research in this area of study.

Keywords: skin lesion, evolution assessment, dermoscopy, image segmentation, image registration.

Glossary

CAD Computer-aided diagnosis.
CPD Coherent Point Drift.

DRA Demons Registration Algorithm.

EM Expectation-Maximization.

FFD Free Form Deformation.

GGR Global Growth Rate.
GMM Gaussian Mixture Models.
GVF Gradient vector flow.

LGR Local Growth Rate.

MSER Maximally stable extremal region.

NIM Number of incorrect matches.

RANSAC RANdom SAmpling Consensus.
ROI Region of Interested.

SIFT Scale Invariant Feature Transform.

UV Ultraviolet.

List of Figures

2.1 Anatomy of skin and subcutenous layer from a sectional view [1]. . . 7
2.2 Clinical and Dermoscopic images of benign pigmented skin lesions: (a) junctional nevus; (b) dermal nevus; (c) blue nevus; (d) dysplastic nevi; (e) congenital nevus; (f) spitz nevus. Original images from DermNet New Zealand [2]. 8

4.1 Pipeline of the algorithm methodology. 23
4.2 Representation of Fotofinder bodystudio ATBM tower (a) and FotoFinder medicam 1000 dermoscopy (b). Adapted from public domain images at https://www.fotofinder.de/en/products/bodystudio/. 25
4.3 Pipeline of preprocessing stage. 26
4.4 Red channel of input image. 27
4.5 Final candidate pixels after noise removal in vertical (a) and horizontal direction (b). 29
4.6 Dermoscopic image before (a) and after (b) the hair removal. 29
4.7 Image after color transformation (a) and contrast enhancement (b). . 30
4.8 Image (a) before and (b) after applying Local Normalization for correcting uneven illumination. 31
4.9 Outcoming binary image after Otsu's Threshold. 32
4.10 (a) Mask of the centered illuminated disk and (b) result of corner's noise removal. 33
4.11 Steps of Morphological Refinement of the proposed skin lesion segmentation algorithm: (a) First Morphological opening, (b) Image Erosion, (c) Second Morphological Opeing, (d) Image Dilation, (e) Morphological Closing (f) Flood-fill operation. 34
4.12 Gaussian Mixture Model centroids moving towards the reference points. The red border refers to the reference image and the blue border with the target image. 36

4.13 Results of four different difference images resulting from the alignment procedure. 38

4.14 An example of unidimensional data set. Adapted from "Fuzzy C-Means Clustering" [3]. 42

4.15 A typical membership function used in exclusive clustering methods such as K-Means. Adapted from "Fuzzy C-Means Clustering" [3]. . . 43

4.16 A typical membership function used in overlapping clustering methods such as Fuzzy C-Means. Adapted from "Fuzzy C-Means Clustering" [3] 43

5.1 Distribution of Jaccard Index values from segmentation results. . . . 50

5.2 Dermoscopic Images where the lesion has a low contrast comparing with the healthy skin. 51

5.3 Example of non acceptable segmentation. (a) Original dermoscopic image and (b) binary mask resulting from our segmentation. 51

5.4 Matrix of scatter plots of all features grouped by the labels. The features GGR, LGR, GA Rate, TA Rate, GA Diff and TA Diff stands for Global Growth Rate, Local Growth Rate, Green Area Rate, Total Area Rate, Green Area Difference, Total Area Difference respectively. 52

5.5 Matrix resulted from the calculation of correlation coefficient between all features. The features GGR, LGR, GA Rate, TA Rate, GA Diff and TA Diff stands for Global Growth Rate, Local Growth Rate, Green Area Rate, Total Area Rate, Green Area Difference, Total Area Difference respectively. 53

5.6 Example of skin lesion where the evaluation of growth was difficult and the labels were voted my minimal margin. The images on the left and right side represent the first image and the last image captured chronologically. 55

5.7 Example of skin lesion where the growth was evident and decided by unanimous vote. The images on the left and right side represent the first image and the last image captured chronologically. 56

5.8 Cluster partition using Fuzzy C-Means with All Features. The colors red, blue, yellow and green represent the cluster of lesion that did not evolve, where a significant evolution has occurred, correct real labels and incorrect real labels. 58

5.9 Cluster partition using Fuzzy C-Means with Top Features. The colors red, blue, yellow and green represent the cluster of lesion that did not evolve, where a significant evolution as occurred, correct real labels and incorrect real labels. 58
5.10 Cluster partition using Gaussian Mixture Model with All Features. The colors red, cyan, yellow and green represent the cluster of lesion that did not evolve, where a significant evolution has occurred, correct real labels and incorrect real labels. 59
5.11 Cluster partition using Gaussian Mixture Model with Top Features. The colors red, cyan, yellow and green represent the cluster of lesion that did not evolve, where a significant evolution as occurred, correct real labels and incorrect real labels. 59
5.12 Average of Precision, Recall, and Accuracy by Clustering method, built with all features and top features. The results are the average of one hundred runs. 60
5.13 Average ROC curve for the dataset with all features (a) and top features (b), with Fuzzy C-Means. The value of AUC was 0.50 and 0.51 respectively. 60
5.14 Average ROC curve for the dataset with all features (a) and top features (b), with Gaussian Mixture Model. The value of AUC was 0.57 and 0.54 respectively. 61
5.15 Average of Precision, Recall, and Accuracy by Classification Tree method, built with all features and top features. The results are the average of one hundred runs. 62
5.16 Average ROC curve for the dataset with all features (a) and top features (b). The value of AUC was 0.51 and 0.62 respectively. 62

List of Tables

2.1 Summary of the 7-point checklist. Based on [4]. 11
2.2 Summary of the Menzies Method. Based on [4]. 11
2.3 Summary of patterns of benign and malignant lesions. Based on [4]. . 12

4.1 Name of the labelled skin lesion and number of images per skin lesion. 24
4.2 Summary of database details. 26
4.3 Summary of database details after image selection for registration step. 36

5.1 Evaluation of segmentation algorithm using the statistical metrics Sensitivity, Specificity, Accuracy, Precision, Dice Coefficient, and Jaccard Index. 50
5.2 Number of skin lesions that the clustering algorithms detect an significant growth. 60

Contents

Glossary **xiii**

List of Figures **xv**

List of Tables **xix**

1 Introduction **1**

1.1 Context 1

1.1.1 The Skin Cancer Problem 1

1.1.2 The Research Problem 2

1.2 Motivation 2

1.3 Structure 3

2 Human Skin Medical Background **5**

2.1 The Human Skin 5

2.2 Cancers of the skin 6

2.2.1 Pigmented Skin Lesions 7

2.3 Melanoma 9

2.4 Skin Imaging Techniques 9

2.4.1 Dermoscopy 10

2.5 Diagnostic Methods 10

3 Literature Review **13**

3.1 Preprocessing 13

3.2 Segmentation 15

3.3 Feature Extraction 17

3.4 Registration and matching 18

3.5 Evolution Assessment 20

4 Methods **23**

4.1 Image Acquisition and Dataset . . . 23
4.2 Preprocessing . . . 25
4.2.1 Resizing . . . 26
4.2.2 Hair Removal . . . 26
4.2.3 Color Transformation and Contrast Enhancement . . . 30
4.2.4 Correction of Non-Uniform Illumination . . . 30
4.3 Segmentation . . . 31
4.3.1 Otsu's Thresholding . . . 32
4.3.2 Morphological Refinement . . . 33
4.3.3 Blobs Filtering . . . 34
4.4 Registration . . . 35
4.4.1 Image Selection . . . 35
4.4.2 Coherent Point Drift . . . 35
4.5 Feature Extraction . . . 37
4.6 Evolution Assessment . . . 40
4.6.1 Data Labeling . . . 41
4.6.2 Data Clustering . . . 41
4.6.2.1 Fuzzy C-Means . . . 42
4.6.2.2 Gaussian Mixture Models . . . 44
4.6.3 Decision Tree Learning - Classification Tree . . . 44

5 Results and Discussion **47**
5.1 Statistical Metrics . . . 47
5.2 Segmentation . . . 48
5.2.1 Ground Truth and Segmentation Metrics . . . 48
5.2.2 Quantitative Results . . . 49
5.3 Feature Selection . . . 51
5.4 Evolution Assessment . . . 54
5.4.1 Data Labeling . . . 54
5.4.2 Clustering . . . 57
5.4.3 Decision Tree Learning . . . 60
5.5 Brief Conclusions . . . 62

6 Conclusions **65**
6.1 Summary of Thesis . . . 66
6.2 Limitations and Future Work . . . 67

Bibliography **69**

1

Introduction

1.1 Context

1.1.1 The Skin Cancer Problem

In the present day, skin cancer represents one of the most common types of cancer among the Caucasian population [5]. Its prevalence has been increasing progressively in the last few decades reaching a total of 2 to 3 million skin cancers and 132.000 melanoma skin cancers each year [6]. To a certain extent, Melanoma comprises only 4% of all skin cancers but is responsible for 80% of skin-cancer related deaths [7]. Advanced melanoma is the fastest growing malignancy in men and the second-fastest growing in women [8]. Every year, more than 100.000 new cases of Melanoma are diagnosed in Europe and more than 22.000 European citizens lose their lives to the disease [9].

The human skin is, nowadays, increasingly subject to vulnerabilities, much of it from overexposure to potentially harmful sun rays. This is mainly due to continuously loss of the atmosphere protective filter - the ozone layer - allowing more Ultraviolet (UV) radiation reaching the Earth's surface [6].

The monitoring of skin lesions should, according to the recommendations of dermatologists, be done regularly by the person himself, with high regularity in most cases. Depending on the degree of risk associated with the type of skin and personal history, some patients should be accompanied by medical specialists, with prescribed regularity. Early diagnosis of melanoma skin cancer results in a survival chance of 95 %, while late detection follows a fatal prognosis with an average survival of 6 to 9 months [10].

However, monitoring a skin lesion is quite a difficult task. When performing skin inspection, dermatologists look for patterns or early signs that are associated with a

malignant growth of the lesion. Such a task is very subjective since it is dependent on the human visual perception. A systematic approach is required, although diagnostic experience can be an asset. However, even the most experienced dermatologist will have occasional difficulty to make the best possible diagnostic decision. An erroneous decision can have negative consequences such as the evolution of the malignant form of the skin lesion - the melanoma.

On account of these statistics, there is the need for an additional tool for complementing traditional diagnosis.

1.1.2 The Research Problem

In the last decades, Computer-aided diagnosis (CAD) has been much explored by researchers in the area of medical dermatology. These are systems that assist doctors in the interpretation of medical images. Various approaches for implementation of CAD systems of skin lesions have been developed. The majority of the objective of these CAD systems is the classification of different skin lesions and especially the detection of Melanoma. However, not much attention has been paid to the development of automated systems for evaluating changes in skin lesions. This would require to have at least two or more images of the same lesion captured at a different time. Thus, much of the work performed outside the main scope focus on performing a precise registration [11, 12] or extracting robust features for evolution assessment [13].

1.2 Motivation

The use of Computer-aided diagnosis (CAD) has been crucial in filling gaps in diagnoses performed by the naked eye by dermatologists. The diagnosis is directly dependent on the visual acuity (the ability of the eye to distinguish two very close points) of the dermatologist, where only the macroscopic characteristics of the lesion are evaluated, ignoring critical microscopic characteristics [14]. In other words, human vision lacks precision and the ability to quantify information from an image.

Dermatologists often have high-capacity visualization instruments, such as cameras and lens systems - dermoscopes, capable of enlarging with great detail the skin's lesions, and still capable of relating these marks in time, making it easier to see images with the characteristics of sharpness, correctness, and resolution required

for short diagnosis. On the other hand, for self-control monitoring, the people themselves do not have the same medical devices to acquire the image and, for this reason, it will be more difficult to obtain images with the same quality and resolution that can alert them to potentially significant alterations and, therefore, to the need to consult a doctor.

According, to the ABCDE method of detection of Melanoma (detailed in Subsection 2.5), evolution is one of the essential parameters to be analyzed when monitoring possible skin lesion malignancies. In order to assess the evolution of a skin lesion, there is a need to acquire images of possible malignant skin lesions for growing quantification. However, due to the lack of availability of databases of this nature, especially of free access, we proposed a protocol with the Portuguese Institute of Oncology Francisco Gentil for the sharing of dermoscopic time-series images.

Still, when images are acquired at different times, their acquisition may be done differently or present artifacts that weren't present in previous acquisitions. For that matter, registration (the process of obtaining the best alignment between two images) plays the first vital step to assess lesion evaluation. Computing a proper registration between images is critical to allow a direct and better assessment of the evolution of skin lesion.

In this work, we pretend to assess the skin lesion evolution through the extraction of border and shape features, in order to provide a warning to the doctor when a lesion may have evolved or not. In order to do so, registration of automatically segmented binary masks was also performed. Furthermore, we desire to understand how the conditions for image acquisition and treatment may influence the primary goal.

1.3 Structure

The remaining of this dissertation is organized as follows:

- **Chapter 2** provides the medical background of the skin lesion as well as image techniques for assessing lesion details. Moreover, a summary of the most used clinical diagnostic methods for melanoma is also described.
- **Chapter 3** describes the state of the art of image analysis techniques typically integrated into automated diagnosis systems. Moreover, a brief description of the attempts done for evolution assessment is provided.
- **Chapter 4** explains in detail all the different stages and methodologies used

towards the evolution assessment.

- **Chapter 5** presents results from different stages of workflow along with discussion analysis.
- **Chapter 6** concludes the dissertation by displaying an overview of the work together with its limitations and future work.

2

Human Skin Medical Background

2.1 The Human Skin

The skin is the largest organ in the human body covering all of its external surface area. It serves both as a mechanism of communication between the internal and the external environment and as a defense structure. In addition, it cooperates with other organs for the proper function of the organism such as thermoregulation and elaboration of important metabolites like Vitamin D. The human skin is a stratified estructure consisting of two main distinctive parts: the dermis and the epidermis.(refer to Figure 2.1).

The epidermis is the most superficial and thinner layer of the skin. It is composed of Stratified squamous epithelium which is a layered, paved and keratinized tissue. The tight junction between these cells makes the epithelium tissue an efficient barrier against the entry of invading agents and the loss of body fluids. Internally, the epidermis is divided into several layers, from bottom to top: stratum basale, stratum spinosum, stratum granulosum, and stratum lucidum. In these four layers, there are presently four of the principal type cells, i.e., Keratinocytes, Langerhans cells, Merkel cells, and Melanocytes.

- Keratinocytes, also known as basal cells, represent the majority of the epidermis and are located in the stratum basale. Within the stratum basale, keratinocytes are continually dividing and forming new cells. On account, they are forced to differentiate and migrate towards the surface of the epidermis. During this migration, these cells become further away from the blood vessels, which are located in the dermis. Since the epidermis does not have blood vessels, these cells will eventually die from lack of nutrients and oxygen and discarded [1].

- Langerhans cells are dendritic cells originated from bone marrow that later

migrated to the epidermis. These cells only represent a small portion of total cells of the epidermis where their primary objective is to detect and report dangerous invasive cells to the immune system. Moreover, UV radiation can cause a reduction in the number of normal Langerhans cells [15]

- Melanocytes are dendritic cells found in the epidermis responsible for the production of the pigmented melanin to the surrounding epidermis cells. Apart from giving color to the skin, it serves as a protection against UV radiation from sunlight. In other words, when the skin is exposed to radiation, the melanocytes produce more melanin in order to protect the skin from damaging, resulting in tanning and, in exceptional cases, the appearance of freckles.

Regarding the objective of this study, Melanocytes represent the most important type of cells on account of malignant transformation potential.

The other important layer of the skin, the dermis, is located just beneath the epidermis. The dermis is formed of dense connective tissue of elastic and collagen fibers. Together it forms a mesh that gives flexibility to the skin. Its composition is essentially collagen (about 70%) and other glycoproteins and fibers of the elastic system.

2.2 Cancers of the skin

The classification of skin lesion into benign and malignant is based on the effects they have on the human body. Benign skin lesions are slow-growing, expansive and well tolerated by the host organism. On the other hand, malignant skin lesions have rapid and unlimited growth and may infiltrate in other tissues. In some individual cases, malignant skin lesions can develop in such a way that can spread to other tissues of the body. This is also known as metastasis. The development of malignant skin lesions are essentially mutations of oncogenes induced by UV radiation.

Skin cancers can develop from any cells, but non-pigmented skin cells are more likely to develop to skin cancer. The most common non-pigmented cancers are Basal Cell Carcinoma and Squamous Cell Carcinoma. They are named from their respective cells such as basal cells and squamous cells respectively. Luckily, these type of cancer is more easily to cure [16].

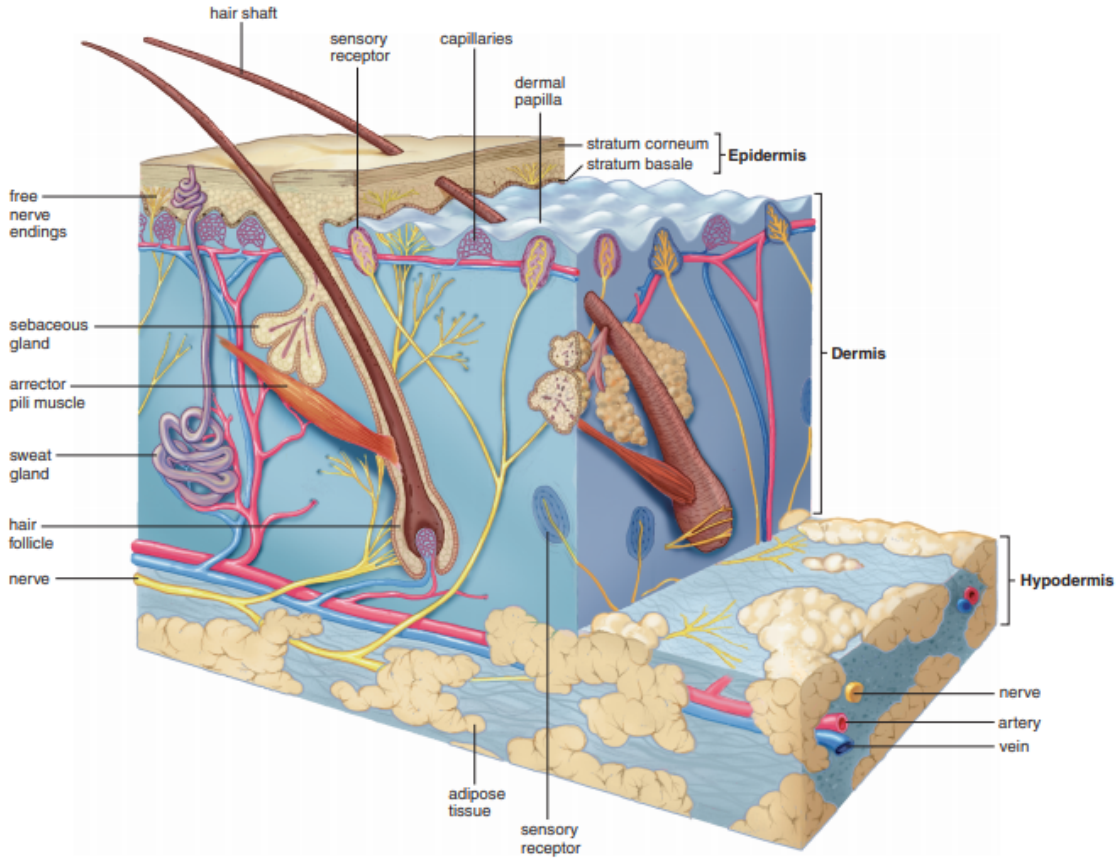


Figure 2.1: Anatomy of skin and subcutenous layer from a sectional view [1].

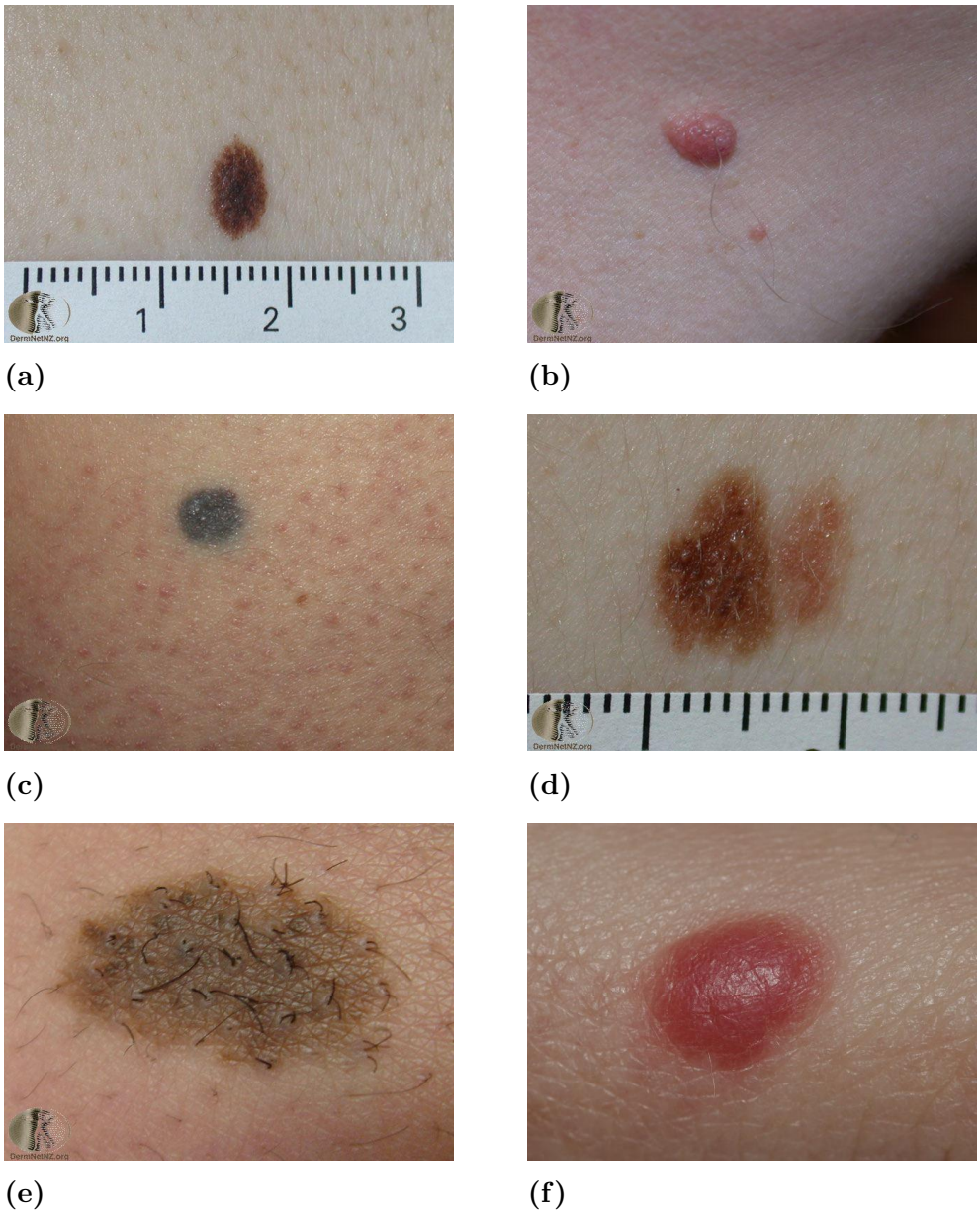
2.2.1 Pigmented Skin Lesions

Pigmented Skin lesion are the other type of skin lesion and, are divided into benign and malign lesions. Bening lesions are commonly known as moles and are characterized by their symmetrical shape and small size. However, atypical moles can develop into their malignant form, also known as Melanoma. Nevertheless, there are many types of pigmented skin lesions that we should know about:

- **Common nevus** (Figure 2.2(a)-(b)) - is the most common mole in the human body and can be further divided in junctional nevus, compound nevus and dermal nevus, depending on which layer of the skin it grows [16].
- **Blue nevus** (Figure 2.2(c)) - are a variant of the common nevus characterized by their rare variant of color blue.
- **Dysplastic nevus** (Figure 2.2(d)) - also known as atypical nevus, is known for its similar appearance to melanoma. Atypical nevus are usually larger than the common nevus, with a big variant of colors. Besides is one of the main

precursors of melanoma [17].

- **Congenital nevus** (Figure 2.2(e)) - they are present since birth and can be larger than 20 cms. They are highly suspicious to evolve to melanoma depending on their size. [17, 16].
- **Spitz nevus** (Figure 2.2 (f)) - is distinguish by its small pinkish modules and are usually seen in children [16].



(a) **(b)**

(c) **(d)**

(e) **(f)**

Figure 2.2: Clinical and Dermoscopic images of benign pigmented skin lesions: (a) junctional nevus; (b) dermal nevus; (c) blue nevus; (d) dysplastic nevi; (e) congenital nevus; (f) spitz nevus. Original images from DermNet New Zealand [2].

Accurately diagnosing a benign skin lesion and distinguishing it from a malignant

condition requires consideration of the physical and histological characteristics of the lesion as well as the patient's attributes and overall condition. Biopsy or surgical excision is commonly performed when a potential malignancy cannot be ruled out [17].

2.3 Melanoma

Lately, the incidence of Melanoma has been rising in a fast pace in New Zealand, Australia, North America, and Europe, precisely in the white population, with highest incident seen in Australia and New Zealand [18]. In contrast, the disease is less prevalent in African Americans and very exceptional in the Japanese population. There are many risk factors that we can control to avoid Melanoma. The main factor is the exposition to sunlight but, there are risk factors that can be prevented such as time outdoors during childhood, proximity to the equator and viruses [16].

Melanoma can appear from normal skin cells or pigmented skin lesion. In case of pre-existent pigmented skin lesion, take place a change of in size, change in color and exhibition of an irregular border. It is up to us to be vigilant and aware of a change of color, size or border and, if this occurs, assessment by a specialist is required. When Melanoma is detected, surgery is always the most indicated treatment. However, in advanced stages, when metastasis has occurred, palliative treatments like radiotherapy, chemotherapy, and immunotherapy are only with the objective to relieve the symptoms [1, 16].

2.4 Skin Imaging Techniques

Nowadays, there are numerous non-invasive skin visualization techniques to help diagnose different skin lesions. Clinical photography was a breakthrough in dermatology since it was the beginning of the use of technology for storing information in favor of medicine. Besides clinical photography, many imaging techniques allow for the visualization of different skin lesion structures. For instance, dermoscopy, confocal scanning laser microscopy, optical coherence tomography, high-frequency ultrasound, magnetic resonance imaging, positron emission tomography, among others [19].

2.4.1 Dermoscopy

Dermoscopy is a non-invasive skin imaging technique that promotes the visualization of small structures not detectable to the naked eye. This technique is described due to its capability of optical magnification up to x10. The technique is based on the interposition between a lens and the skin of some drops of water, oil or ultrasound gel. In this way, the skin becomes translucent allowing the dermatologist to observe up to the dermo-epidermal junction numerous peculiar characteristics for the early diagnosis of melanoma that are not visible to the naked eye [20, 21]. With the development of the technology, new-generation dermoscopes do not need the use of immersive fluid but, instead, an integrated polarized light, allowing the rapid screening of multiple skin lesions in a short period.

Overall, dermoscopy has proven to be a reliable and important tool, providing additional accurate analysis, that would be inconceivable with only the naked eye, allowing the physician's experience in the diagnosis of pigmented skin lesions [21].

2.5 Diagnostic Methods

Over time, with the crescendo of information, there was a need of well defined criterias to help dermatologist, physicians and clinicians diferentiate pigmented skin lesions in melanocytic and non-melanocytic and later melanocytic in benign and malignant. The most known diagnostic methods that are currently use are: the ABCDE rule of dermoscopy, the 7-point Check List, the Menzies Method and finally the Revised Pattern Analysis [4]. ABCDE rule is the most commonly used in dermoscopy and was first introduced by Friedman et al. in 1985 [17]. Initially, in ABCD rule, A stands for assymetry, B for irregularity or sharpness of the border, C for color variation and D for diameter (or differential sctructures). Later, Abbasi et al. [22] proposed the inclusion of evolving as one additional criteria, since suspicious pigmented skin lesions may evolve in different aspects over time. The 7-point Checklist, the Menzies Method and the Revized Pattern Analysis are summarized in the tables 2.1, 2.2 and 2.3 respectively.

After all, in the majority of cases, these clinical diagnostic methods show low accuracy on account of different measurements consensus, subjective of diagnosis and poor correlation of standards [23].

7-point checklist	
(1)	Atypical pigmented network
(2)	Blue-white veil
(3)	Atypical vascular pattern
(4)	Irregular streaks
(5)	Irregular pigmentation
(6)	Irregular dots/globules
(7)	Regression structures

Table 2.1: Summary of the 7-point checklist. Based on [4].

The Menzies Method
Negative features
Point and axial symmetry of pigmentation
Presence of a single color
Positive features
Blue-white veil
Multiple brown dots
Pseudopods
Radial streaming
Scar-like depigmentation
Peripheral black dots-globules
Multiple blue/gray dots
Broadened network

Table 2.2: Summary of the Menzies Method. Based on [4].

	Benign	Malignant
Dots	Centrally located or situated right on the network	Unevenly distributed and scattered focally at the periphery
Globules	Uniform in size, shape, and color, symmetrically located at the periphery, centrally located, or uniform throughout	Globules that are unevenly distributed and when reddish are highly suggestive of melanoma
Streaks	Radial streaming or pseudopods tend to be symmetrical and uniform at the periphery	Radial streaming or pseudopods tend to be focal and irregular at periphery
Blue-white veil	Tends to be centrally located	Tends to be asymmetrically located or diffuse almost over entire lesion
Blotch	Centrally located or may be diffuse hyperpigmented area that extends almost to periphery of the lesion	Asymmetrically located or there are often multiple asymmetrical blotches
Network	Typical network that consists of light to dark uniform pigmented lines and hypopigmented holes	Atypical network that may be nonuniform with black/brown or gray thickened lines and holes of different sizes and shapes
Network borders	Either fades into the periphery or is symmetrically sharp	Focally sharp

Table 2.3: Summary of patterns of benign and malignant lesions. Based on [4].

3

Literature Review

This chapter aims at contextualizing and show the state of the most recent studies within the area of interest of this dissertation. For this reason, different approaches are reviewed in this chapter focusing on the following analysis steps: preprocessing, segmentation, feature extraction, registration and matching and finally evolution assessment. Thus, this chapter attempts to demonstrate the obstacles that the state of the art has encountered to overcome the problem of the evolution and comparative analysis of skin lesions.

3.1 Preprocessing

Preprocessing aims at improving the image by correcting any defect arising from its acquisition and highlighting important details for the analysis. In order for the segmentation step to have satisfactory results, it is necessary that the image be with the minimum of imperfections. There are many artifacts present in pigmented skin lesions, such as hair, air bubbles, ruler markings, shadows and reflections. Preprocesing can be divided in two main groups of methods: artifacts removal and image enhancement.

Hair represents one of the biggest obstacles before segmentation. The presence of hair, especially thick hair, causes a considerable loss of accuracy during segmentation. The detection of the border can become very difficult or even impossible, not reflecting the reality. Previously, before the use of specialized software, the shaving of the lesion area was needed. To avoid this problem, the first method most widely used and on which later proposed methods are based is the dull razor [24]. Sultana et al. use the technique of Top Hat Transform (described in detail in Subsection 4.2.2) to detect hairs resulting in a mask of hair that is later removed using an inpainting method based on DullRazor [23]. Recently [25], made a study of different

hair removal methods based on three techniques: linear interpolation, inpainting by Partial Differential Equation (PDE)-based diffusion and exemplar-based. They realized that fast inpainting methods were faster and with better performance. In the same study, [25] used a Derivative of Gaussian (DOG) for hair detection and fast marching inpainting scheme. Therefore, [25] proposed new hair removal based on a multi-resolution inpainting method which uses a multi-resolution coherence transport inpainting method [26]. In addition, methods based on morphological operations such as erosion and dilation were also proposed. Møllersen et al. [27] applied a top hat operation using a long and thin structuring element (SE), in both horizontal and vertical directions. In the end, the detected hair were replaced using pixels values from morphological closing using an SE disc.

Even though hair represents the artifact which may influence more negatively, there are others that can also affect further algorithms such as air bubbles, reflections, ruler (scale) marks, small pores, among others. Median Filter is one of the most used techniques to deal with this type of artifacts due to its ability to remove noise while preserving edges. In 1995 and later in 2003, Kjoelen et al. [28] and Maglogiannis [29] used a [3x3] size filter for noise reduction due to the presence of hair, scales and light reflections. In further studies, Celebi et al. [20] state that the appropriate median filter size should be proportional to the image size for better results, being the size determined by this rule of thumb:

$$n = floor(5.\sqrt{((M/768).(N.512))}) \tag{3.1}$$

The choice of 3.1 is based on good results from a 768 x 512 image with an n=5 median filter. One article, in particular, applied the technique Gaussian Smoothing which consisted of a 3x3 size Gaussian filter to minimize the effects of small artifacts [30].

Regarding Color Transformation, various approaches have been adopted. Both [25] and [31] changed color space from Red-Green-Blue (RGB) to CIE L*a*b due to its relative perceptual uniformity and better human perception. On the contrary, [32] converted RGB to gray-scale for better functioning and accuracy of further algorithms. In addition, [33] used Shades of Gray as a Color Constancy method in order to retrieve the real colors that are affected by the acquisitions imperfections. The objective of this method is "to transform the colors of an image I, acquired under an unknown light source, so they appear identical to colors under a canonical light source."

Furthermore, the problem of unbalanced illumination has been an issue that authors have been trying to correct. Cavalcanti et al. [34] transformed RGB to Hue, Saturation, Value (HSV), and retrieve the four corners of the Value channel to estimate the new local illumination in order to minimize shading effects. Differently, [32] decided that Local Normalization (LN) would be appropriate to correct non-uniform illumination and shading artifacts, and increase local contrast.

Finally but not least, black frames are a visual characteristic of dermoscopic images. Since dermoscopy contains different sizes and shape lens, these images may contain dark bands with different widths. Sultana et al. [23] proposed an implementation of a circle mask with variable size to establish an ROI. Since most of the skin lesions are not centered in the image, starting from the middle, the circle mask was iteratively increased until all lesion was contained.

3.2 Segmentation

The study of skin lesions segmentation has been one of the areas that research have dedicated more efforts for improvement and performance. Within the area of computer vision, segmentation refers to the process of decomposition of a digital image into several segments (regions) that form it. In other words, segmentation can be considered a process of classification of pixels, where they are grouped based on the sharing of certain characteristics such as color, intensity or texture. When dealing with dermoscopic images of pigmented skin lesions, the objective aims to locate the exact boundary between the surrounding skin and lesion area. An accurate lesion segmentation is very important, especially for border and shape quantifications. The problem of attaining an accurate border detection by automatic segmentation is that dermatologist do not usually delimitate the skin lesion resulting in a ground truth problem [35]. On top of that, certain pigmented skin lesions are not well defined due to depigmentation, low lesion-to-skin gradient, multiple lesion regions, among others, making the process of skin lesion segmentation, not an easy task [36].

For this reason, plenty of algorithms have been proposed, and in the following paragraphs, a brief discription of some of them is presented.

One of the most usual approaches is the Threshold-based methods. Based on the fact that the values of pixels that belong to a skin lesion differ from the values of the background, thresholding methods will classify each pixel, taking into account whether the calculated value exceeds a certain threshold or not. The discriminant

value varies in the literature, varying from using the luminance value calculated from RGB channels [37], to calculate which component from RGB channel presents the highest entropy [38]. Within thresholding algorithms, segmentation can range from simple to adaptive or hybrid thresholding. For instance, [39] decided for simple thresholding, where Zortea et al. [31] and Pereira et al. [32] decided for adaptive thresholding. Zortea et al. proposed a weighted thresholding method where it uses the Otsu's Thresholding [40] in a new intensity image estimated by the size of classes from background skin and lesion with a histogram analysis. Pereira et al., on the other hand, first computes a Region of Interested (ROI) by varying the Local Normalization scale and later, applying threshold depending on the maximum intensity of the ROI.

Region growing or region-based methods are probably the second most used approach and consists of grouping pixels or sub-regions into larger regions. [30] is one typical example where initial seed points are empirically decided, and the segmentation stops when the skin lesion has been isolated. Moreover, Celebi et al. [20] decided for an unsupervised approach for color segmentation using Statistical region merging algorithm based on region growing and merging. Rajab et al. [41] even compared a region-based segmentation with neural network edge detection. In this study, [41] demonstrates that the region-based method provided the best performance over a range of signal to noise ratios in lesions with different border irregularity properties.

Segmentation can also be achieved by the application of Active-contours which consists in detecting the object contours using curve evolution techniques. Erkol et al. [42] applied Gradient vector flow (GVF) snake to determine the lesion border. Then an automatic initializing of the snakes by smoothing the image using a 15x15 Gaussian filter and later applying the Otsu's Thresholding to provide the under-segmentation of the skin lesion region. Later, Tang et al. [43] proposed a Multi-direction GVF snake which consists of an extension of the basic GVF snake. The method traced the lesion boundary expertly despite the presence of other objects near the skin lesion area.

Segmentation methods based on soft computing was also investigated. These methods involve the classification of pixels using soft-computing techniques such as fuzzy logic, neural networks, and evolutionary computing. Therefore, [44] proposed to segment the lesion based on fuzzy c-means (FCM) thresholding while [41] used a neural network for edge detection.

Another method which is becoming extremely popular in computer vision is super-

pixel segmentation. This consists of clustering a group of pixels with similar characteristics. Navarro et al. [11] proposed an adaptation of Simple Linear Iterative Clustering (SLIC) with the following detail: instead of initializing the centers using a regular grid, they replace the center initialization with Scale Invariant Feature Transform (SIFT) feature points, forcing the superpixels to be smaller and more precise around the lesion.

There is an abundant number of methods for segmentation with, every one of them, has its advantages and drawbacks. Some authors compared some methods to understand which one was adequate. In 2009, Silveira et al. [38] evaluated 6 methods for the segmentation of skin lesions in 100 dermoscopic images: Adaptive thresholding (AT), Gradient Vector Flow (GVF), Adaptive Snake (AS), Level Set Method of Chan et al. [45] (C-LS), Expectation-maximization level set (EM-LS) and fuzz-based split-and-merge algorithm (FBSM). The results obtained were ground truth images, and the errors of segmentation were evaluated using four different methods. TDR was considered of the 4, the most relevant method of evaluation from the point of view of the clinic, and the best methods of segmentation were AS and EM-LS (which are semi-supervised, with TDR = 95.47% and TDR = 95.20% respectively) and FBSM (which is a fully automatic method, with TDR = 93.67%).

3.3 Feature Extraction

There are many works concerning feature extraction in skin lesions. We exposed an overview of extraction of features from the three main groups: shape, color, and texture.

The irregularity of shape is one of the characteristics referred to in the ABCDE rule since this features may infer patterns of melanocytic development or unstable regression of melanoma. Among the various types of analysis in the literature for the extraction of form, the regional descriptors are the most used, but other like Fourier descriptors are also present. These regional descriptors can be used to assess geometrical properties such as area, perimeter, equivalent diameter, compactness, circularity, solidity, rectangularity, aspect ratio, and eccentricity. Regarding asymmetry of the lesion, average, variance and standard deviation are some of the features used to describe it [46].

Concerning color, the authors state that color identifiers are mainly statistical parameters calculated from different color channels, such as the mean value, standard

deviation, variance, minimum and maximum colors, and color skewness of color channels. RGB color space is usually used although some authors preferred to convert to other color space such as HSV, HSL, CIE Luv, CIE Lab since they are based on the human perception of colors [46]. Other authors have used an optical model of the skin to interpret the colors that appear in a lesion, finding that all normal colors of the skin lie on a two-dimensional surface fragment within a three-dimensional (3-D) color space [47].

The texture is a descriptor that provides measures like smoothness, roughness, and regularity and it is frequently used because it assists in discriminating between malign and benign lesions. Depending on the technique used, several characteristics can be obtained from the texture of the lesion. One of the methods that allow obtaining dozens of descriptive statistical data of the texture is based on the gray-level co-occurrence matrix (GLCM) which characterizes the texture of an image by calculating how many pairs of pixels exist with a specific value and a specific spatial relationship [48, 30]. There are alternatives texture descriptors that were also used for better discrimination of skin lesions. For example, [36] made an exhaustive review of other features adopted, for instance, Wavelet-based descriptors, Gabor filter descriptors, Intensity distribution descriptors Haralick descriptors, Local binary pattern (LBP), SIFT and color SIFT, among others.

3.4 Registration and matching

Registration consists of transforming different sets of images into one coordinate system, by computing a transformation between two or more representations of the same subject. This transformation is executed in one image in order to optimally fit in the other image, called the reference image [49]. When dealing with skin lesion images, registration is crucial for monitoring the evolution and calculating the changes in the skin lesions. As a matter of fact, image registration allows the calculation of border features that can be used to monitoring the progress of skin cancer [29].

Concerning single lesion analysis, some studies have been done. Maglogiannis et al. [29] proposed the Log-polar representation of the Fourier Transform. The Fourier Transformation spectrum is independent of horizontal and vertical shifting, whereas the log-polar transform eliminates the dependency on magnification and rotation. This algorithm showed good overall results with better performances for rotation up to 10 degrees and scale factor up to 1.0. For bigger rotations, the algorithm

tended to over calculate rotation corrections, whereas scale correction tended to undercalculate above 1.0 scale factor.

Montin et al. [50] proposed two methods with different deformation models: a b-spline-based model called Free Form Deformation(FFD) and a Demons Registration Algorithm (DRA). The latter is a fast and accurate method for deformable mono-modal registration of non-segmented images, while FFD is a combination of Mean Square Error as the similarity metric, b-splines as transformation model and quasi-Newton as an optimizer. After tuning the models with simulated images, DRA showed to be more reliable, stable and robust than FFD with evaluation results of 77.14%, 93.88% and 100% for skin, border, and inner nevus respectively. However, the DRA algorithm is more adequate for quantitive measurements of human skins' in vivo biomechanical properties.

Anagnostopoulos et al. [12] projected the problem of registration with the ROI-SIFT algorithm, a modification of SIFT algorithm. ROI-SIFT consisted in calculating SIFT points using a "hard" value of SIFT parameter h and then, ROIs are defined according to this points. Later, the h is increased, producing a larger number of key points; however, only key points belonging to ROIs are kept. Then RANdom SAmpling Consensus method (RANSAC), is used to identify outliers and maximize the registration accuracy. Finally, rigid image registration is performed based on the best homography found.

Navarro et al. [11] proposed a Superpixel(SP)-SIFT technique. This technique resides in previously segmentation of the image in superpixels isolating important skin lesion area. Next, computes the SIFT descriptor using only the pixels within the superpixel containing the skin lesion, resulting in a full skin lesion segmentation mask feature descriptor. The detected feature is used to authenticate a correct match between two images and evaluate possible lesion change.

In addition to registration regarding single pigmented skin lesion, there have also been studies concerning mapping and registration of multiple lesions. This reason is due to the difficulty of doctors in examining and monitoring the entire body of the patient. Huang and Bergstresser [51] proposed a new hybrid method for finding correspondences between dermatological images based on a bipartite graph matching. It was from Voroni cells and distances between points that the authors were able to convert the point registration problem in images to bipartite graph matching problem.

Korotkov et al. [52] proposed a non-rigid point cloud registration. The skin lesions

are detected using Maximally stable extremal region (MSER). Next, a homography transformation is computed between ROIs (that were previously detected using MSER). SIFT is used to detect important key points between ROIs while RANSAC is used to find correspondences and eliminate any possible outlier.

Eventually, Mirzaalian et al. [53] tested five matching algorithms in 56 pairs of dermatological images: Coherent Point Drift (CPD), shape contexts (SC+TPS), the spectral technique of (Spect) and Hypergraph matching (d = 2) (Hyp). The matching was evaluated using the number of incorrect matches (NIM) being that CPD presented the least NIM among all tested algorithms.

Regardless of the registration process being applied to single or multiple skin lesion images, registration is an important step since, in the majority of cases, evolution assessment is dependent on image registration.

3.5 Evolution Assessment

According, to the ABCDE method of detection of Melanoma, evolving is one of the important parameters to be analyzed when monitoring possible skin lesion malignancies. Nevertheless, there is still much work ahead concerning the development of automated systems for evaluating changes in skin lesions.

Prigent et al. [54] proposed two methods to measure the global area of changes and the local changes: a binary change detection area and multi-level change detection, in other words, homogeneity. The global area of change is calculated by subtracting two consecutive image series while the homogeneity is approached by modeling the change binary mask as a realization of a Gaussian random field under the null hypothesis. Later, the model is split into regions, and a probability is assigned to each region.

Navarro et al. [11] proposed to measure skin lesion evolution by comparing the segmented areas and calculate a pixel-level difference. Knowing that the scale of the images was known, the author was able to provide the skin lesion evolution in a comprehensive metric.

Korotkov et al. [52] attempt a similar approach by computing the difference image using each channel individually of the previously converted HSV representation. The change mask is obtained by combining the difference images and binarized using a threshold provided.

Mendes et al. [13] experiment with a different approach by using a mathematical model based on planar linear transformations. In this study, relevant information such as the variation on the growth along the border, the deformation and the symmetry of the lesion through time were measured. The models were first validated in synthetic binary images of skin lesions and then, tested in real dermoscopic melanoma images, resulting in similar results.

Furthermore, [55] from two segmented images that were previously aligned using Principal Component Analysis (PCA) and stochastic gradient descent obtained a change map. This same researchers investigated by comparison, how computed features would change at different time points [56].

Despite all these attempts, evolution assessment in pigmented skin lesion is still raw and more studies will be needed. The development of an algorithm capable of detecting small and accurate changes could be an essential tool for the dermatologist on the early detection of melanocytic lesions.

4

Methods

The methods used in this study were carefully chosen after the study of the literature. The proposed methodology is divided into 6 main stages with all of them described in detail throughout this chapter. The pipeline of our algorithm methodology is outlined in the Figure 4.1.

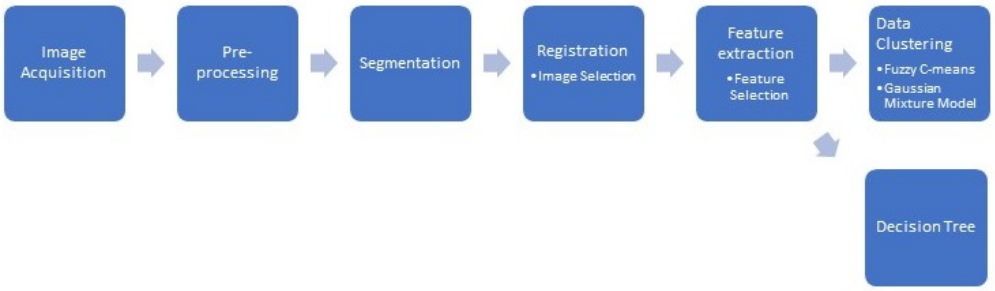


Figure 4.1: Pipeline of the algorithm methodology.

The algorithms used in each stage of the pipeline were implemented using MATLAB 2018a with the exception of Registration that was implemented in Python 3.6.

4.1 Image Acquisition and Dataset

The first stage of study began with the acquisition of images, and a crucial step for the continuity of this study. By rule, dermatologist often uses high capacity visualization instruments to monitoring pigmented skin lesions from time to time to inspect possible evolutions. For this matter, we request time series of dermoscopic images to the Portuguese Institute of Oncology of Coimbra Francisco Gentil.

The dataset consisted of 89 pigmented skin lesions with a total of 444 time series dermoscopic images. The images are acquired by rule, in the interval of 6 in 6 months; however, the interval between some images is higher than six months because of postponing of medical appointments or when the atypical nevi became too suspicious and needed to be removed. This led to a non uniform of images per skin lesion. In addition, the dataset is retrieved from 9 patients who are monitored by the dermatologist. The information of the patients was anonymous, and therefore no information about the name, age, and gender was stored.

The most crucial point when describing the image data set provided by the Portuguese Institute of Oncology of Coimbra Francisco Gentil was that no ground truth about the lesion area or any label regarding the evolution of the skin lesion was provided except for eight skin lesions. These 8 lesions were the only that the dermatologist had a record of having evolution, having no information if other lesions could or could not have evolved. Unfortunately, the label regarding the evolution of the skin lesion was vague since we only knew that the skin lesion had evolved somewhere between the time series images. This led to only 8 labels in our study being all of them in the same class - a significant evolution had occurred. Besides, no ground truth about the segmentation area was provided in this eight lesions. This represented the biggest drawback of the entire study since our primary objective was to provide a warning trigger when a significant evolution had occurred in the skin lesion. In order to give more insight of the concerning lesions, we display in Table 4.1 more detail information.

Lesion	**Nº of Images**
P01L01	7
P01L17	2
P02L02	7
P02L04	8
P03L02	9
P04L08	5
P05L02	2
P07L03	3

Table 4.1: Name of the labelled skin lesion and number of images per skin lesion.

The images were acquired using a digital dermoscopy of Fotofinder Body Studio which has an automatic focus leading to no information about the real size of the skin lesion. The only information about the acquisition was that the magnification was always set to 20x. This was also a drawback when analyzing the evolution in a

further chapter.

Figure 4.2: Representation of Fotofinder bodystudio ATBM tower (a) and FotoFinder medicam 1000 dermoscopy (b). Adapted from public domain images at https://www.fotofinder.de/en/products/bodystudio/.

Due to the existence of excessive undesirable artifacts or skin lesion which wouldn't fit entirely within the image frame, a total of 11 lesions were not considered in our study, not being reliable for comparison and evolution analysis. In the end, 379 images were suitable and proceeded to be analyzed. A summary of the final dataset can be shown in Table 4.2.

All the image files were JPG format, with an 8-bit RGB color graphics and a spatial resolution of 1920x1080 pixels.

4.2 Preprocessing

The pipeline of our study begins with the preprocessing stage. Knowing that the dermoscopic images have in their presence an abundant type of artifacts, it is essential to remove these unwanted artifacts since it may affect border detection when segmenting the skin lesion. Besides, image enhancement plays a critical role allowing to a better distinction of the skin lesion with the background skin. The proposed pipeline of preprocessing is depicted in Figure 4.3.

Total number of patients	9
Total number of skin lesions	78
Total number of images	379
average number of images per skin lesion	5
Minimum number of images per skin lesion	2
Maximum number of images per skin lesion	10

Table 4.2: Summary of database details.

Figure 4.3: Pipeline of preprocessing stage.

4.2.1 Resizing

We began our preprocessing stage by resizing our image frame by a scale factor of 1/3 of the original size. The image frame was initially 1920x1080 pixels and, as a result of the downsizing, the scale image was 640x360 pixels. This procedure allowed the following steps, especially the segmentation, to be faster and have similar performances. Transforming the images before resizing them was slower and did not improve the results.

4.2.2 Hair Removal

The presence of hair is a considerable obstacle for a correct segmentation, showing to be an impediment for border detection. Thick and dark hairs are the main interferences but also thin hair when located next to the lesion boundary. In this study, the hair removal algorithm was based on the one proposed by [27] due to its

logical and fast implementation. The algorithm was divided into the following steps:

1. Generation of Candidate Pixels;
2. Noise Removal;
3. Hair pixels Replacement.

The first step in the algorithm was detecting hair in the dermoscopic image. The first approach for detecting the hair was to convert the RGB image to a grayscale image since only one RGB channel is used for hair detection. However, the contrast between hair and skin seemed to be greater when we only used the red layer of the RGB image. In the end, the red channel was chosen and can be seen in Fig. 4.4.

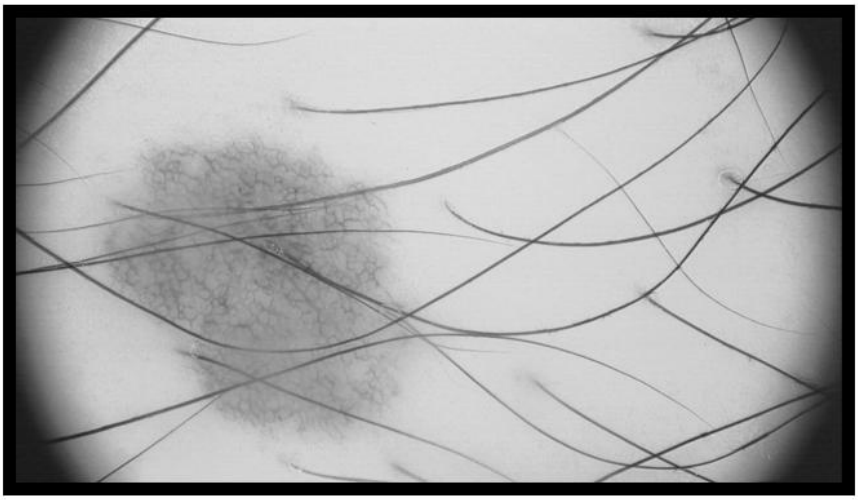


Figure 4.4: Red channel of input image.

Assuming that the hair is darker than the surrounding skin, a grey level threshold is applied in the red layer of the input image using the Otsu's threshold which selects the optimal threshold that maximizes the interclass variance between two classes [40]. However, instead of using only one threshold calculated with the Otsu's method, we used 5 variations of the Otsu's Thresholding by multiplying it by 0.85, 0.90, 0.95, 1.00 and 1.05. Knowing that sometimes the Otsu's threshold over-segment or under-segment the areas of interest, these allowed detecting from the darkest hairs to the more lighter, without missing any possible candidate in the dermoscopic image.

The detection of candidate hair pixels started by applying a morphological operation

named Top Hat Transform filtering. The method computes the difference between the image and its morphological closing [1] :

$$T_i = I - I_c(I, e_i) \tag{4.1}$$

In 4.1, I represent the input image and I_c describes the morphological closing of input Image with a structuring element e_i [2]. The structuring element adopted in this procedure is used to identify the hair, which is long and thin and therefore, a line-shaped structuring element is used. When detecting possible hair pixels in the vertical direction, the top-hat algorithm with a horizontal structuring element is implemented in the binary image. Afterward, a morphological opening [3] is applied to eliminate possible objects that are too short to be a hair. This same procedure was repeated in the horizontal direction yet, in this case, a vertical structuring element is used instead. Besides repeating this procedure for vertical and horizontal direction, there was a need to use a more extended structuring element for thick hair detection. To sum up, the top-hat filtering was applied using two different sizes of the structuring element for thin and thick hair detection in both horizontal and vertical direction.

After the previous step, there was a lot of noise present in the binary image such as objects that were not hair or even texture from the skin lesion. In order to eliminate the noise present in the binary image, the properties eccentricity and major axis length were calculated for all the objects shown in the binary image. The eccentricity is the ratio distance between the center of the ellipse and its major axis length, taking the values of 0 (circle) to 1 (line segment). The elimination of excess objects followed the condition exhibit in equation 4.2.

$$MajorAxis < t_1 \vee Eccentricity < t2 \tag{4.2}$$

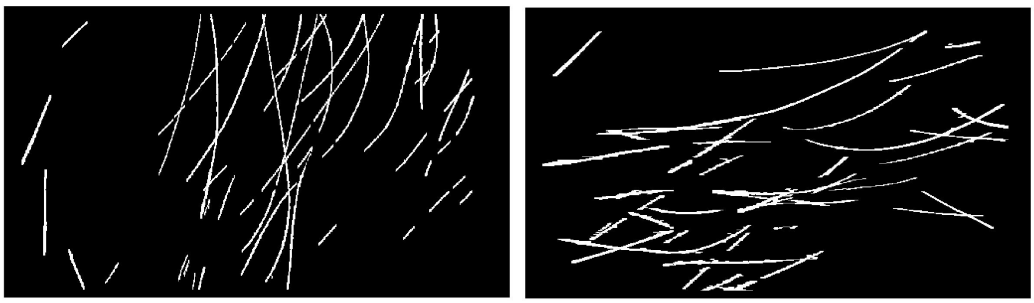
The thresholds values displayed in the above equation were chosen empirically and by comparison with ones used by other authors [24, 27]. Apart from this, it is important to note that different thresholds had to be chosen to filter thin and thick hair, whereas, for eccentricity, the threshold had to be close to 1. For the final step

[1] Morphological closing refers to the mathematical morphology process of erosion following dilatation of an image set using a structuring element.

[2] In image processing, structuring element is a group of pixels of different shapes used in morphological operations such as dilation, erosion, opening, and closing

[3] Morphological opening refers to the process of dilatation followed by an erosion of an image set using a structuring element.

of noise removal, a dilation process using a 3x3 square was applied. This step was necessary in order to smooth the border of the hairs to be removed.

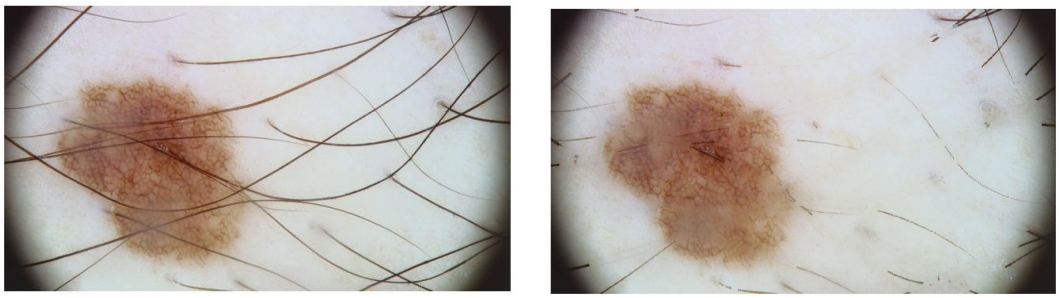


(a) **(b)**

Figure 4.5: Final candidate pixels after noise removal in vertical (a) and horizontal direction (b).

The end of hair removal step resides on eliminating the final candidate pixels followed by the replacement of the hair pixels with non-hair-valued pixels. The replacement of the hair pixels is done by linear interpolation. The pixels used for the interpolation process are located at a distance of 3 pixels perpendicular to the length of hair. However, the replacement of the region of hair is not only applied in the red channel but also in the green and blue of RGB, combining all three channels and forming a correct, no hair, color image.

One important aspect that was not mentioned previously was that the image was rotated 4 times with angles of 0, 22.5, 45 and 67.5 degrees. This was a necessary procedure in order to detect hair in every possible direction since the real direction of the hair was unknown. In short, this process covered 8 directions (counting with vertical and horizontal direction) with 5 different thresholds. The result correction is displayed in Figure 4.6.



(a) **(b)**

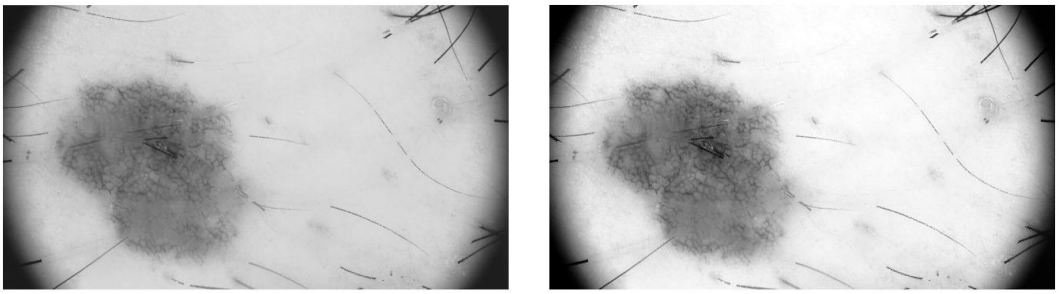
Figure 4.6: Dermoscopic image before (a) and after (b) the hair removal.

4.2.3 Color Transformation and Contrast Enhancement

Once the main artifacts were removed, we decided to convert our RGB image to a one-layer image. Data reduction to a one-layer image was needed to fit our segmentation method and also for performance matters. Therefore, we decided to convert the true color RGB image to a grayscale intensity image. The conversion is based on a weighted sum of the R, G, and B components (see Equation 4.3).

$$I_{grayscale} = 0.299 \times R + 0.587 \times G + 0.114 \times B \tag{4.3}$$

As a result of this process, the converted image was a non-normalized with an uneven distribution of intensities, presenting eligible areas that were too bright or too dark. So as to recover the information of eligible areas, contrast enhancement was applied to normalize the converted image. Firstly, we found the lower and upper limits that could be used for the contrast enhancement. Then a contrast stretch method is applied by stretching the intensity values of the converted image to the new values calculated in the previous step. The resulting image is presented in Figure 4.7.



(a) **(b)**

Figure 4.7: Image after color transformation (a) and contrast enhancement (b).

4.2.4 Correction of Non-Uniform Illumination

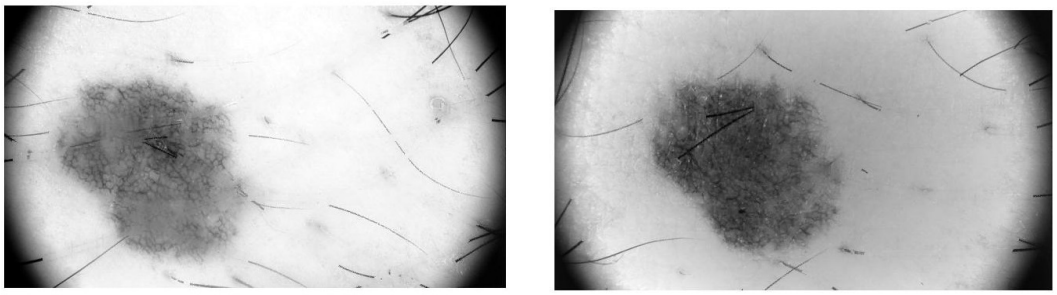
At this point, we have a grayscale image followed by contrast enhancement. However, unbalanced illumination is still visible in the images almost due to the acquisition process of the dermoscope. Modern dermoscopes have integrated polarized light positioned in the edge of the circular disk which, sometimes may cause some uneven illumination areas or brighter artifacts such as air bubbles or light hairs.

Therefore, we perform the Local Normalization algorithm developed by [57]. The algorithm aim at uniformizing the mean and variance of the image around a local

neighborhood, being suitable for correct uneven illumination or shading artifacts. The algorithm is defined by:

$$g(x,y) = \frac{f(x,y) - m_f(x,y)}{\sigma_f(x,y)} \tag{4.4}$$

where f(x,y) represents the input image and $m_f(x,y)$ is the estimation of local mean of the input image, while σ_f is an estimation of the local variance. Local Normalization uses fast recursive Gaussian filters for the estimation of local mean and variance. The parameters used for estimation of the local mean and variance were 230 and 200 respectively, which represent the size of the smoothing window. These parameters were decided empirically. Finally, the final result is displayed in Figure 4.8.



(a) **(b)**

Figure 4.8: Image (a) before and (b) after applying Local Normalization for correcting uneven illumination.

4.3 Segmentation

At this point, we have a preprocessed image with the major artifacts removed plus, vital areas of the dermoscopic image, especially the boundary, highlighted using image enhancement techniques. Thus, in this step, using Thresholding segmentation along with morphological operations, we intend to isolate the pigmented skin lesion area from the background skin. This is a crucial step for the further stage since an accurate segmentation is determinant for image registration. Our segmentation method is described in detail below.

4.3.1 Otsu's Thresholding

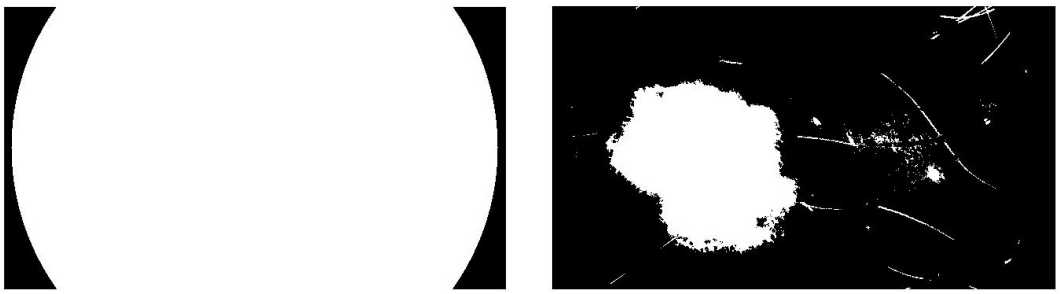
Our segmentation method started by calculating the Otsu's Thresholding in the preprocessed image. This method assumes the existence of two classes, in this case: skin lesion and clear skin and, thoroughly select the optimal threshold that minimizes the intra-class variance, defined as the weighted sum of variances of the two classes [40]. However, as it is known from dermoscopic images, the ROI is represented as a disc-shaped illuminated center with a dark surrounding.

Moreover, the dark surrounding is not totally black, fading to the illuminated area and, the disc-shaped illuminated center is not entirely contained in the image. Due to these facts, the Otsu's Threshold algorithm will consider that the classes are the dark surrounding and the bright center, leaving the skin lesion undetected. We thus computed a circular mask with the center in $(x,y) = (320,180)$ and radius of 300 pixels that was chosen empirically. As a result, the group of pixels used in the Otsu's Threshold was the ones located inside the illuminated disc, helping to balance the classes sizes. The outcoming image after the Otsu's Thresholding is displayed in Figure 4.9.

Figure 4.9: Outcoming binary image after Otsu's Threshold.

4.3.2 Morphological Refinement

Observing the final result of the segmentation and after all the morphological operations (that will be explained next), the final lesion mask would be connected with a thin layer of the dark surrounding. Seeing that, we applied once again the same circular mask to eliminate any noise located in the frame corners as shown in Figure 4.10(a).



(a) **(b)**

Figure 4.10: (a) Mask of the centered illuminated disk and (b) result of corner's noise removal.

After analyzing the Image 4.10 it was clear that there was much noise presented that needed to be removed as well as the irregular shape of the skin lesion boundary. Seeing that, we started by performing a morphological opening to remove small unwanted objects that were considered positive pixels in the thresholding segmentation.

Consequently, for the following steps, we created a disk-shaped structuring element, that was used to slightly erode the edges of the structures in the image in small disk shapes. This way the margins were smoothed. Then, dilatation [4] is applied to recover the volume of the border lost after application of image corrosion. However, before this operation, an opening operation is performed to enable the elimination of small undesirable objects avoiding its ample growth resulting from the dilatation.

Due to the inconsistency of certain blobs, it was necessary to arrange a solution for the standardization after treatment. For this purpose, the closing function was initially applied, which join blobs at a certain distance from each other. Finally, in order to fill possible holes left within the main blob, we performed a flood-fill operation on the negative class of pixels of the skin lesion binary mask. The results of the steps detailed above are presented in Figures 4.11.

[4]Dilation adds a layer of pixels to both inner and outer boundaries depending on the size and shape of the structuring element

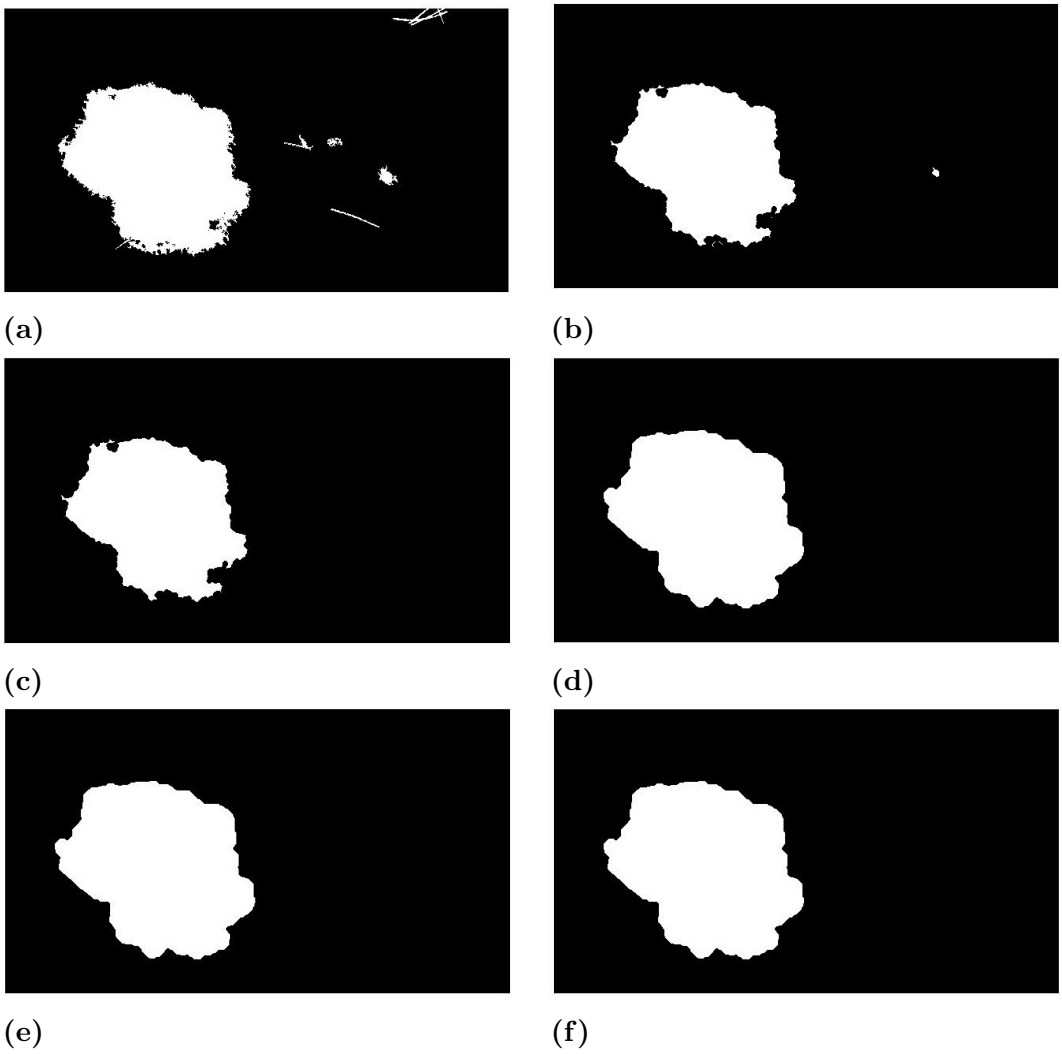


Figure 4.11: Steps of Morphological Refinement of the proposed skin lesion segmentation algorithm: (a) First Morphological opening, (b) Image Erosion, (c) Second Morphological Opeing, (d) Image Dilation, (e) Morphological Closing (f) Floodfill operation.

4.3.3 Blobs Filtering

At this point, the final result of the skin lesion segmentation mask had been reached. Still, in some cases, beyond the skin lesion segmentation, there were some isolated areas visible in the binary image. For instance, the presence of large hair pores or even the existence of furuncles (infection of the hair follicle) was one of the cases of the isolated areas. Besides, dermoscopic images with a vast tangle of hair away from the lesion could also contribute to this.

For this matter, we decided to label all the blobs visible and compute measurements of the following set of properties: area and centroid. Then, for each labeled blob, we

calculate the distance from their centroid to the center of the image frame, dividing by its area. The goal is to keep the largest blob within the image frame which is closest to the center of the image frame. A final binary image was obtained after removing the blobs which had a larger distance to the center of the image frame per area.

4.4 Registration

In order to access the progress of a skin lesion, a registration step was performed. Registration is required when images are captured at different times, and in this way, a direct comparison is possible for monitoring the evolution and quantification on any significant changes.

4.4.1 Image Selection

Before initializing the registration process, we performed a second selection of images that would be suitable for evolution assessment. This selection was more strict since variations in change are very sensitive to external factors. For this reason, skin lesions with abysmal results in segmentation were rejected since the binary mask did not accurately represent the lesion area. Besides, images that were not entirely contained within the image frame were also denied even though their binary mask were a very good representation of the skin lesion. Therefore, of the 379 images which were used up to this step, 59 were not suitable and consequently removed for further studies. Once the image dataset was carefully chosen, we could initialize our registration step. A summary of the dataset for image registration after image selection can be shown in 4.3.

4.4.2 Coherent Point Drift

At this point, we had an accurate binary mask of the skin lesion area resulting from the segmentation phase. Since the goal was the alignment of segmented binary structures that were previously extracted, the choice of our registration method had to take this into account. For that matter, an intrinsic registration method was elected as it relies on the patient-generated image content only [58], that is, the anatomical information and features generated by the subject image itself.

Total number of patients	9
Total number of skin lesions	64
Total number of images	320
average number of images per skin lesion	5
Minimum number of images per skin lesion	2
Maximum number of images per skin lesion	10

Table 4.3: Summary of database details after image selection for registration step.

Seeing that, we decided to use a probabilistic method for point set registration named Coherent Point Drift (CPD) that was proposed by [59]. The CPD consists of aligning two point sets based on probability density estimation. The target points set is modeled as a Gaussian Mixture Model (GMM) centroids and forced to move coherently, towards the reference set of points by maximizing the likelihood (see Figure 4.12).

The transformation is obtained by optimizing the Maximum A Posteriori estimation and constraining the GMM centroid locations with specific parameters. These parameters are directly regulated by the type of transformation to be performed. As was mentioned in Subsection 4.1, the digital dermoscope automatically focuses the skin lesion. Also, the variation of the lesion area over time is too low, to cause a variation of the dermoscopy workout distance. Since the goal is a direct evaluation of the lesion evolution, a non-deformable registration is needed. For that reason, rigid parameters such as rotation and translation for the alignment of the two images were applied using CPD.

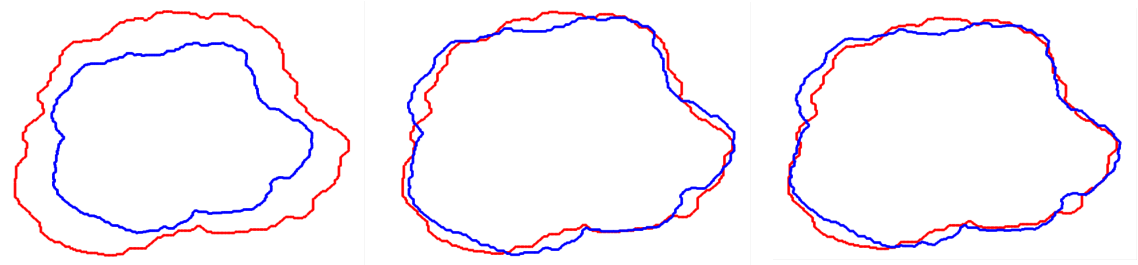


Figure 4.12: Gaussian Mixture Model centroids moving towards the reference points. The red border refers to the reference image and the blue border with the target image.

For the alignment of the two images, we did not use the whole binary mask of the skin lesion. Using the binary image as a whole would require the optimization of many points in space, leading to a decrease in the algorithm performance. This way, we decided to extract the border of the image to be used as input to the CPD algorithm. Nevertheless, when we applied the CPD algorithm it does not align two borders but two sets of points of the lesion border. In order to transform the image to a set of points, we make use of the Cartesian coordinate system to transform the lesion border pixels to a pair of numerical coordinates. Once the transformation matrix is calculated between the two borders, we then apply the transformation matrix to the whole binary mask. However, the result transformation not only includes integer points but also float points in the cartesian coordinates. As we know, an image is a representation of a set of pixels equally distant from each other. Due to this fact, a rounding operation was performed to the resulting set of a pair of coordinate points to ensure that the transformed binary mask was filled to its exact location in the image frame.

Besides, the transformed binary mask undergo a closing procedure to eliminate possible holes. These holes may appear when the rounding operation is applied, skipping a pixel without being fulfilled.

Finally, to accomplish a systematic evaluation of the skin lesion evolution, we use the first captured image of each lesion as the image of reference. Then, the more recently image was used to align with the skin lesion when it was first detected. The result is a registration between the image when was last acquired with the initial captured image. The decision of only aligning two images of each lesion on account of the limited label data for each skin lesion. More details are depicted in Subsection 4.6.1.

4.5 Feature Extraction

In this section, we aim at exploiting the extract of some features to analyze the lesion evolution. However, in this study we do not focus on extracting features for characterization of the skin lesion but instead, obtaining essential features in aligned images to assess the lesion evolution.

In order to extract features from the registration previously performed, a direct subtraction from the two aligned images was computed. Through the subtraction of the two binary masks, we aim to identify the exact location where the skin lesion

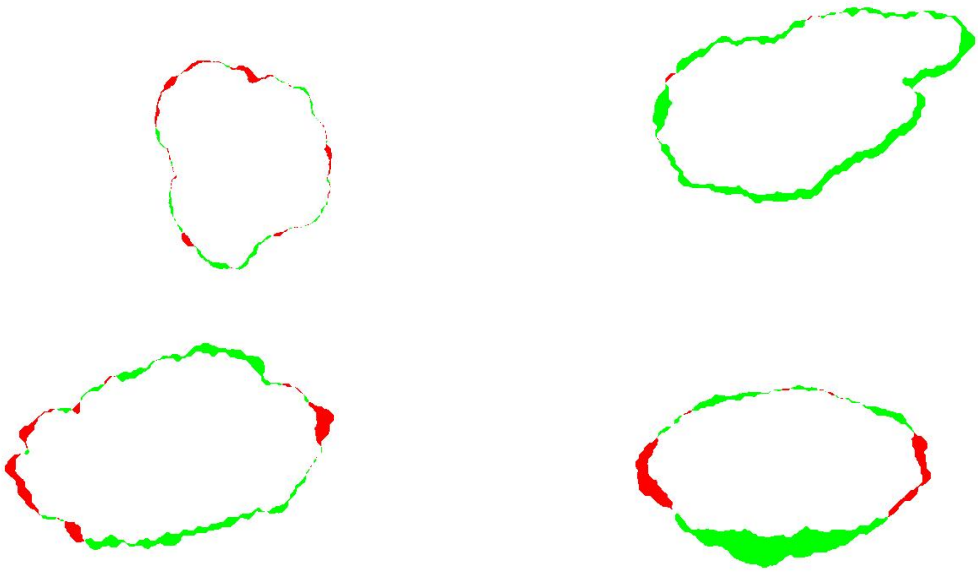


Figure 4.13: Results of four different difference images resulting from the alignment procedure.

had evolved. Besides for a better identification where the lesion has changed, a color identification was used to identify the following changes:

- Green - Areas where the reference image was smaller than the target image.
- Red - Areas where the reference image was bigger than the target image.
- Black - Areas where there were no changes between the reference image and the target image.

These changes are shown in Figure 4.13.

After analyzing the difference image, it appears that areas where the skin lesion had evolved, were of the same magnitude as the areas where it supposedly decreased. These may be due to the incorrect segmentation and border detection in some areas or even perhaps inconsistency during the alignment of images. Seeing that, we decided to extract the area where the skin lesion had grown, but also the total area resulted from the difference image, this is, both areas where had increased and decreased.

In order to expand the assessment of the skin lesion evolution, a growth area rate was computed. The growth area rate is the rate of change of area from one period to another, that is the difference between the reference and target image. This growth area rate was both applied to the total area of difference between aligned images,

but also to the area where the lesion had increased - the green area.

Also, we decided to extract two features which could characterized by how the skin lesion grows locally and globally based on the distance between the border points of the two lesions. The assessment of the local growth is used to determine if the growth through time has been proportional in all directions of the skin lesion. In other words, we wanted to assess if the lesion suffers any deformation through time. The feature is named Local Growth Rate (LGR) and was proposed by Mendes et al. [13]. The feature is described as follows:

$$LGR(i) = \frac{\sqrt{(x_i^{'} - x_i)^2 + (y_i^{'} - y_i)^2}}{d_{i0}} \tag{4.5}$$

where $d_{i0} = \sqrt{(x_i - x_0)^2 + (y_i - y_0)^2}$ and $i = 1,...,M$ for a sampling with M points where the points (x_i, y_i), $(x_i^{'}, y_i^{'})$ and (x_0,y_0) are respectively points in the border of the reference image, the corresponding points in the border of the target image and the centroid.

The deformation is designated as ΔLGR and measured as the deviation of $LGR = LGR_i : i = 1,...M$. ΔLGR is defined in Equation 4.6.

$$\Delta LGR = \frac{max(LGR) - min(LGR)}{mean(LGR)} \tag{4.6}$$

The assessment of the global growth is done through a feature named Global Growth Rate also proposed by Mendes et al. [13]. The Global Growth Rate (GGR) provides an overall insight into the overall lesion growth and can be described as follows:

$$GGR(i) = \frac{\sqrt{(x_i^{'} - x_i)^2 + (y_i^{'} - y_i)^2}}{max(d_{i0})} \tag{4.7}$$

The deviation of GGR can provide a better understanding if the lesion grows globally at the same distance along the border, and can be calculated by:

$$\Delta GGR = \frac{max(GGR) - min(GGR)}{mean(GGR)} \tag{4.8}$$

Finally, we had extracted a perimeter growth rate of the skin lesion. The perimeter growth rate allows determining how the shape had changed by measuring the length of the outline shape over time. More precisely, is the difference between the perimeter

of the target image and the perimeter of the reference image in order to the reference image. This was the only feature that was not dependent on the image registration.

Before assessing the lesion evolution, we evaluate if there was any correlation between all the features extracted from the alignment of lesions, from the border points and the perimeter rate. In other words, we decided to inquire if features could be redundant in the presence of another relevant feature. Seeing that, a correlation coefficient was computed to measure their linear dependence. The coefficient of correlation is a statistical method used in medical research to assess a possible linear association between features since it is simple to calculate and interpret [60]. The Pearson's Coefficient is the most common and used in this study and is defined as

$$\rho(A,B) = \frac{1}{N-1} \sum_{N}^{i=1} \left(\overline{\frac{A_i - \mu_A}{\sigma_A}} \right) \left(\frac{B_i - \mu_B}{\sigma_B} \right) \tag{4.9}$$

where μ_A and σ_A are the mean and standard deviation of A respectively, where μ_B and σ_B are the mean and standard deviation of B. A and B are two random variables.

Once the Pearson's coefficient was computed, it was clear that some features were strongly correlated. Seeing that, we decided to remove the Global Growth Rate and features related to the total area, that is, the rate of the total area growth and the difference of the total area between skin lesions.

4.6 Evolution Assessment

Evolution Assessment defined the last step of our algorithm, by having the complicated task to determine an optimum threshold in order to warn dermatologist of a significant lesion evolution using features described in Subsection 4.5. We proposed in this step two different approaches:

1. Since no label was provided, through unsupervised machine learning, we attempt to infer a group which had a significant evolution by looking at the extracted features.

2. In case of Unsupervised Learning demonstrating be unable for such task, through the extracted labels (see Subsection 4.6.1) apply a supervised learning algorithm that could map a correct model to the appropriate output.

4.6.1 Data Labeling

The detection of significant evolution of a skin lesion is not an easy task even for dermatologists. During the evaluation of the possible growth of the skin lesion, dermatologists have to measure if the lesion has evolved clinically. However, the procedure is sensitive to intra and inter dermatologist diagnosis.

Since no label about the evolution of the skin lesion was provided, we decided to create our labels of the 64 skin lesions. Owing to the difficulty of assessing the skin lesion evolution visually, we resort to the technique of Crowdsourcing [61]. In crowdsourcing, a group of multiple annotators are asked to contribute by performing a task that cannot be individually done with the same ease. For that matter, we asked seven persons to compare the first and last dermoscopic images of each skin lesion and label if a significant growth has occurred. The strategy for deciding the final label was by Majority Voting. This was a crucial step in trying to incorporate more information into the problem in order to try to create a more sophisticated and reliable automatic solution than clustering.

4.6.2 Data Clustering

The choice of data clustering was not by chance but by being unsupervised learning since no real labels about the lesions evolution were provided. These methods are used when all data is unlabeled and the algorithms learn to inherit the structure from the input data. In other words, we are looking for a threshold that allows dividing lesions where a significant growth has occurred.

The goal of clustering is to create groups of data points such that points in different clusters are dissimilar while points within a cluster are similar. That way, our primary desire was to present a distinct cluster that allowed a proper detection of significant growth, since dermatologist have difficulties in quantifying it by the naked eye.

In our study, two clustering methods were evaluated to solve this problem: Fuzzy-Means and Gaussian Mixture Models. We started by giving a brief description of the used algorithms.

4.6.2.1 Fuzzy C-Means

The fuzzy Clustering method is a soft form of clustering which, unlike K-Means [5], assigns degrees of membership in various clusters to each data point. Fuzzy C-Means was the second clustering method we tested. What is unique in this method is that, rather than entirely belonging to one cluster, each data point has a probability of belonging to each cluster. Like K-means, Fuzzy C-means still converges by minimizing the square error function; however, it uses a weighted sum based on the membership function.

The Fuzzy C-means algorithm is similar to the k-means and presents as follows: an initial fuzzy partition of N objects into K clusters is selected by picking a membership matrix U. Later, a probability of each data point belonging to a given cluster K is computed. The new centroid of the cluster is recalculated as a weighted centroid given the probabilities of membership U of all data points. The algorithm stops when entries in the membership U does not change significantly or when a specific number of iterations is reached.

In other words, the convergence is met when the square error function is minimized, as shown in Equation 4.10.

$$E^2(U) = \sum_{i=1}^{N} \sum_{k=1}^{K} u_{ij} \left\| x_i - c_k \right\|^2 \tag{4.10}$$

Where $c_k = \sum_{i=1}^{N} u_{ik} xi$ is kth fuzzy cluster center and u_{ij} represents the grade of membership of object x_i in cluster c_j wherein $u_{ij} \in [0,1]$.

To better understand the difference between both clustering algorithms, we display two images of how a group of data points is divided using a hard and a soft clustering technique.

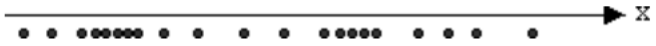


Figure 4.14: An example of unidimensional data set. Adapted from "Fuzzy C-Means Clustering" [3].

[5]K-Means Clustering involves the partition of n observations into k clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster.

The first clustering presented in Figure 4.15 is an example of the K-Means algorithm where each data point belongs to a specific centroid.

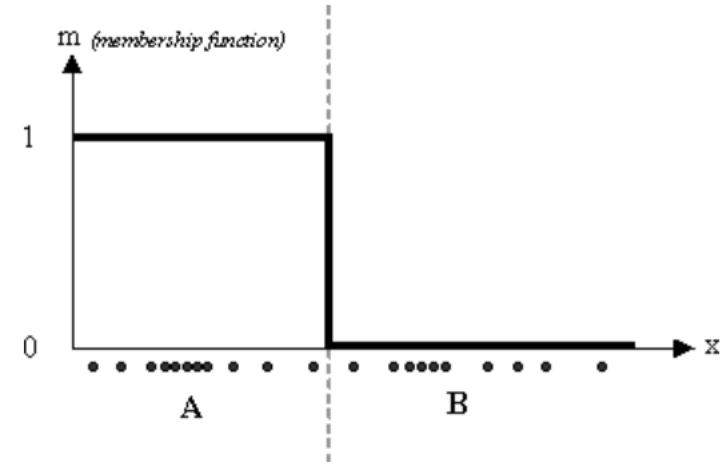


Figure 4.15: A typical membership function used in exclusive clustering methods such as K-Means. Adapted from "Fuzzy C-Means Clustering" [3].

On the contrary, in the Fuzzy C-Means, each data point has a probability of belonging to a specific cluster. This probability is modeled by the smoother line of the membership function. For example, the red dot pointed in Figure 4.16 has a probability of 0.2 to be part of the cluster A for this data set. That way, this point will be part of cluster B, contrary to what would happen in K-Means.

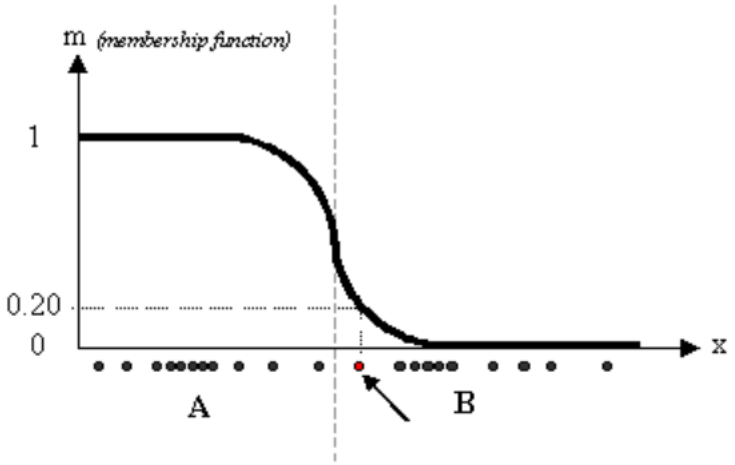


Figure 4.16: A typical membership function used in overlapping clustering methods such as Fuzzy C-Means. Adapted from "Fuzzy C-Means Clustering" [3]

4.6.2.2 Gaussian Mixture Models

Gaussian Mixture Models (GMM) is a clustering method that attempts to find a mixture of multi-dimensional Gaussian probability distributions that best model the input data points. GMM can be classified as a probabilistic method that considers that all data points are created as a mixture of a determined number of Gaussian distributions with unknown parameters [62]. In order to find the parameters of the GMM for each cluster, estimation is performed by training the data by adopting the iterative expectation-maximization (EM) algorithm. The most popular and more highly regarded is maximum likelihood (ML) estimation and was the one used in this work [63]. The main idea of how the GMM is fit into the input data using the EM algorithm can be explained in the following steps:

- The EM algorithm begins with a set of components of means, covariance matrices and mixing proportions and are initialized using a binary Vector quantizer estimation.
- Then for each data point, the algorithm computes the posterior probabilities of component memberships. In other words, a calculation of the expectation of the likelihood function regarding unabsorbed data points is computed.
- The Maximization step of the EM algorithm started by using the component-membership posterior probabilities as weights. The maximum likelihood is applied to estimate the components means, covariance matrix and mixing proportions.

These last two steps are repeated until convergence is met.

Analysing how GMM works it is clear that it is much more flexible concerning cluster covariance. The clusters can assume elliptic shapes rather than restricted circles.

4.6.3 Decision Tree Learning - Classification Tree

For the second approach of evolution assessment, we decided to use the Classification Tree algorithm due to its popularity, easy to interpret and low complexity.

A decision tree is a decision support tool used to visually and explicitly represent decisions and decision making. In machine learning, decision tree learning is a method of approximating the discrete-value target function represented in a decision tree [64]. The idea of a decision tree is to divide the data set into smaller data sets based on the characteristic features until you reach a small enough set that

contains data points that fall under one label. This algorithm is recursive by nature as each node executes a test over a feature, and each descending branch from that node represents a decision rule, with the leaves representing the outcome.

The decision regarding which feature to be considered in each node is calculated using a cost function. The feature chosen is the one where the split costs the least. This makes the root of the decision tree, the feature with the high purity and the most discriminatory. In classification trees, the Gini's Diversity Index is used as the cost function to evaluate the splits in the dataset. It gives an idea of how good a split is by how mixed are the response classes in the groups created by the split - Gini is a measure of purity. The Gini index is calculated by:

$$gdi = 1 - \sum_{i} p^2(i) \tag{4.11}$$

where the sum is over the classes i at the node, and $p(i)$ is the observed fraction of classes with the class i that reach the node [65].

Just like any classifier, the performance can be optimized by reducing the complexity of the tree and thus reducing the overfitting in the classification tree. This is known as pruning. The most used pruning techniques are all of them pre-pruning. They consist in setting parameters for a manual choice of the stopping point. Pruning can be performed by setting the maximum tree depth, the maximum number of terminal nodes, minimum samples for a node split, size of the resultant nodes, among others. Some of them were tested during this study in order to optimize our classifier.

5

Results and Discussion

In this chapter, we presented all the experimental results for the comparative and evolutive analysis of dermoscopic images, and segmentation of pigmented skin lesions based on thresholding segmentation. Furthermore, the entire steps and methodology to achieve these results are described in detail in Chapter 4.

5.1 Statistical Metrics

In order to evaluate the effectiveness of the proposed algorithms, the statistical metrics sensitivity, specificity, accuracy, and precision have been chosen to evaluate the performance of our classification, from segmentation to evolution assessment. Since the segmentation step can be seen as a pixel classification task, the performance can also be measured with a statistical metric. In the case of evolution assessment, the confusion matrix was analyzed in order to understand the disparity between the predictions and ground truth. The definition of the statistical metrics can be observed below:

$$Sensitivity = \frac{TP}{TP + FN} \tag{5.1}$$

$$Specificity = \frac{TN}{FP + TN} \tag{5.2}$$

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} \tag{5.3}$$

$$Precision = \frac{TP}{TP + FP} \tag{5.4}$$

Where the TP (true positive) is the outcome where the model correctly predicts

the positive class. Similarly, a TN (true negative) is the outcome where the model correctly predicts the negative class. The FP (false positive) is the outcome where the model incorrectly predicts the positive class, while FP (false negative) is the outcome where the model incorrectly predicts the negative class.

It is important to note that, from a medical point of view, some metrics can present misleading results when dealing with an unbalanced class dataset. Seeing that, it is known that the accuracy, which measures the rate of correct classified classes, would not be the best metric to evaluate the performance of our algorithms since it can be strongly influenced in favor of the largest class. For that matter, we focus more in both sensitivity and precision metrics since they focus in evaluating the positive class in terms of medical diagnosis such as the boundary of the skin lesion in segmentation or if a significant growth has occurred in evolution assessment.

- The sensitivity, also known as Recall, was used to calculate the fraction of relevant instances that have been retrieved over the total amount of relevant instances.
- The Precision was used to calculate the fraction of relevant instances among the retrieved instances.

5.2 Segmentation

The first step of our study began with an accurate segmentation of our pigmented skin lesions based on thresholding segmentation. Although segmentation does not represent the primary goal of our study, an important detection of pigmented skin lesion area is crucial for a better and correct extraction of features. Segmentation side by side with registration are essential for the performance of lesion evolution.

5.2.1 Ground Truth and Segmentation Metrics

For each dermoscopic images of the initial dataset referred in Section 4.1 a manual segmentation was performed in full-size images displayed in the computer screen. Manual segmentation was performed by us, with some prior clarification by Dr. Raquel Cardoso of the precise location of the lesion boundary. Since it was not performed by a clinical expert (despite the help of the dermatologist), it is almost certain that a residual error will be present when evaluating the precision of the image segmentation.

However, as it is known, dermatologists usually have difficulty at delineating the exact transition between the lesion and the skin mainly due to fading of color between the lesion and the skin. Finally, the ground truth images have the same dimension as their corresponding lesion image, in order to have a direct comparison.

The evaluation of the lesion segmentation was performed using general statistical measurements for binary classification like sensitivity, specificity, accuracy, and precision. However, for the study of segmentation, we select two similarity measurements commonly use in segmentation: Dice Coefficient and Jaccard index. Both of the similarity measurements range from 0 (meaning no overlap) to 1 which means complete congruence. Although these measurements are insensitive to under or over area estimation, they are reactive to misplacement of segmentation labels. For our study, we decided to focus more on Jaccard index since it is numerically more sensitive to mismatch when there is a fairly robust overlap, even though Dice Coefficient is now more common because of the higher results for the same cases of segmentation. If we consider an application to binary data, both coefficients measure the similarity between two finite simple sets and, can be described in equation 5.5 for the Dice Coefficient and 5.6 for Jaccard index.

$$DSC(A,B) = \frac{2|A \cap B|}{|A| + |B|} \tag{5.5}$$

$$J(A,B) = \frac{|A \cap B|}{|A| + |B| - |A \cap B|} \tag{5.6}$$

In the above equations, the A and B are the binary masks of the segmentation algorithm and ground truth respectively.

5.2.2 Quantitative Results

The evaluation of the segmentation was performed by comparing the final segmentation mask with the ground truth binary mask. Table 5.1 gives the overall segmentation result of mean and standard deviation of the statistical measurements described above.

Observing the table 5.1, it can be inferred that the performance of the segmentation algorithm was good. The means of the statistical measurements are above 0.90 except for the Jaccard Index. On the other hand, when looking to the standard deviation, the values can range from a 0.0242 in the specificity to a 0.1148 in

	Mean	Standard Deviation
Sensitivity	0.9084	0.0979
Specificity	0.9865	0.0242
Accuracy	0.9627	0.0354
Precision	0.9303	0.1148
Dice Coefficient	0.9129	0.0943
Jaccard Index	0.8487	0.1056

Table 5.1: Evaluation of segmentation algorithm using the statistical metrics Sensitivity, Specificity, Accuracy, Precision, Dice Coefficient, and Jaccard Index.

precision.

Observing the distribution of Jaccard index values (see Figure 5.1), it is clear that the distribution resembles a normal distribution centered between 0.90 and 0.95. According to the standard normal model, in 95% of the cases, our segmentation algorithm will score value of 0.8487 of Jaccard index. The reason for the algorithm not being successful in the other cases is that the lesions have very low contrasts. It is difficult, even for the human eye to detect the edges of the lesions (Figure 5.2) and thus it was not possible to define them precisely since the colors inside the lesion have little or no difference with skin color surrounding. In other cases, the segmentation was not satisfactory due to the presence of illumination artifacts or for the incorrect acquisition of the pigmented skin lesions where the preprocessing stage was incapable of correcting (see Figure 5.3).

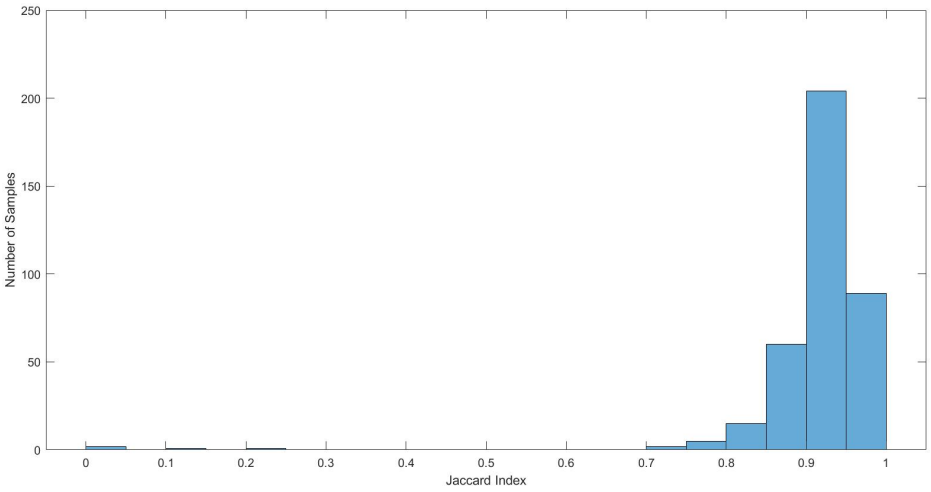
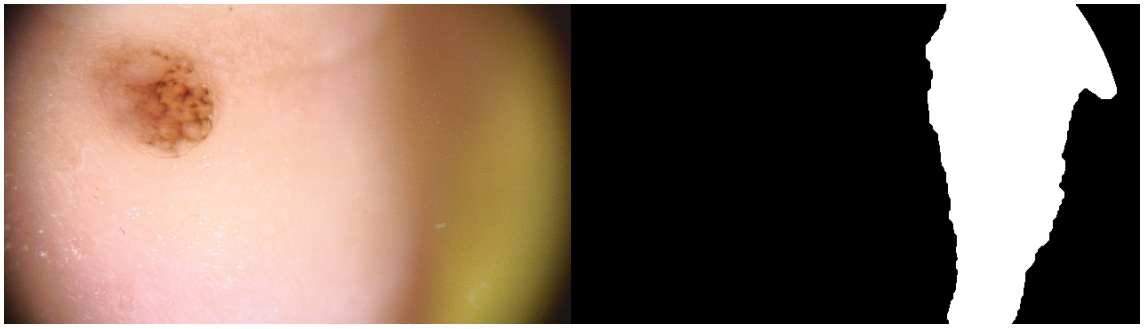


Figure 5.1: Distribution of Jaccard Index values from segmentation results.

Figure 5.2: Dermoscopic Images where the lesion has a low contrast comparing with the healthy skin.



(a) **(b)**

Figure 5.3: Example of non acceptable segmentation. (a) Original dermoscopic image and (b) binary mask resulting from our segmentation.

5.3 Feature Selection

We begin to examine our data distribution (see Figure 5.4) to see if it is possible to observe distinctive groups in the 2D space. By looking at our data distribution it is not possible to observe a distinctive groups between features in the 2D space. However, it is possible to observe two different behaviours between features:

- A possible linear association between the features GGR and LGR, TA Rate and GA Rate and TA Diff and GA Diff. Tha data points clearly concentrated around the identity line (y=x).

- A more dispersed distribution of data points with no clearly visible grouping.

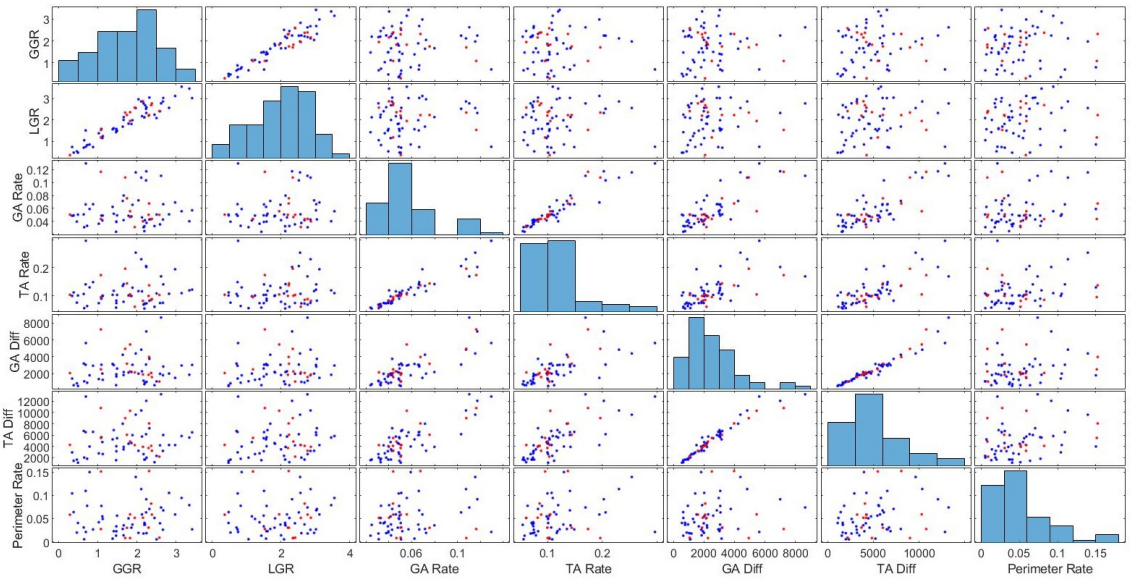


Figure 5.4: Matrix of scatter plots of all features grouped by the labels. The features GGR, LGR, GA Rate, TA Rate, GA Diff and TA Diff stands for Global Growth Rate, Local Growth Rate, Green Area Rate, Total Area Rate, Green Area Difference, Total Area Difference respectively.

To confirm our assumptions we used the Pearson's coefficient as feature selection technique to understand if the data contained some features that were either redudant or irrelevant, in order to be removed and not having a big loss of information as mentioned in Section 4.5. This was a necessary procedure to gain a better understanding of the features and their relationship to the response variables. In Figure 5.5 we depicted the resulting correlation coefficient matrix.

Taking a more in-depth look at the Correlation Coefficient Matrix, it is clear that there are many correlated features, mainly the ones with similar characteristics. Starting with the features extracted from the distance between border points of the lesion, the strong linear correlation could be predictable. By analyzing the mathematical equations of both features (see Equations 4.5 and 4.7) the only difference between them is the denominator. If the lesion border is not well defined, the difference between the two metrics could not be distinguishable, although the segmentation step presents good results. Seeing that, we decided to remove the Global Growth Rate. Features with the same behavior, i.e., with similar extraction formula are those from the area of difference between the alignment of the two binary images. There are features in the area that grew (Green Area, see Figure 4.13 and explain in detail in Subsection 4.5) and the total area. This aspect had already been

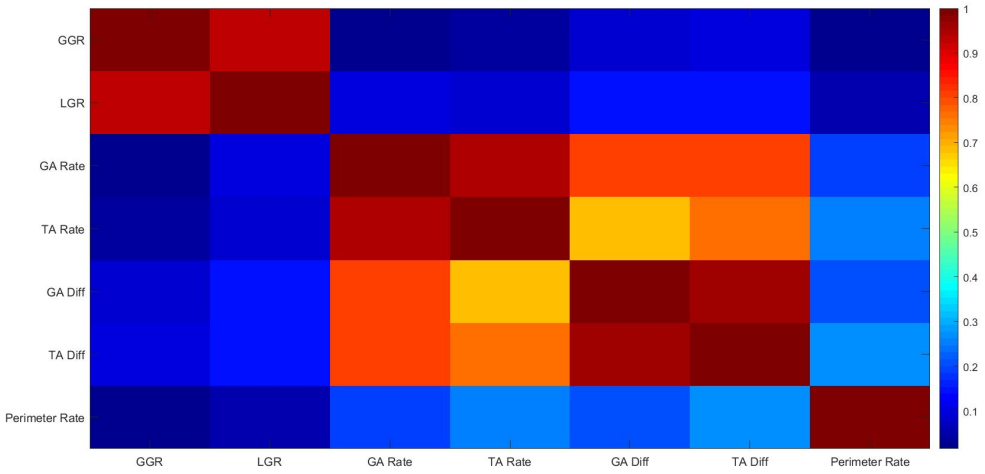


Figure 5.5: Matrix resulted from the calculation of correlation coefficient between all features. The features GGR, LGR, GA Rate, TA Rate, GA Diff and TA Diff stands for Global Growth Rate, Local Growth Rate, Green Area Rate, Total Area Rate, Green Area Difference, Total Area Difference respectively.

mentioned in the Subsection 4.5 that it seemed that both areas were of the same magnitude. The registration parameters were optimized in order to maximize the likelihood between the reference and target images. However, it seems that in some images, the areas in green and red were located in opposite directions, leading to believe that the registration was not correctly performed. Besides, since lesions have different sizes features of the Total Area can influence the performance of the evolution assessment. Seeing that, features related to the Total Area from the alignment of images were removed.

The removal of some of the features was based on the Pearson'a correlation coefficient. However, this metric has some limitations that need to be taken into account. The Pearson's correlation coefficient being a univariate feature selection, shows to be simple to understand and is overall particularly useful for gaining a better understanding of the data. Nevertheless, it does not make it necessarily appropriate for optimization of the feature set. Also, one obvious drawback of Pearson correlation, as a feature ranking mechanism, is that it is only sensitive to a linear relationship. If the relationship is non-linear, Pearson correlation can be close to zero even if there is a 1-1 correspondence between the two variables. In short, the decisions of removing the features based on the Pearson's correlation coefficient may be subject to errors which can only be confirmed in the clustering algorithm.

In the end, the top features used for further analysis were: Local Growth Rate,

Growth Area Rate, Green Area Difference and Perimeter Growth Rate.

5.4 Evolution Assessment

In this study, the primary goal is, through the analysis of evolution in a skin lesion, present a warning sign to the dermatologists and patients if a significant growth has occurred. With the segmentation mask of skin lesions captured at different times, we can benefit from the alignment of the skin lesions to assess when a significant growth has occurred in the skin lesion. However, as it is known, by observing the skin lesions, it is hard to deduce if evolution has occurred and, without having labels to identify, we resort to clustering. The primary goal is to cluster the data into two groups: where a significant growth has occurred and when has not.

5.4.1 Data Labeling

The labeling of data was based on crowdsourcing [61], which consisted of a group of people performing the label of images where a significant growth has occurred. However, the people that participate in this task were not experts in the area which could lead to a biased and erroneous label. As a result, there were skin lesions where people were in doubt. In Figure 5.6 we present some cases where the choice of the label was by minimum "margin". Examples, where the growth was evident, are presented in Figure 5.7.

In order to identify the skin lesion evolution, a comparison is carried out of the lesions' visual properties such as size and shape. Human perception normally relies on this information when the lesion evolves, because of its unique properties. Lesions may have similar appearance over time, so we tended to observe the location of details such as pores and spots within the lesion. This was a way to understand how difficult is to evaluate the lesion growth. Thus, some possible explanations can be given to justify such facts:

- Most of the times dermatologist rely on the digital dermoscopy only for better observation of the lesion details, so there is no major concern of the lesion pose. Thus, in some images, the lesion was in different perspective, difficulting to quantify if the lesion had growth and sometimes even tell if the lesion was the same.

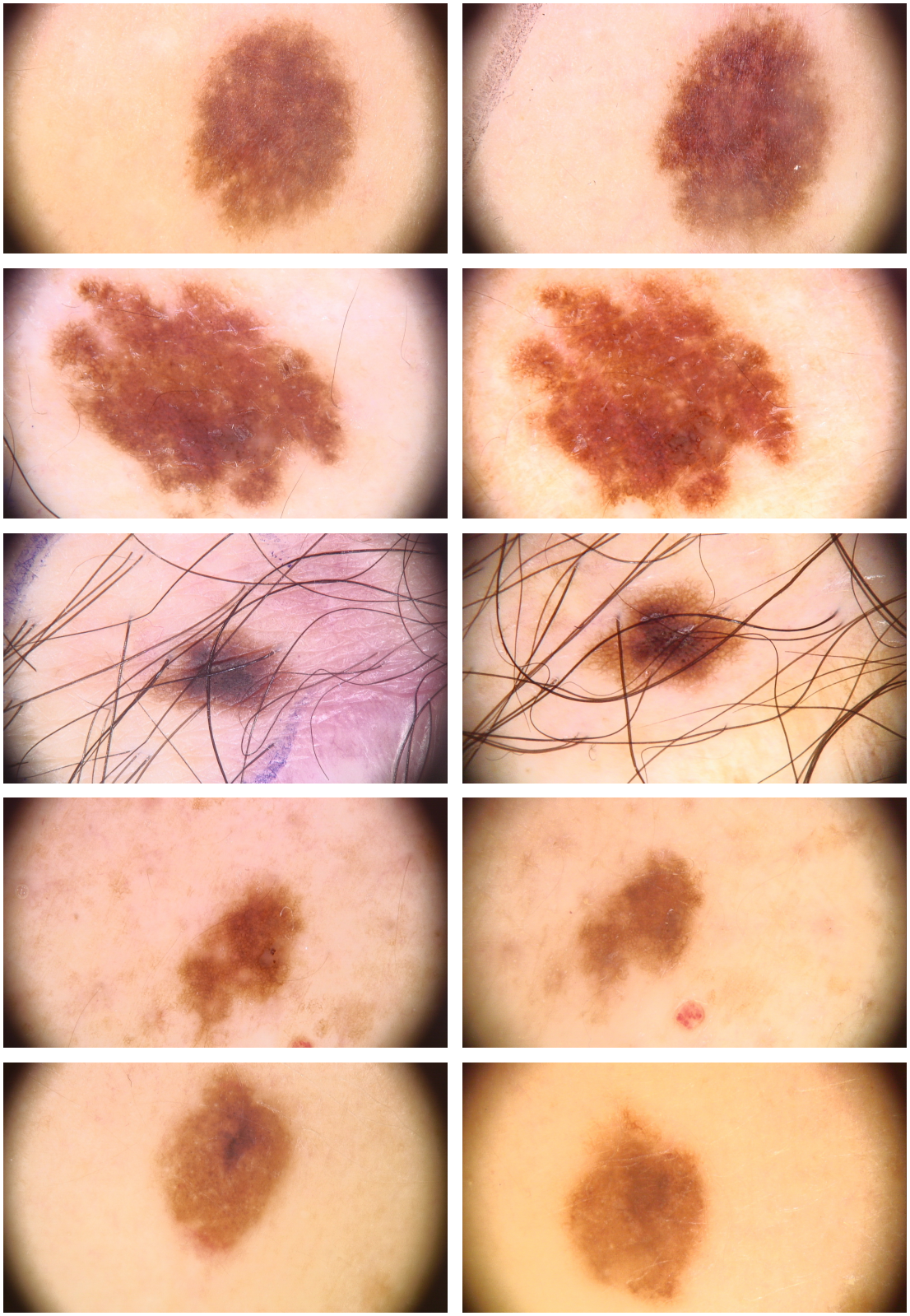


Figure 5.6: Example of skin lesion where the evaluation of growth was difficult and the labels were voted my minimal margin. The images on the left and right side represent the first image and the last image captured chronologically.

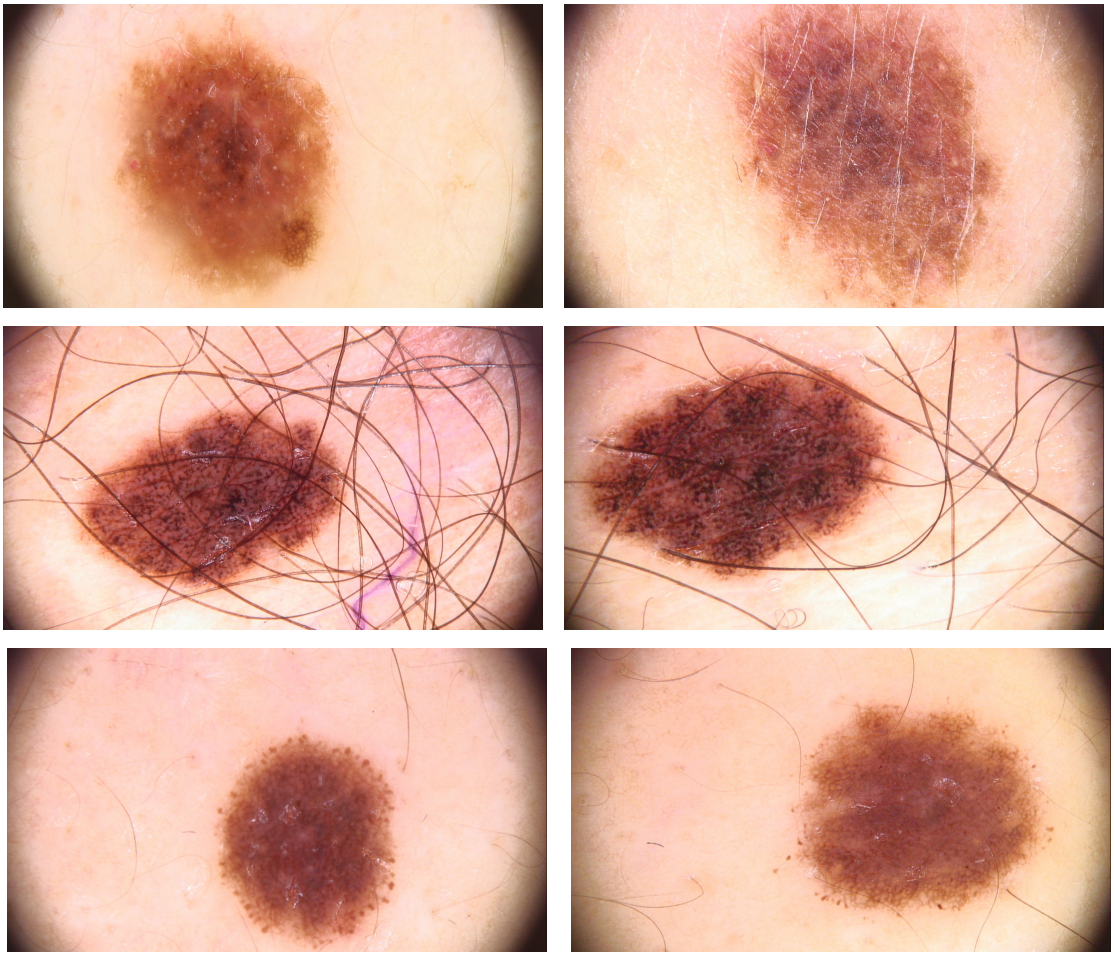


Figure 5.7: Example of skin lesion where the growth was evident and decided by unanimous vote. The images on the left and right side represent the first image and the last image captured chronologically.

- The dermoscope automatically focuses the skin lesion which causes some doubts when assessing the real size of the lesion.

- Since images are acquired during long intervals of time. It is possible that people may have gained weight or lost, causing the skin to stretch or shrink.

- The use of polarized light in dermoscopy is important since it reduces the amount of light reflected off the skin surface allowing the observation of deeper structures and the exact location of the lesion boundary. Although the digital dermoscopy used in this study had a built-in polarized light, it appears that in some lesions this function was not turned on, compromising the decision.

5.4.2 Clustering

We started by understanding which of Clustering Methods better divide the groups in question. As was mentioned in 4.6.2 we resort to Fuzzy C-Means and Gaussian Mixture Models clustering methods to understand the effects to categorize the data points.

However, for this problem in question, some important aspects of machine learning should be referred before taking any conclusions. The dataset used for this problem was extremely small, only having 64 pairs of images, from which about 73% were negative class - no significant growth has occurred - leading to an unbalanced dataset too. Moreover, with a few data points in clustering, the risk to overfit the learning model is far higher, although the concept of overfitting is not usually applied to unsupervised learning algorithms. Nonetheless, in a statistical setting, the finite dataset is assumed to be sampled from the reality. That is, the goal is to approximate the cluster groups to the underlying reality [66]. Moreover, the presence of outliers can become troublesome, although for this problem we examined possible outliers that could identify a significant growth.

Since the features and algorithms are decided, the plan of testing can now be formulated. Two different data sets were used for this problem. These are the full dataset containing all features extracted from the alignment of images and binary masks and, the dataset only containing the top features selected after discarding possible redundant features. For the plan of testing mentioned above, we run the clustering algorithms one hundred times, and afterward, we analyzed the average of the resulting precision, recall, and accuracy. The final results of the different metrics as well as the ROC curves [1] are displayed in Figure 5.12 and, Figures 5.13 and 5.14 respectively. In addition, we display the partition of the data in cluster from Figure 5.8 to 5.11. Note that, the figures present the separation of the cluster with two-to-two combinations of the features in question. In the case of cluster methods using all features, we only displayed the behavior among the top features.

Looking at each cluster separation, we notice that there isn't a notable separation of the data, reached the point where some data points are mixed within the two clusters. Furthermore, it is clear that a reduction of the initial features has none or almost no impact on the clustering division. Since we knew the correct real labels of eight skin lesions, all of them positive, we identify with a different color the ones

[1] ROC curve is a graphical plot that illustrates the diagnostic ability of a binary classifier system as its discrimination threshold is varied

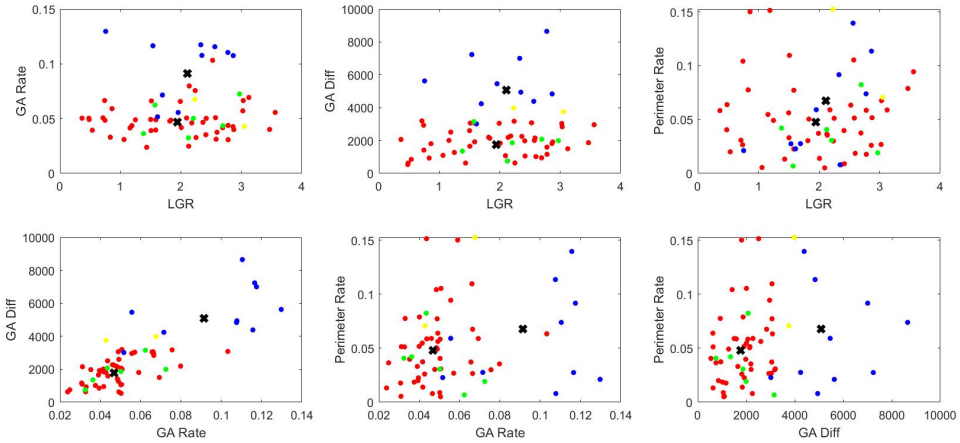


Figure 5.8: Cluster partition using Fuzzy C-Means with All Features. The colors red, blue, yellow and green represent the cluster of lesion that did not evolve, where a significant evolution has occurred, correct real labels and incorrect real labels.

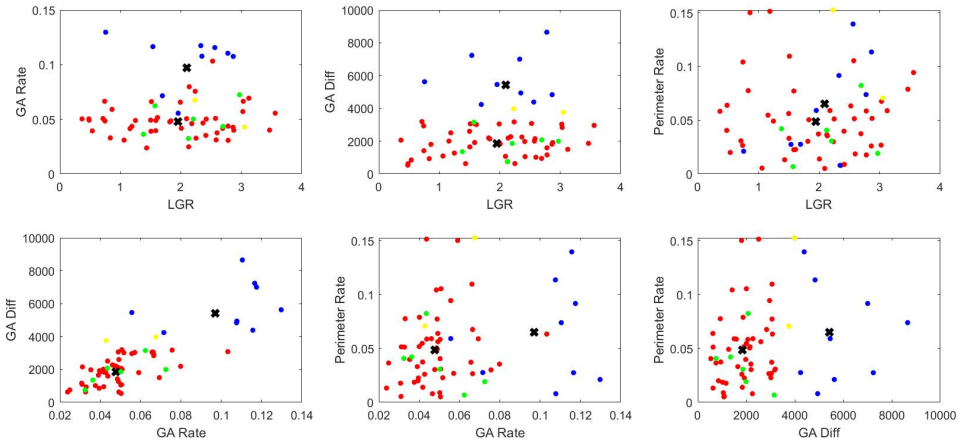


Figure 5.9: Cluster partition using Fuzzy C-Means with Top Features. The colors red, blue, yellow and green represent the cluster of lesion that did not evolve, where a significant evolution as occurred, correct real labels and incorrect real labels.

right and wrong. From the eight lesions labeled by the Dermatologist, only two were corrected identify in the Fuzzy C-Means and three in the Gaussian Mixture Models.

Taking a more in-depth look at overall results in Figure 5.12 it is evident that both clustering algorithms fail do distinguish both groups. The metrics presented barely exceed the barrier of 0.50 except for accuracy. The high number of accuracy compared with the other metrics, especially in GMM can be explained by high specificity of the GMM (0.78 with all features and 0.74 with top features) due to a large number of negatives class.

The results presented in the ROC curves come to confirm our assumptions. The

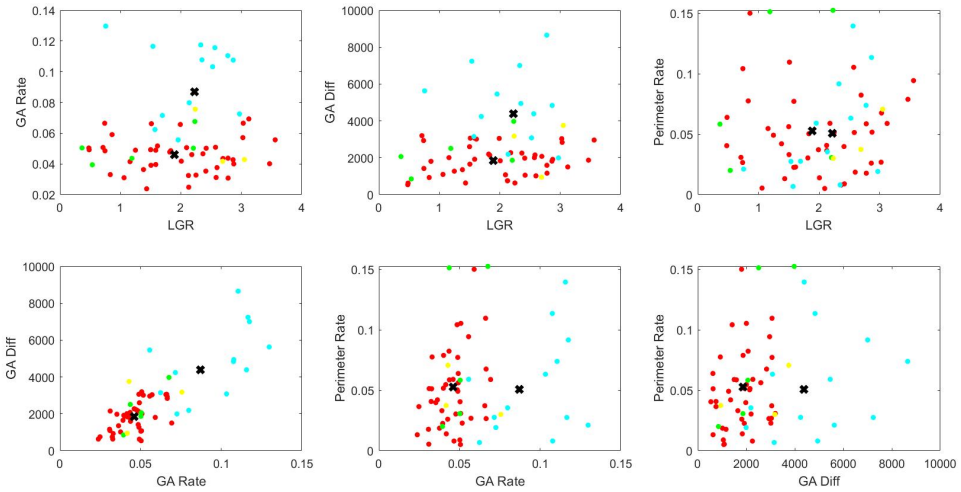


Figure 5.10: Cluster partition using Gaussian Mixture Model with All Features. The colors red, cyan, yellow and green represent the cluster of lesion that did not evolve, where a significant evolution has occurred, correct real labels and incorrect real labels.

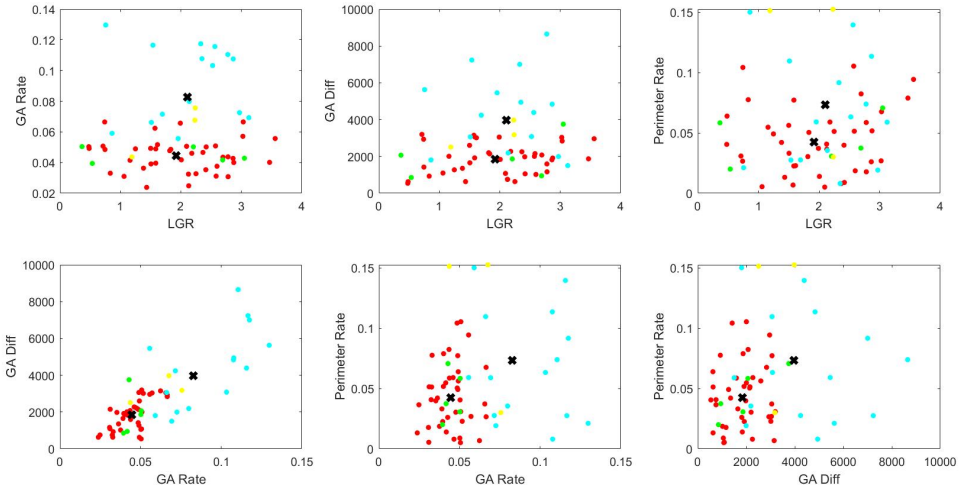


Figure 5.11: Cluster partition using Gaussian Mixture Model with Top Features. The colors red, cyan, yellow and green represent the cluster of lesion that did not evolve, where a significant evolution as occurred, correct real labels and incorrect real labels.

Area under the curve extracted from the ROC curve quantifies the overall ability of the classifier to discriminate between those individuals which have significant growth and those which haven't. Since the values of Under the Curve do not exceed the 0.60, the behavior of the clustering is similar to a Random guess(0.5).

	FCM	GMM
All	2	3
Top	2	3

Table 5.2: Number of skin lesions that the clustering algorithms detect an significant growth.

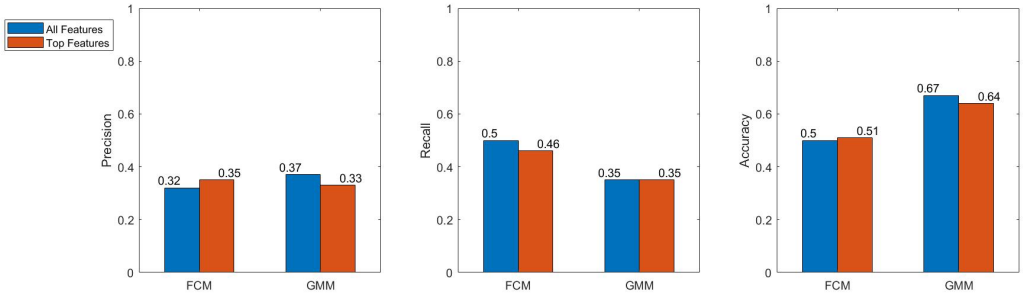


Figure 5.12: Average of Precision, Recall, and Accuracy by Clustering method, built with all features and top features. The results are the average of one hundred runs.

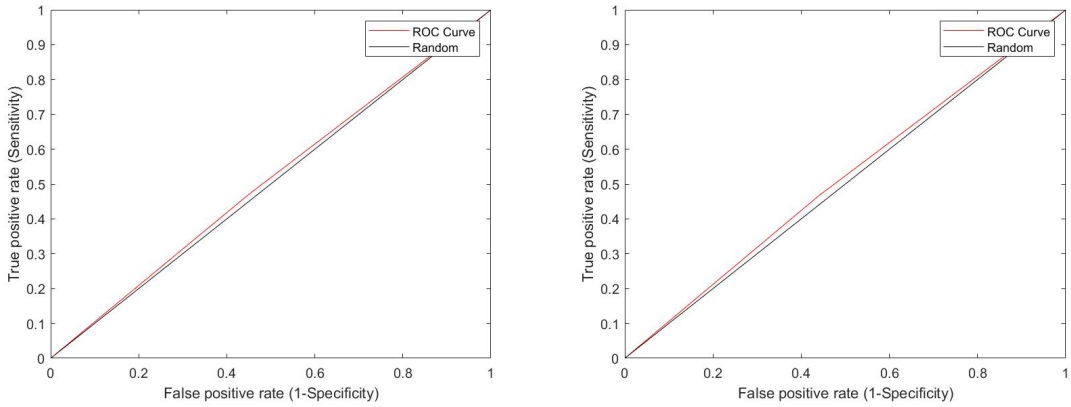


Figure 5.13: Average ROC curve for the dataset with all features (a) and top features (b), with Fuzzy C-Means. The value of AUC was 0.50 and 0.51 respectively.

5.4.3 Decision Tree Learning

Through the application of clustering, we were not able to distinguish the group where a lesion had evolved and where had not. Thus for our second experiment, we decided to use supervised learning to understand if providing the extracted labels, the algorithm can predict the distinctive groups from the input features.

Again, for this second experiment, we use the same dataset which led us to some precautions of how we would test it. Since the dataset was small the risk of over-

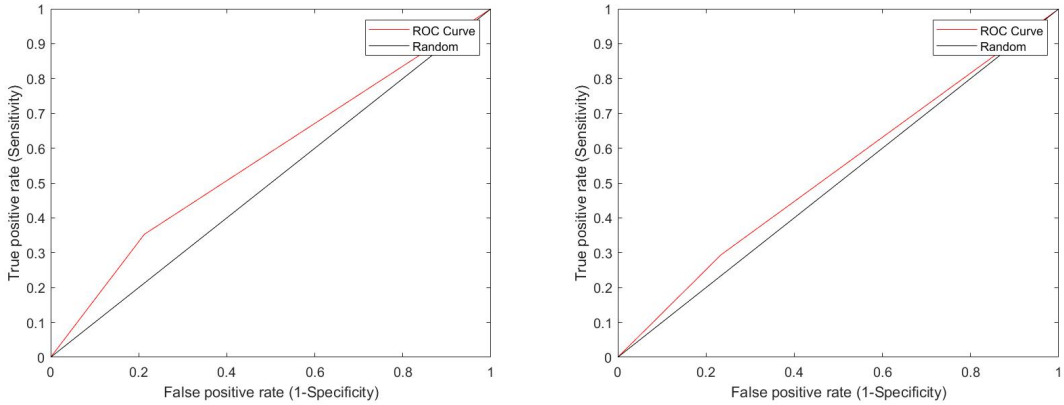


Figure 5.14: Average ROC curve for the dataset with all features (a) and top features (b), with Gaussian Mixture Model. The value of AUC was 0.57 and 0.54 respectively.

fitting was much higher (as already mentioned in Subsection 5.4.2). Thus, to avoid overfitting, a low complex method with few parameters was chosen to test our hypothesis - Classification Tree. Seeing that, we decided to verify our results with a Leave-one-out Cross-Validation. This is a special case of the K-fold cross-validation where each observation is considered as a validation set, and the rest $n-1$ observations are a training set, repeating this for n times for each observation as a validation set. This testing method is appropriate for fewer data since it uses almost the entire data set leading to a reduce bias and reduced over-estimation of the test error. The training set is previously balanced by undersampling the bigger class.

There are many combinations of parameters which can change the performance of the classifier. For Classification Tree the parameters were the following: *SplitCriterion* - as split criterion of nodes; *MinParent* - minimum number of observations of branch node observations; *MinLeaf* - minimum number of leaf node observations; *MaxSplits* - maximal number of decision. The optimized classifier used 'gini' criterion, at least 15 leaf node observation, a maximum of 2 splits with a total of 30 observations. Therefore, for each dataset, we run one hundred times with Leave-One-out cross-validation with their average values being computed. The final results of the different metrics as well as the ROC curves are displayed in Figure 5.15 and, figure 5.16 respectively.

After taking a glimpse of the average results, it is evident that is far from desired. The decision tree has similar results to both clustering algorithms Fuzzy C-Means and Gaussian Mixture Model. Still, the values of Recall with Classification Tree were a little higher. This means that this algorithm was capable of correctly guessing a higher number of lesion that has evolved among the ones present in the target

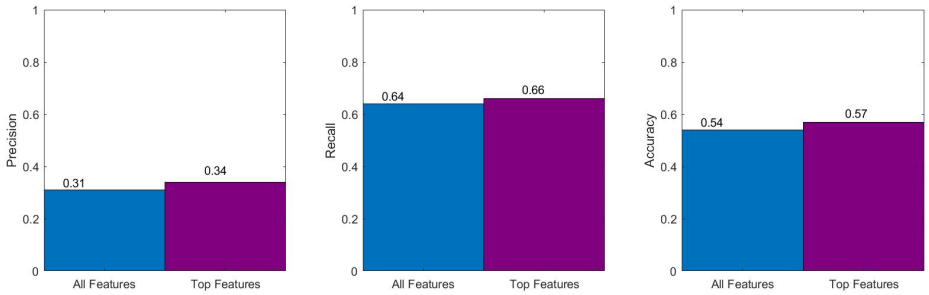


Figure 5.15: Average of Precision, Recall, and Accuracy by Classification Tree method, built with all features and top features. The results are the average of one hundred runs.

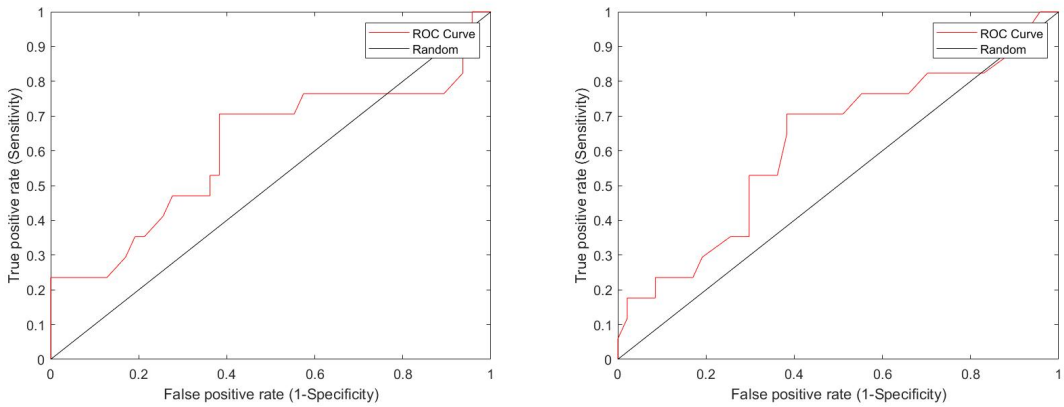


Figure 5.16: Average ROC curve for the dataset with all features (a) and top features (b). The value of AUC was 0.51 and 0.62 respectively.

values comparing with both clustering algorithms. Taking a deeper look at how the classifier responds to the different group of features, the results were the same for all metrics.

5.5 Brief Conclusions

In the context of this dissertation, the objective of detecting groups where a significant growth has occurred were partially achieved.

The influence of a reduction of the number of features was invalid since there was no discrepancy between results from the Top Features and All Features, leading to the conclusion that these were not good features for this problem. By analyzing the dermoscopic images of the skin lesions by the naked eye, we believed that these features were very reasonable for the characterization of evolution thus we only included features from the border and shape. PCA was also considered, but with the

reduction of dimensionality, we would lose interpretability of features, not creating a possible threshold where a significant evolution had occurred. For instance, GMM fails to cluster the data accurately due to a large number of free parameters in the covariance matrix [67].

As far as I am concerned the low precision values are directly dependent on the labels extracted. Having a low precision means that the number of false positives was very high, that is, there were many predictions which both clustering methods and Classification Tree predict a significant evolution when in fact it did not happened. However, it is essential to understand the type of dataset we are dealing with. The dataset consists of dermoscopic images of people who are at risk of developing a malignant lesion and need to be monitored regularly. From the outset, there is a higher probability of having more cases of lesion evolution than the average population. However, the group of people who participate in the extraction of labels are not part of this group, not being accustomed to observe these suspicious lesions on a regular basis. Therefore, when extracting the labels, there are only a few cases where the majority vote in significant growth. Thus we believed if an area expert provided the labels the results would be much different. Also, for lack of time we didn't had the opportunity to ask these labels to the partners of IPO. Probably on the other way around, a higher number of real significant growth would exist and consequently lowering the number of false positives. Therefore, the nature of labels had a significant impact on this study.

6

Conclusions

The thesis development aimed at contributing to the exploration of problems related to the analysis of evolution in a skin lesion by comparing significant features that could describe a significant growth. However, the problem of early detection of malignancies of human skin is still a difficult task, and the monitoring of human skin is fundamental. For that matter, the choice of this study aims at exploring meaningful features that could characterize a significant growth over time.

The dermatologists look for early signs that would indicate a possible malignant evolving of the skin lesion in order to remove it and avoid a possible metastasis of cancer. However, the detection of early signs is strongly dependent on the professional experience of the dermatologist. This was proven during our study when a group of people was asked to annotate if a significant evolution has occurred in the skin lesion. The appearance of the lesions over time and early detection can be misleading.

Thus, the role of computer vision is to provide the necessary tools to help the dermatologist monitoring a possible evolution of the human skin. Therefore, for the aim of this study, a review of the literature was performed to understand the application of computer vision in the early detection of malignant lesions. However, there were few information in the literature regarding evolution detection. Most of state of the art focus on developing CAD for classification of dermoscopic images. Despite all of the efforts made in this area that are already strong and robust CAD, that are achieving performance on par with expert dermatologists [68]. Regarding the analysis of skin lesion evolution few attempts have been made, most of them focusing more on extracting features that could characterize a significant evolution [13] or performing an accurate registration for a better comparison [11]. This gave the motivation to study an automatization of detecting a significant evolution of the skin lesion, although the objective was only partially achieved and there is much work to be done yet.

6.1 Summary of Thesis

In the Introduction we briefly context the Melanoma skin cancer, presenting statistics of its incidence and the importance of regularly monitoring of skin. In particular, we exposed the difficulties that dermatologist experience everyday and also the lack of tools researchers, which lead to an establishment of protocol with the Portuguese Institute of Oncology of Coimbra Francisco Gentil. The pipeline of this study, it was not chosen by chance but carefully determine for a better understanding of the problem. As a result, the framework was based on the literature review that is present in Chapter Two of this work.

In this chapter, we look deep into different approaches implemented for comparative and evolution analysis of skin lesions and present the standard workflows of similar studies. Seeing that, in Chapter Three, the proposed pipeline was as follows: Acquisition of skin lesions, Pre-processing of images, Segmentation of the region of interest of our problem - the skin lesion, Registration of the segmented area for comparison purposes and finally the evolution assessment of lesion. The preprocessing stage and the segmentation were chosen together. The former process was an algorithm based on Otsu's Thresholding and Morphological operations. This process was chosen since it has proven to be a simple and effective algorithm when the preprocessing stage is done correctly. Regarding, the preprocessing stage it was fundamental for correction and enhancement of essential details. Consequently, in order to make a correct evaluation and direct comparisons of spaced time series images registration was needed. Coherent point drift algorithm was preferred due to its robustness to noise, outliers and missing points, fastening and reduction of its complexity computation and finally for its accurate results. The final step of this study consisted in evolution evaluation of registered images concerning the presence of growth rate, local growth rate, perimeter rate, and total area difference. We proposed the adoption of clustering algorithms to distinct a cluster where a significant growth has occurred, based on features extracted from the alignment of the skin lesions and the shape and border of the skin lesion.

Finally, in Chapter Four, we carefully justified and limit to present relevant results from evolution assessment. Overall the results regarding evolution assessment were not high enough to achieve the goal of a complete and automatic evolution assessment.

6.2 Limitations and Future Work

The limitations presented in this study were practically all related to the dataset most of them by lack of information. The first limitation was the lack of datasets containing evolutive images of dermoscopic images. The majority of the dataset available aims at classifying different skin lesions diseases especially Melanoma, only providing single images of each lesion. An effort was necessary to provide useful databases and more information on this subject, mainly for free access.

The second limitation of this dissertation was the inexistence of true labels regarding the lesion evolution. Relying on the decision of people with no experience led to uncertainty and possible errors when assessing the evolution. A better record of dermatologists diagnosis regarding the evolution must be performed. Besides, with a reduced dataset without ground truth labels, we were unable to draw consistent and significative conclusions.

Also, it is essential to define standard conditions for capturing images of lesions and treatments of them in order to give an alert to the dermatologists of when the lesion may have evolved or not. Many dermoscopic images were acquired using different poses, which makes the analysis difficult. Essentially doctors focus on the immediate observation of microscopic details of the dermoscopic images. Therefore, a joint work between dermatologists and computer science researchers must be established.

In addition, the assessment of evolution must rely on other features rather than only border and shape of skin lesions. Although other types of features such as color and texture have been explored for lesion classification, none of them focused on characterizing the change of the lesion over time.

Finally, by addressing the problem of evolution assessment in this dissertation, is safe to say that is indeed a challenging task. Still much work needs to be done, and we hope this dissertation attracts more attention to the problem of assessing the evolution of the skin lesion.

Bibliography

[1] S. S. Mader, *Understanding Human Anatomy and Physiology*. WCB/McGraw-Hill, 2004.

[2] D. N. Zealand, "DermNet New Zealand." https://www.dermnetnz.org.

[3] P. di Milano, "A Tutorial on Clustering Algorithms." http://home.deib.polimi.it/matteucc/Clustering/tutorial_html/index.html.

[4] J.-H. Braun, Ralph Peter and Rabinovitz, Harold S and Oliviero, Margaret and Kopf, Alfred W and Saurat, "Dermoscopy of pigmented skin lesions," *Journal of the American Academy of Dermatology*, vol. 52, pp. 109—-121, 2005.

[5] E. S. C. Foundation, "ABOUT US." http://www.escf-network.eu/en/welcome/about-us.html, 2018.

[6] W. H. O. (WHO), "How common is skin cancer?." http://www.who.int/uv/faq/skincancer/en/index1.html, 2017.

[7] J. M. Gibson, "Review article," *Journal of Advanced Nursing*, vol. 30, no. 3, pp. 773–773, 1999.

[8] N. C. C. N. (NCCN), "NCCN clinical practice guidelines in oncology(TM) - melanoma." https://www.nccn.org/professionals/physician_gls/, 2010.

[9] I. A. f. R. o. C. WHO, "GLOBOCAN 2012: Estimated Cancer Incidence, Mortality and Prevalence Worldwide in 2012." http://globocan.iarc.fr/Pages/fact_sheets_population.aspx, 2012.

[10] B. J. Coventry, D. Baume, and C. Lilly, "Long-term survival in advanced melanoma patients using repeated therapies: Successive immunomodulation improving the odds?," *Cancer Management and Research*, vol. 7, pp. 93–103, 2015.

[11] F. Navarro, M. Escudero-Vinolo, and J. Bescos, "Accurate segmentation and registration of skin lesion images to evaluate lesion change," *IEEE Journal of Biomedical and Health Informatics*, vol. 2194, no. c, pp. 1–1, 2018.

[12] C. N. E. Anagnostopoulos, D. D. Vergados, and P. Mintzias, "Image registration of follow-up examinations in digital dermoscopy," *13th IEEE International Conference on BioInformatics and BioEngineering, IEEE BIBE 2013*, no. December, 2013.

[13] A. I. Mendes, C. Nogueira, J. Pereira, and R. Fonseca-Pinto, "On the geometric modulation of skin lesion growth: A mathematical model for melanoma," *Revista Brasileira de Engenharia Biomedica*, vol. 32, no. 1, pp. 44–54, 2016.

[14] I. Maglogiannis and D. I. Kosmopoulos, "A system for the acquisition of reproducible digital skin lesions images.," *Technology and Health Care*, vol. 11, no. 6, pp. 425–41, 2003.

[15] G. Tortora and B. Derrickson, *Principles of Anatomy and Physiology*. John Wiley & Sons, Inc., 2010.

[16] H. L. Kaufman, *The melanoma book: a complete guide to prevention and treatment*. Gotham, 2005.

[17] A. W. Friedman, Robert J and Rigel, Darrell S and Kopf, "Early detection of malignant melanoma: the role of physician examination and self-examination of the skin," 1985.

[18] D. Griffiths, Christopher and Barker, Jonathan and Bleiker, Tanya and Chalmers, Robert and Creamer, *Rook ' s Textbook of Dermatology*, vol. 1. John Wiley & Sons, 2010.

[19] L. Smith and S. Macneil, "State of the art in non-invasive imaging of cutaneous melanoma," *Skin Research and Technology*, vol. 17, no. 4, pp. 257–269, 2011.

[20] M. E. Celebi, H. A. Kingravi, H. Iyatomi, Y. A. Aslandogan, W. V. Stoecker, R. H. Moss, J. M. Malters, J. M. Grichnik, A. A. Marghoob, H. S. Rabinovitz, and S. W. Menzies, "Border detection in dermoscopy images using statistical region merging," *Skin Research and Technology*, vol. 14, no. 3, pp. 347–353, 2008.

[21] G. K. Hamblin, Michael R and Avci, Pinar and Gupta, *Imaging in dermatology*. Academic Press, 2016.

[22] H. M. Shaw, D. S. Rigel, R. J. Friedman, W. H. Mccarthy, and A. W. Kopf, "Early diagnosis of cutaneous melanoma: revisiting the ABCD criteria," *Jama*, vol. 292, no. 22, pp. 2771–2776, 2015.

[23] A. Sultana, I. Dumitrache, M. Vocurek, and M. Ciuc, "Removal of artifacts from dermatoscopic images," *IEEE International Conference on Communications*, 2014.

[24] T. Lee, V. Ng, R. Gallagher, A. Coldman, and D. McLean, "Dullrazor®: A software approach to hair removal from images," *Computers in Biology and Medicine*, vol. 27, no. 6, pp. 533–543, 1997.

[25] Q. Abbas, M. E. Celebi, and I. F. García, "Hair removal methods: A comparative study for dermoscopy images," *Biomedical Signal Processing and Control*, vol. 6, no. 4, pp. 395–404, 2011.

[26] M. T. B. Toossi, H. R. Pourreza, H. Zare, M. H. Sigari, P. Layegh, and A. Azimi, "An effective hair removal algorithm for dermoscopy images," *Skin Research and Technology*, vol. 19, no. 3, pp. 230–235, 2013.

[27] K. Møllersen, H. M. Kirchesch, T. G. Schopf, and F. Godtliebsen, "Unsupervised segmentation for digital dermoscopic images," *Skin Research and Technology*, vol. 16, no. 4, pp. 401–407, 2010.

[28] A. Kjoelen, M. Thompson, S. Umbaugh, R. Moss, and W. Stoecker, "Performance of AI methods in detecting melanoma," *Engineering in Medicine and Biology Magazine, IEEE*, vol. 14, no. 4, pp. 411–416, 2002.

[29] I. Maglogiannis, "Automated segmentation and registration of dermatological images," *Journal of Mathematical Modelling and Algorithms*, vol. 2, no. 3, pp. 277–294, 2003.

[30] R. Sumithra, M. Suhil, and D. S. Guru, "Segmentation and classification of skin lesions for disease diagnosis," *Procedia Computer Science*, vol. 45, no. C, pp. 76–85, 2015.

[31] M. Zortea, E. Flores, and J. Scharcanski, "A simple weighted thresholding method for the segmentation of pigmented skin lesions in macroscopic images," *Pattern Recognition*, vol. 64, no. November 2016, pp. 92–104, 2017.

[32] J. Pereira, A. Mendes, C. Nogueira, D. Baptista, and R. Fonseca-pinto, "An Adaptive Approach for Skin Lesion Segmentation in Dermoscopy Images Using a Multiscale Local Normalization," vol. 1, pp. 537–545, 2015.

[33] M. E. Barata, Catarina and Marques, Jorge S and Celebi, "Improving dermoscopy image analysis using color constancy," *International Conference on Image Processing(ICIP)*, pp. 3527–3531, 2014.

[34] P. G. Cavalcanti and J. Scharcanski, "Macroscopic Pigmented Skin Lesion Segmentation and Its Influence on Lesion Classification and Diagnosis," *Springer*, vol. 6, 2013.

[35] R. Day, GR and Barbour, "Automated skin lesion screening - a new approach," *Melanoma Research*, vol. 11, pp. 31–35, 2001.

[36] K. Korotkov and R. Garcia, "Computerized analysis of pigmented skin lesions: A review," *Artificial Intelligence in Medicine*, vol. 56, no. 2, pp. 69–90, 2012.

[37] K. Ramlakhan and Y. Shang, "A Mobile Automated Skin Lesion Classification System," *2011 23rd IEEE International Conference on Tools with Artificial Intelligence*, 2011.

[38] M. Silveira, J. C. Nascimento, J. S. Marques, A. R. S. Marçal, T. Mendonça, S. Yamauchi, J. Maeda, and J. Rozeira, "Comparison of Segmentation Methods for Melanoma Diagnosis in Dermoscopy Images," *IEEE Journal of Selected Topics in Signal Processing*, vol. 3, no. 1, pp. 35–45, 2009.

[39] N. Alfed, F. Khelifi, A. Bouridane, and S. Member, "Pigment network - based skin cancer detection," *2015 37th Annual International Conference of the IEEE*, pp. 7214–7217, 2015.

[40] N. Otsu, "A Threshold Selection Method from Gray-Level Histograms," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.

[41] M. I. Rajab, M. S. Woolfson, and S. P. Morgan, "Application of region-based segmentation and neural network edge detection to skin lesions," *Computerized Medical Imaging and Graphics*, vol. 28, no. 1-2, pp. 61–68, 2004.

[42] B. Erkol, R. H. Moss, R. J. Stanley, W. V. Stoecker, and E. Hvatum, "Automatic lesion boundary detection in dermoscopy images using gradient vector flow snakes," *Skin Research and Technology*, vol. 11, pp. 17–26, 2005.

[43] J. Tang, "A multi-direction GVF snake for the segmentation of skin cancer images," *Pattern Recognition*, vol. 42, pp. 1172–1179, 2009.

[44] S. Supot, "Border Detection of Skin Lesion Images Based on Fuzzy C-Means Thresholding," *2009 Third International Conference on Genetic and Evolutionary Computing Border*, pp. 5–8, 2009.

[45] T. F. Chan, B. Y. Sandberg, and L. A. Vese, "Active Contours without Edges for Vector-Valued Images," *Journal of Visual Communication and Image Representation*, vol. 141, pp. 130–141, 2000.

[46] R. B. Oliveira, A. S. Pereira, and J. M. R. Tavares, "Skin lesion computational diagnosis of dermoscopic images: Ensemble models based on input feature manipulation," *Computer Methods and Programs in Biomedicine*, vol. 149, pp. 43–53, 2017.

[47] S. D. O. Cotton, "Developing a predictive model of human skin colouring Developing a predictive model of human skin colouring," *Imaging*, vol. 2708, pp. 814–825, 1996.

[48] R. Oliveira, J. Papa, A. Pereira, and J. Tavares, "Computational methods for pigmented skin lesion classification in images: review and future trends," *Neural Computing and Applications*, vol. 225081445, pp. 1–36, 2016.

[49] K. D. Toennies, *Guide to medical image analysis*. Springer, 2017.

[50] E. Montin, E. Cutrì, G. Spadola, A. Testori, G. Pennati, and L. Mainardi, "Tuning of a deformable image registration procedure for skin component mechanical properties assessment," *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, vol. 2015-Novem, pp. 6305–6308, 2015.

[51] H. Huang and P. Bergstresser, "A New Hybrid Technique for Dermatological Image Registration," *2007 IEEE 7th International Symposium on BioInformatics and BioEngineering*, pp. 1163–1167, 2007.

[52] K. Korotkov, J. Quintana, S. Puig, J. Malvehy, and R. Garcia, "A new total body scanning system for automatic change detection in multiple pigmented skin lesions," *IEEE Transactions on Medical Imaging*, vol. 34, no. 1, pp. 317–338, 2015.

[53] H. Mirzaalian, T. Lee, and G. Hamarneh, "Spatial Normalization of Human Back Images for \nDermatological Studies," *IEEE Journal of Biomedical and Health Informatics (IEEE JBHI)*, vol. 18, no. 4, pp. 1494–1501, 2013.

[54] S. Prigent, X. Descombes, D. Zugaj, L. Petit, P. Martel, and J. Zerubia, "Assessing skin lesion evolution from multispectral image sequences," 2015.

[55] S. O. Skrovseth, T. R. Schopf, K. Thon, M. Zortea, M. Geilhufe, K. Mollersen, H. M. Kirchesch, and F. Godtliebsen, "A computer aided diagnostic system for

malignant melanomas," *2010 3rd International Symposium on Applied Sciences in Biomedical and Communication Technologies, ISABEL 2010*, 2010.

[56] M. Geilhufe and F. Godtliebsen, "Digital monitoring of changes in skin lesions," pp. 229–233, 2010.

[57] D. Sage and M. Unser, "Teaching Image-Processing Programming in Java," *IEEE Signal Processing Magazine*, vol. 20, no. 6, pp. 43–52, 2003.

[58] J. B. A. Maintz and M. a. Viergever, "An Overview of Medical Image Registration Methods (Cited by: 2654)," *Nature*, vol. 12, no. 6, pp. 1–22, 1996.

[59] A. Myronenko and X. Song, "Point-Set Registration: Coherent Point Drift," *IEEE transactions on pattern analysis and machine intelligence*, pp. 1–14, 2009.

[60] M. M. Mukaka, "Statistics corner: A guide to appropriate use of correlation coefficient in medical research," *Malawi Medical Journal*, vol. 24, no. 3, pp. 69–71, 2012.

[61] J. Howe, *Crowdsourcing: How the power of the crowd is driving the future of business*. Random House, 2008.

[62] S. learn, "2.1. Gaussian mixture models." http://scikit-learn.org/stable/modules/mixture.html.

[63] R. Chellappa, A. Veeraraghavan, and N. Ramanathan, *Encyclopedia of Biometrics*. Springer, 2016.

[64] T. G. Dietterich, *Machine learning in ecosystem informatics and sustainability*. McGraw, 2009.

[65] MathWorks, "fitctree." https://www.mathworks.com/help/stats/fitctree.html.

[66] S. Bubeck and U. V. Luxburg, "Overfitting of clustering and how to avoid it," *Preprint*, pp. 1–39, 2007.

[67] A. Krishnamurthy, "High-dimensional clustering with sparse Gaussian mixture models," *Unpublished paper*, pp. 1–8, 2011.

[68] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, 2017.