UNIVERSITAT DE BARCELONA

Final Degree Project

Biomedical Engineering Degree

Evaluation of different segmentation-based approaches for skin disorders from dermoscopic images

Barcelona, 6 June 2023

Author: María Plana Dato

Director/s: Christian Mata

Josep Munuera

Tutor: Roser Sala

ACKNOWLEDGEMENTS

First and foremost, I would like to express my sincere appreciation to my supervisor Christian Mata for his continuous support along the development of the project. Without his assistance and encouragement, this work would not have been possible. His feedback, expertise and dedication has been instrumental in shaping my current knowledge about the topic and in guiding me towards a high-quality assignment. I would like to extend my gratefulness to my second director, Josep Munuera.

My deep gratitude also goes to my project tutor Roser Sala for her guidance and follow-up of the project. I am immensely thankful for her availability and her commitment to its correct fulfilment. She encouraged me to continue my Final Degree Project in my lowest moments during its development.

Last but not least, I would like to thank my family and friends for supporting me throughout the completion of the work. I am sincerely pleased for their unwavering help and understanding. Their constant motivation were the driving forces behind my determination to the fulfilment of the work to the best of my capabilities.

To all those who have contributed directly or indirectly to this commitment, I am absolutely grateful.

ABSTRACT

Skin disorders are the most common type of cancer in the world and the incident has been lately increasing over the past decades. Even with the most complex and advanced technologies, current image acquisition systems do not permit a reliable identification of the skin lesion by visual examination due to the challenging structure of the malignancy. This promotes the need for the implementation of automatic skin lesion segmentation methods in order to assist in physicians' diagnostic when determining the lesion's region and to serve as a preliminary step for the classification of the skin lesion. Accurate and precise segmentation is crucial for a rigorous screening and monitoring of the disease's progression.

For the purpose of the commented concern, the present project aims to accomplish a state-of-the-art review about the most predominant conventional segmentation models for skin lesion segmentation, alongside with a market analysis examination. With the rise of automatic segmentation tools, a wide number of algorithms are currently being used, but many are the drawbacks when employing them for dermatological disorders due to the high-level presence of artefacts in the image acquired.

In light of the above, three segmentation techniques have been selected for the completion of the work: level set method, an algorithm combining GrabCut and k-means methods and an intensity automatic algorithm developed by Hospital Sant Joan de Déu de Barcelona research group. In addition, a validation of their performance is conducted for a further implementation of them in clinical training. The proposals, together with the got outcomes, have been accomplished by means of a publicly available skin lesion image database.

LIST OF FIGURES

LIST OF TABLES

LIST OF EQUATIONS

LIST OF ABBREVIATIONS

AI	Artificial Intelligence
AHE	Adaptive Histogram Equalisation
AK	Actinic Keratosis
ANN	Artificial Neural Network
BCC	Basal Cell Carcinoma
CAD	Computer-Aided Diagnosis
CLAHE	Contrast Limited Adaptive Histogram Equalisation
CNN	Convolutional Neural Network
DF	Dermatofibroma
DSC	Dice similarity coefficient
FDA	Food and Drug Administration
FMM	Fast March Method
FN	False Negative
FP	False Positive
GDPR	General Data Protection Regulation
GMM	Gaussian Mixture Model
GUI	Graphical User Interface
GVF	Gradient Vector Flow
IAD	Interactive Atlas of Dermoscopy
IoU	Intersection-over-Union
ISIC	International Skin Imaging Collaboration
JAAD	Journal of the American Academy of Dermatology
JSC	Jaccard Similarity Coefficient
MDSW	Medical Device Software
MRI	Magnetic Resonance Imaging
PDE	Partial Differential Equation

PH	Pedro Hispano
RCM	Reflectance Confocal Microscopy
ROI	Region of Interest
SAM	Segment Anything Model
SCC	Squamous Cell Carcinoma
SK	Seborrheic Keratosis
SWOT	Strengths, Weaknesses, Opportunities and Threats
TN	True Negative
TP	True Positive
UMCG	University Medical Center Groningen
US	Ultrasound
UV	Ultraviolet
VL	Vascular Lesion
WBS	Work Breakdown Structure

TABLE OF CONTENTS

1. INTRODUCTION

1.1 MOTIVATION

According to global statistics, skins disorders affect between 30% and 70% of people worldwide, and they are one of the most frequent consultation in primary care. There have been described more than 3000 skin diseases that affect population no matter age or ethnic group. Besides, they are considered the fourth leading cause of disability worldwide. They are a heterogeneous group of conditions including chronic diseases, such as psoriasis and eczema, and skin cancer, which may be life-threatening [1].

The Journal of the American Academy of Dermatology (JAAD) [2] reported prevalence, economic impact and mortality for skin disease in the U.S. in 2013 and claimed that 84,5 million Americans were impacted by skin diseases, causing the health care system $75 billion cost.

Diagnosis of dermatological disorders mostly relies on a visual examination by the physician. That can lead to a mis-diagnosis of the case since it depends on the naked eye experience of the dermatologist and its visual perception. Conventional diagnostic approach not only is subjective but also can even be more challenging when the skin disorder is infectious or the individuals suffer from an infectious disease. Examination should then not be performed closed to the patient due to spreading threat. Then, poorly diagnostic accuracy results in inadequate patient outcome such as unsuitable treatment. [3]. Regarding this problem, any helpful advance in skin disorder treatment and early diagnosis is valued.

To provide an early and objective diagnosis, lesions should be identified automatically or semi-automatically from the image since touching the patient is not required and cost-effective approach ratio is greater. However, skin lesion segmentation remains an unsolved challenge due to the variability in colour, texture, and shapes. Although significant research efforts have been made into developing algorithms to reduce image artefacts, such as hair or bubbles, or detect the indistinguishable boundaries of the lesion, segmentation step remains a critical challenge [4]. For this reason, it is of vital importance to dig into this problem and try to find alternative solutions that could assist physicians.

As any medical speciality, dermatology is evolving towards the application of the newest technologies for the most accurate results in diagnosis. The implementation of such innovative solutions to characterise skin lesions has a promising future, so a further digging in the field is a must. Regarding the fact that society is governed basically by the employment of mobile phones, making use of these devices is the easiest way to assist the population. In the current project, various algorithms for segmentation of dermoscopic images are implemented, that could be employed through easily accessible devices such as mobile phones. Thus, we believe that the

ultimate application of the project is the development of apps that can be used on-site to help diagnosis and follow-up of people with skin disorders.

1.2 OBJECTIVES OF THE PROJECT

The current project aims to evaluate the efficiency of three different image processing algorithms for the segmentation of skin lesions by conducting a quantitative analysis of the segmented binary results. The three methods considered in this project are: the level set method, an algorithm combining GrabCut and k-means clustering segmentation methods and an intensity threshold automatic approach developed by Hospital Sant Joan de Déu research group. The first two algorithms depicted are publicly available while the last one is in-house developed by the hospital institution. This work aims to segment melanocytic lesions, both malignant and benign, from 48 images taken from the ISIC-2016 database challenge. The segmented lesions are compared with ground truth images and different statistical metrics are calculated as a means of validating the performance of each method. For the purpose of accomplishing the main goal, the different sub-objectives are:

- To conduct a literature review to analyse current skin disorders present in population and check for available segmentation-based algorithms to evaluate their efficacy for a further implementation in clinical training.
- To study different algorithm options and justify the three segmentation methods chosen to carry out the validation stage.
- To implement and test the pipelines and to use them to automatically and semi-automatically segment images of malignant and benign melanocytic lesions from a dermoscopic public image database, ISIC-2016 dataset challenge.
- To evaluate the performance of the segmentation algorithms by means of a statistical analysis to quantitatively carry out a proper appraisal of the approaches considered.
- To generate a plan for a time-to-come clinical application of the research developed as a means of leveraging the entirely documentation tracked down and reviewed.
- To propose future improvements of the project and the research work executed.

1.3 LOCATION OF THE PROJECT

The origin of this project was born from the supervisors Dr. Christian Mata and Dr. Josep Munuera, members of the Hospital Sant Joan de Déu research team. The idea conceived consisted of assessing the operation of three different segmentation pipelines for skin lesion segmentation. The images used for the purpose of the project were firstly thought to be taken from the paediatric hospital's database. Nonetheless, the research group proposed an alternative solution which gives consideration to the use of a public dataset. This decision was settled on because of the major difficulties it takes nowadays to work with patients information due to legal aspects and data privacy. The aim of the work is then to determine how efficient and reliable are the methods

proposed for the segmentation of the ISIC-2016 database (https://challenge.isic-archive.com/data/#2016) to subsequently use the apprenticeship got from the research project as a tool for the clinicians to easier carry out their diagnosis.

1.4 SCOPE AND LIMITATIONS

This report presents three different methods for automatic and semi-automatic skin lesion segmentation. The database utilised is the so-called ISIC-2016, which consists of a set of two types of skin lesions: malignant, known as melanomas, and benign, noncancerous and often harmless abnormal growths of the skin. The algorithms considered consist of a segmentation stage which differ from one other. Level set method and the one combining GrabCut and k-means are publicly available algorithms. Regarding intensity threshold segmentation, the method has been applied by means of a medical MATLAB App developed by Hospital Sant Joan de Déu research group. The application consists of a graphical user interface which enables the generation of segmented images throughout the implementation of different algorithm options. A later quantitative study of the outcome reached is performed in order to evaluate how the performance of each method is regarding the lesion segmentation. It is highly important to mention that the scope of the project does not consider machine learning-based methods.

The project has been accomplished in Hospital Sant Joan de Déu and it took 17 weeks; the first conception was started during first week of February 2023 and it is expected to be finished by the beginning of June 2023. The timeframe comprises the completion of the project, including the first conception of the topic, the conduction of a previous literature review on the subject, the assessment of the segmentation methods after their implementation and the writing of the final degree project.

In order to complete the project satisfactorily, the different limitations that could be presented during the development of the project must be set out. One of the major constraints of the project is, without a doubt, the time. This factor is a really valuable resource because of the existence of deadlines. In the present case, not only the academic deadline had to be satisfied but one has to bear in mind the need to deal with project limitations, which delay the compliance of the delivery term. So it is necessary to highlight the importance of an appropriate organisation and definition of what it is aimed to achieve.

Segmentation, considered to be one of the most critical tasks in medical image, involves the extraction of a given object from the regions of no interest. Humans are able to rapidly split images into distinct region areas of interest just by simply taking a look at them. This cannot be achieved as easily as a human does when requested to an algorithm since it fails to reproduce the human visual system. Firstly, image contrast, quality and noise may differ from image to image so the same pipeline can fail to identify the region of interest. Poor image quality has a negative impact on the segmentation outcome, so a reliable acquisition device must be contemplated. In addition,

some skin images can present artifacts, such as hair or bubbles, that must be mitigated to guarantee a successful segmentation. The current lack of standardisation to skin image segmentation may produce distinct results when applying different algorithms even though the image employed is the same. This originates is a challenging task when trying to validate segmentation results among studies.

Furthermore, one should be aware of that the images implemented during the progress of the work have been extracted from a public dataset rather than the Hospital Sant Joan de Déu database. The images which the project is going to work over are dermoscopic, high-quality close-up images of the skin captured using a epiluminescence microscope. It has been lately justified the aim of avoiding the clinical data.

Another consideration which ought to be taken into account is the fact that no artificial intelligence tools have been implemented throughout the development of the project. Even though machine learning is currently though of to be the future of medical imaging segmentation, conventional approaches should also be contemplated since they have demonstrated over the past decades high precision and accurate rates although their computational complexity may resemble simpler. Complexity not always means outperformance results. In addition, the main purpose of the work aims to evaluate the segmentation results of different techniques, so the implementation of more than one artificial intelligence may result in a more prolonged research, which confronts one of the main limitations of the project, the time. Yet another reason not to make usage of artificial intelligence has been the restriction of the project only to the segmentation step but not the classification of the lesion whether it is benign or malignant. Many machine learning tools combine in their pipeline both steps at the same time, which for the current case is unnecessary.

2. BACKGROUND

2.1. DERMATOLOGIC DISORDERS

Dermatological diseases are significant illnesses affecting people in any age group and ethnic group [3]. Skin and subcutaneous diseases rank as the fourth leading cause of nonfatal disease affecting 30%-70% of individuals. Skin disease is also the major complaint in primary care, with 8%-36% of patients indicating at least one skin grievance. A recent study has proved that this kind of disorders are actually more widespread than we assume since those affected by the disease do not tend to consult a specialist. Furthermore, consistently lack of dermatologists, particularly in rural areas, results in a burden of diagnosis falling on non-specialists such as primary care physicians or novice nurses. The knowledge limitation and need of training in a field of study plenty of conditions, causes an accuracy diagnostic about 24%-70% [5]. People with dermatological disorders have higher risk for mental illnesses such as anxiety or depression [3]. Regarding the prevalence of the different skin disorders sub-types, the most common diseases are skin infections (8.9%), acne (5.4%), and atopic dermatitis or eczema (5.5%) [6].

Skin cancer is characterized by an abnormal and uncontrolled growth of cells from skin and can be categorised as benign or malignant. It is considered the most common cancer and it is estimated to have 96,480 new cases and 7,230 deaths in 2019 only in the U.S. There are two main types of skin cancer, melanoma and non-melanoma (basal cell, squamous cell, merkel cell carcinomas, etc). Melanoma only accounts for about 2% of malignant skin cancer [7] but causes most deaths so can be fatal if untreated; These tumours develop changes in skin colour and texture but, when treated, they can be cured in more than 90% of cases.

Psoriasis is a common chronic, recurrent inflammatory skin disease that is characterized by sharp and symmetrical erythematous plaques. Psoriasis is found worldwide; in the U.S., approximately 2% of the population is affected and it can appear at any age, suggesting that ethnic groups, environmental factors and genetic background affect the progression of the disease. For instance, the prevalence of psoriasis is low in certain ethnic groups as the Japanese, and may not be present in Australians and Indians from South America [8].

Atopic dermatitis, commonly known as eczema, is a chronic and remitting skin disease affecting approximately 17.8 million people in the United States [9] and it involves patients from infancy to their adulthood. It can lead to significant morbidity. Atopic dermatitis is caused by a complex interaction of gene mutations, dysregulation of immune system and environmental factors that concern the epidermis causing an intense itching skin lesion condition [10].

Skin lesions can be classified into two broad categories based on their cellular origin: melanocytic and non-melanocytic lesions. Melanocytic lesions can be either benign, the so-called nevus, or malignant, melanomas. The latter is a skin cancer produced by an uncontrolled growth of cells that

accumulate pigment (melanocytes). It can be widely spread on the body through the lymphatic and the blood vessels, giving it its main characteristic of mortality [11]. Non-melanocytic lesions is the term used for describing changes in the epidermis, outer layer of the skin, due to different factors. They tend to be benign, not cancerous, but, in some cases, they can evolute to a cancerous stage and become lethal if untreated. *Figure 1* depicts a brief classification about skin lesions based on their cellular origin.

SKIN LESIONS

MELANOCYTIC

BENIGN
- Congenital nevus (mole)
- Spitz nevus
- Atypical nevus
- Lentigo
- Halo nevus

MALIGNANT
- Melanoma

NON-MELANOCYTIC

BENIGN
- Dermatofibroma
- Seborrheic keratosis
- Cherry angioma

MALIGNANT
- Basal cell carcinoma
- Squamous cell carcinoma
- Merkel cell carcinoma
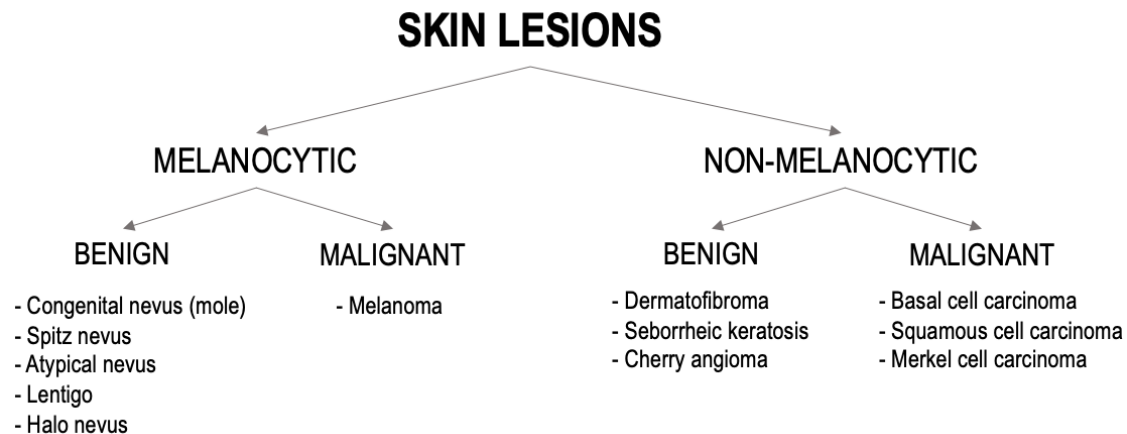
Figure 1. Skin lesions classification based on their cellular origin [own source].

The research of this work is centred on the application of a set of medical segmentation pipelines in order to obtain a binary mask of skin disorders. As to be more specific, the lesions considered regards melanocytic since those alterations are the ones that worry the most and cause the highest morbidity rates all around the world. In the coming section, an intensive exploration regarding clinical melanocytic lesions is performed in addition to a brief introduction to the non-melanocytic ones in order to extend knowledge.

2.1.1 MELANOCYTIC LESIONS

2.1.1.1 MELANOMA

Melanoma, the most serious type of skin cancer, consists of an uncontrolled growth of melanocytes, the cells that produce melanin, the pigment that gives your skin its color. The disease could be asymptomatic for the patient, especially in its early stages, so it is of vital importance to go through dermatological inspection when suspecting. It is a very deadly disease accounting for 75% of skin cancer deaths even though it only makes up 5% of skin cancer cases. In 2019, it were expected 96,480 new cases of melanoma, and 7,230 people died in the United States alone [12]. Melanocytes uncontrolled spread and the asymptomatic nature of the disease makes it the most lethal skin cancer. The determining factors are given by continuous exposure to UV radiation, global warming or genetic predisposition. An early diagnosis of the disease is crucial due to the lack of treatments in advanced stages of the disorder [11].

Malignant melanoma can develop from benign nevi or de novo. Sun rays are the most significant risk factor for melanoma since it directly damages melanocytes genetic information. Nevi carry, at high frequency, a UV-induced mutation of BRAF, a human gene that encodes a protein in charge of directing cell growth called B-Raf. This provides evidence of the etiological factor behind the disease. Melanoma can also be developed for non UV-exposed skin, so the characteristic BRAF mutation is not found in those cases. In addition to genetic mutations, cellular changes can promote the pathogenesis of melanoma; melanocytes may acquire the ability to avoid immune system or they can even secrete growth factors or other signaling molecules that promote cells and blood vessels growth and survival [13].

New methods of non-invasive diagnosis have recently been developed in order to detect early melanocytic lesions to reduce unnecessary biopsies. Early diagnosis and treatment are important for reducing morbidity and mortality. It currently involves a visual examination and a biopsy of the suspected skin region. Then, the treatment options for melanoma depend on diverse factors such as the stage of the cancer, the location and the size of the tumour, the health of the patient itself, etc. The medical procedures utilised include surgery, the main treatment for melanoma, radiotherapy, immunotherapy, and chemotherapy. Treatment decisions are made by a multidisciplinary team of dermatological professionals based on the patient's condition and specific needs. [14] *Figure 2* depicts some examples of melanomas.
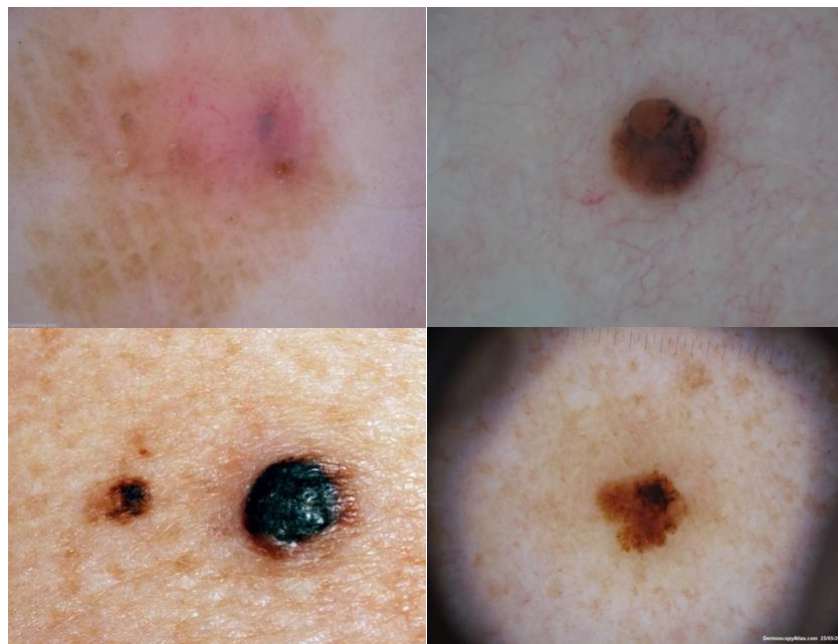
Figure 2. Examples of melanomas dermoscopy images. They vary in colour, shape and texture. The images are publicly available in [15].

2.1.1.2 NEVUS

Nevus is a type of benign melanocytic tumour, containing nevus cells, that is produced due to pigment-producing skin cells. The high concentration of melanin, pigment agent, is responsible for the dark colour. Most of the moles appear during the first two decades of the patient's life, with a

prevalence incidence of one in every 100 babies being born with nevi. Benign moles are usually brown, pink or black and their size varies among small circular or oval ones. The cause leading to the accumulation of melanocytes is not clearly understood, but is thought to be caused by a defect during the development of the embryo. Clinical diagnosis is performed with the naked eye using dermatoscopy. [16].

2.1.1.3 SPITZ NEVUS

Spitz nevus is an uncommon melanocytic mole. This lesion rarely appears in adults, it is more frequently diagnosed in childhood as a benign growth that appears as a dome-shaped lesion on the skin and can clinically be presented either in the classical, reddish pink or the pigmented, brownish black, variant. There can be differentiated three different types of Spitz nevus based on their histology; classical Spitz nevus, the most common type of Spitz nevus, atypical Spitz nevus and spitzoid melanoma. Dermoscopy Spitz nevus images demonstrate that the pigmented variant is much more commonly diagnosed than the classical one. Nevertheless, none of them show clear patterns which may distinguish those benign lesions from melanoma. That is why, even histopathologically, an evident differentiation between benign and potentially malignant Spitz nevus is often challenging. Surgical excision is recommended for clinically atypical Spitz nevus of childhood and for all kind of spitzoid lesions of adulthood [17].

2.1.2 NON-MELANOCYTIC LESIONS

2.1.2.1 BASAL CELL CARCINOMA

Basal cell carcinoma (BCC) is the most common cancer, with approximately 3.6 million cases diagnosed each year [18]. It rarely results in death or metastasis but it can cause significant morbidity. BCC is named after the basal cells that form the lowest layer of the epidermis. The high accessibility to skin and the high prevalence of that cancer have allowed an exhaustive characterisation of the disease. Early basal cell carcinoma is commonly small with some raised areas through which vessels might expose. The tumour grows slowly but, when neglected, it can spread out deeply to cause a huge destruction. BCC can take place at different anatomical locations of the patient. Approximately, 80% occur on the head and neck while the remaining percentage on the trunk and lower limbs [19].

There is a direct link between UV exposure and the development of BCC, however, the precise relation between risk of basal cell carcinoma and the amount, timing, and pattern of exposure to ultraviolet radiation remains unclear. What is clear is that UV exposure damages DNA and induces mutations in some suppressor genes. To diagnose basal cell carcinomas, a skin biopsy, combined with a clinical examination, is performed [18].

Regarding BCC treatment, this can be surgical or non-surgical. Surgical techniques include curettage and cautery, cryosurgery, excision and Mohs' micrographic surgery. An additional

treatment is radiotherapy, which is generally used for elderly people rather than young patients, as the cosmetic outcomes are inferior to those of surgery [19].

2.1.2.2 SQUAMOUS CELL CARCINOMA

Squamous cell carcinoma (SCC) is the second most common non-melanoma skin cancer, characterized by abnormal, accelerated growth of squamous cells. It accounts for 20% of skin cancer and results in 1 million cases in the U.S. yearly, resulting in up to 9,000 deaths. It presents as a red scaly plaque, normally around sun-exposed areas. The risk factors involve UV exposure, older age, fair skin and immunosuppression. Diagnosis is made by a skin biopsy and treatment includes surgical excision [20].

2.1.2.3 SEBORRHEIC KERATOSIS

Seborrheic keratosis (SK) is a common benign skin proliferation of immature keratinocytes that appears as a small roundish reddish to brownish scaling lesion ranging in size from a few mm to many mm and can be presented alone or be one of many lesions. It commonly presents in adult and elderly patients. This lesion does not require treatment. However, it is crucial to be able to differentiate these lesions from other benign and malignant skin disorders. Since it is commonly found on the face, trunk, or extremities, it is infrequently biopsied [21].

Figure 3 exemplifies four different cases of non-melanocytic lesions in order to get familiarized with the field. They can be presented in such different shapes, textures, colours, etc.
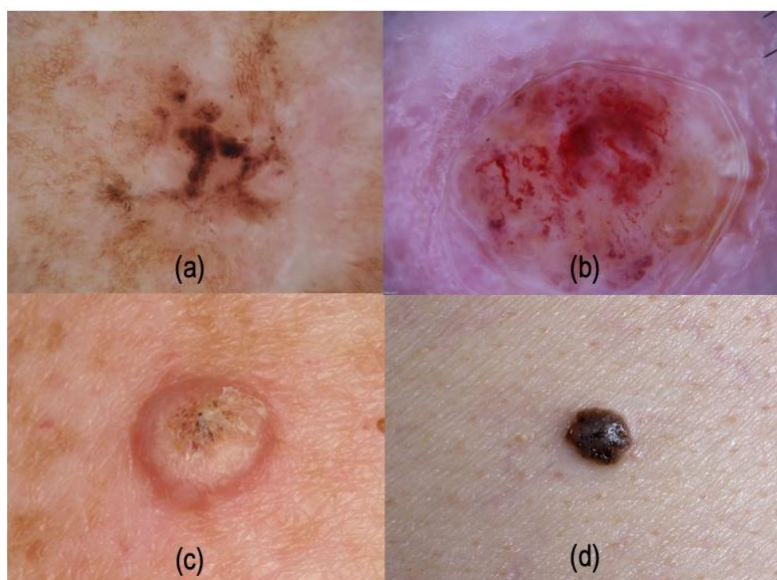
Figure 3. Examples of non-melanocytic lesions. (a) and (b) consist of Basal cell carcinoma, (c) Squamous cell carcinoma and (d) Seborrheic keratosis. Images extracted from [15].

2.2. IMAGE ACQUISITION MODALITIES

Skin lesion images processing requires a previous step: the acquisition of the image. Different approaches are widely implemented as a means to obtain the data such as dermoscopy, digital imaging or confocal reflectance microscopy. Clinical images are usually obtained using digital image cameras. The conditions under which the images are taken are frequently inconsistent, meaning that images are acquired from distances that may vary from one other and under different illumination conditions. Further improvements over skin lesion image acquisition are a must.

Clinical photography has a widespread usage in medicine, ranging from clinical care to teaching aims since the ease of approachability to cameras and documentation provides the opportunity for all fields to use this technology [22]. Photographs should be taken following a certain protocol to ensure high quality results. Camera, lens, lighting, background and photographic technique are among the factors necessary to achieve an accurate image. In addition, ethical and legal aspects should be considered given that clinical photography involves patients privacy [23]. In pediatric dermatology, clinical photographs are taken to judge if a skin lesion has changed or persisted similar along time. This kind of acquisition is not only essential for a precise documentation of physical findings but important in sharing knowledge with other physicians.

Dermoscopy has become one of the most important techniques in the melanoma and pigmented skin lesions diagnosis. It refers to a non-invasive skin imaging methodology that consists of the examination of the skin using an epiluminescent microscopy. It involves optical magnification along with an optic system uncharged of minimizing the reflection of the surface to facilitate visual examination of surface structures. This reduces screening errors and provides greater differentiation between non-affected tissue and skin lesion. Nevertheless, it has been demonstrated that dermoscopy can lower the accuracy of the diagnosis when in the hands of unexperienced physicians. For that reason, the implementation of computerized image analysis techniques are highly recommended in order to minimize diagnostic errors [24].

Reflectance confocal microscopy (RCM) is a high-resolution, non-invasive imaging modality for in vivo skin imaging that uses a near-infrared laser beam of 830 nm. This technique enables the clinician to examine efficiently epidermal and superficial dermal skin lesions, minimizing or avoiding skin biopsy, which may prove to be difficult in pediatric population [25]. RCM is implemented in the diagnosis of both melanoma and nonmelanoma skin tumors, particularly in those challenging cases in which dermoscopy is not so helpful. It can also be used in the interpretation and management of inflammatory skin disorders. RCM has as advantage that it can create images without the use of exogenous contrast agents. The image contrast, however, widely varies between tissues and imaging depths within the same tissue [26].

Current high-resolution transducers allow a detailed ultrasound (US) evaluation of skin tumours. This technique complements clinical visualization, dermoscopy, reflectance confocal microscopy,

surgical planning and follow-up of patients affected by skin disorders. Ultrasounds examination provides a fast and efficient assessment of the tumour diameter and thickness so it is crucial in order to provide physicians additional information of the malignancy [27].

Other techniques such as MRI, spectroscopy or optical coherence microscopy are often used as acquisition systems for skin lesions. Those are, however, underemployed by dermatologists in the assessment of skin tumours since previously reviewed approaches provide greater advantages and more favourable outcomes.

2.3 SEGMENTATION-BASED METHODS

Image segmentation aims to split an image into segments as a means of locating objects and boundaries. In some vision applications, it is pretty useful to separate those regions of the image corresponding to the object of interest from the regions corresponding to background. Image regions are expected to have homogeneous characteristics, such as colour, indicating they belong to the same object. The current different approaches are widely used in pattern recognition, machine vision, face object detection and medical imaging. Some image segmentation techniques like threshold-based, region-based and edge-based segmentation methods are explored next [28].

2.3.1 EDGE-BASED SEGMENTATION

Edge-based segmentation consists of the segmentation of an image by identifying the boundaries of the objects. Boundaries are identified as those locations with strong intensity contrasts, so the approach is applied when there are substantial variations of colour intensity. Edge detection is then based on the detection of sharp changes in the brightness of the image, which is done by different operators such as Prewitt, Sobel, Roberts, Laplacian and Canny [29]. The main purpose of edge detection is to reduce the data amount to be processed while preserving structural information. The Canny operator is the most common used edge detector in image segmentation. It detects the changes in pixels intensity based on the magnitude of the gradient operator while suppressing noise [28].

Canny's edge detection algorithm is a classical and robust method for edge detection in gray-scale images. The algorithmic steps for the Canny edge detector are the following ones: firstly, there is applied a noise reduction technique so the image is convolved with a Gaussian filter. Then, the information about the edges is extracted by computing the gradient of an intensity function of the image, the derivatives in the horizontal and vertical directions. The edges candidates are determined by the suppression of any pixel that is not at the maximum. Hysteresis is then used to eliminate the remaining pixels that have not yet been removed. [30]

Two other examples regarding edge based-segmentation are Sobel edge detection and Prewitt edge detection. Sobel detection is a popular image processing method used to detect edges within an image. The technique employs the concept named convolution, which consists of the application

of specific kernels to highlight those areas of significant intensity changes. The Prewitt operator is another edge detection technique similar to Sobel; Prewitt calculates the gradient magnitude and orientation to identify the boundaries within the image too. However, they mainly differ in the kernel filter employed; while Sobel uses a filter with weighted coefficients emphasizing the central pixel, Prewitt's filter has equal coefficients. [31]

2.3.2 THRESHOLD-BASED SEGMENTATION

Thresholding-based, or intensity-based, method has been commonly used in several skin lesion segmentation methods proposed in literature. The method is the simplest and most effective approach to separate objects from background. The gray levels of the pixels in the image that belong to the object are substantially different from those that belong to the background. Threshold-based segmentation mainly extracts foreground based on the pixel gray value, so it is particularly useful for the segmentation of images with strong contrast between foreground and background. For threshold segmentation of images with low contrast, it is necessary to enhance the contrast of the images first. Pixels are categorized depending on the their range of values; if the pixel value is lower than an certain threshold, it is classified in a category different from those pixels with a value higher than the threshold. The output of the thresholding process is a binary image whose one state means foreground object and can be represented by gray level 1 (white). Otherwise, complementary state corresponds to the background and it is represented by gray level 0 (black) [28]. *Figure 4* illustratively shows how method operates; those pixels with an intensity lower than the threshold T selected are considered background while those higher are the object of interest.
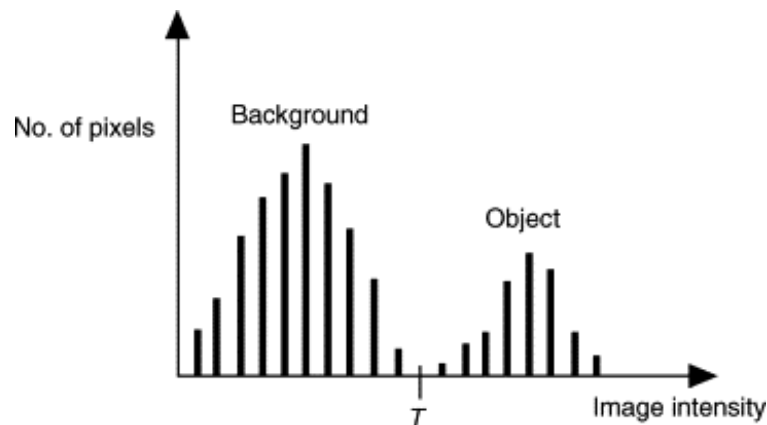
Figure 4. Bimodal histogram with selected threshold T. [32]

Two commonly used threshold segmentation methods are global threshold segmentation and adaptive local threshold segmentation. Global threshold segmentation consists of setting the pixels whose gray level is higher than a given threshold as white and the pixels less than or equal to the threshold value as black. Or inversely, those gray levels greater than threshold are set as black and the remaining ones as white. Nonetheless, in the case of an uneven distribution of gray values, if global threshold is used, the segmentation could not be accurate. The thought strategy is then to

set a certain threshold for the gray value at each position, and the establishment of this threshold is influenced by its neighborhood. This approach is known as adaptive threshold segmentation [33].

2.3.3 REGION-BASED SEGMENTATION

Region-based segmentation algorithms operate iteratively by grouping neighbor pixels which have a similar value into larger homogeneous regions and splitting those different. For example, in a given region of an image, pixels with similar properties, such as grey-level, colour or texture, are grouped together. The algorithm aims to divide the input image into smaller regions recursively. The process finishes when there is no further splitting to be performed. A merging process is done after each split if there is a similarity among pixel's properties. Region-based segmentation methods involve the selection of an initial seed point where the region will begin at. Then, the region grows from this seed point to adjacent points and pixels are merged into larger regions based on certain criteria [28].

Two different examples based on region-based segmentation are region-growing and watershed segmentation. On the one hand, region-growing relies mainly on the assumption that neighboring pixels within a region have similar values. What is consists of is the comparison among neighbour pixels. If a similarity criterion is satisfied, the pixel is set to belong to a certain cluster. For example, difference between pixel's intensity and region's means can be used as a similarity criterion. Watershed, on the other hand, is a region-based technique that employs image topography. It is necessary the selection of at least one seed point interior to each object of the image. The seeds are taken by a user manually or automatically performed. Then, the segmentation process involves flooding the basins and determining the limits where basins merge, which are the boundaries of the image. So, watershed segmentation identifies objects based on the topography of the image while region-growing focuses on similarity criteria [28].

2.3.4 ACTIVE CONTOUR-BASED SEGMENTATION

Active contour-based segmentation, or snakes, is an iterative image segmentation algorithm. An initial curve is specified which will evolve towards the object boundaries that define the region of interest. The method minimizes the energy function from external and internal forces; external forces has the functioning of driving the curve to a desired boundary while the internal forces try to resist that deformation by controlling the degree of shrinkage of the curve. The energy should be defined by a user or through the implementation of an automatic process. Its primary aim in image processing is to define smooth shapes from an uneven distributed image. Active contours are widely used in medical image processing, and specially to delimit object boundaries. Nevertheless, there are still some limitations related to curve initialization and poor convergences [29]. Some active contour-based models examples are such as adaptative snake, gradient vector flow, level set or region-based active contour algorithm.

Gradient vector flow (GVF) is a computer vision framework which main goal is to guide the evolution of a curve towards the desired object boundaries in an image. GVF incorporates gradient information to achieve as accurate as possible segmentation results. The approach is based on Partial Differential Equations (PDEs) in order to obtain the vector field. [34] Adaptive snakes are an extension of the classical snake model with an improved contour's ability to capture boundaries. Then, in adaptative snakes, the idea is to introduce adaptivity into the classical model to intensify its performance. Adaptivity is accomplished by allowing the model's curve to modify size, shape or any characteristic based on local image information. [35]

2.3.5 CLUSTERING-BASED SEGMENTATION

Clustering is an unsupervised study with its application in almost every field of science and engineering. The methodology is based on the process of grouping together data points presenting similarities and label them as a same cluster or group. The characteristics defining the classification widely vary among texture, colours, or intensity. Data items are partitioned into different clusters by keeping in mind two properties; the first one is a high cohesion, which means that data items belonging to a certain group must show similarities. The second property is low coupling, which refers to the fact that data items of a cluster must be different from the data pertaining to other clusters. Clustering can be divided in two different categories: Hard Clustering and Soft Clustering. Hard Clustering consists of the classification of pixels in an exclusive way, that is, if a data item belongs to a certain cluster, it cannot be included in another cluster. In contradistinction, Soft Clustering case considers that pixels can exhibit membership values in more than one cluster. [36]

K-means algorithm is a famous hard clustering algorithm due to its low computational complexity. The algorithm is usually employed in data mining and pattern recognition in the research community. Its main goal is to partition a dataset into k distinct clusters based on feature similarities. The algorithm depends on the value of k, which has to be always specified. That is, depending on the value selected, the segmentation results produced will be different. Then, K-means minimizes the within-cluster sum of squared distances and each pixel is assigned to only one cluster based on its proximity to the centroid. [37] *Figure 5* consists of a visual representation on how k-means algorithm works; the points are considered as similar data in function of their proximity to the centre of the cluster. Fuzzy C-Means algorithm is a soft clustering approach, where each data point is assigned a probability score to belong to a certain cluster. Thus, pixels on the edge of the cluster, with a lower probability grade, are considered in the cluster to a lesser degree than points located in the centre of the cluster. [38] In summary, while both K-means and C-means are clustering approaches, but K-means assigns pixels to a single cluster based on proximity to the cluster while C-means assigns a probability to each pixel across all clusters.
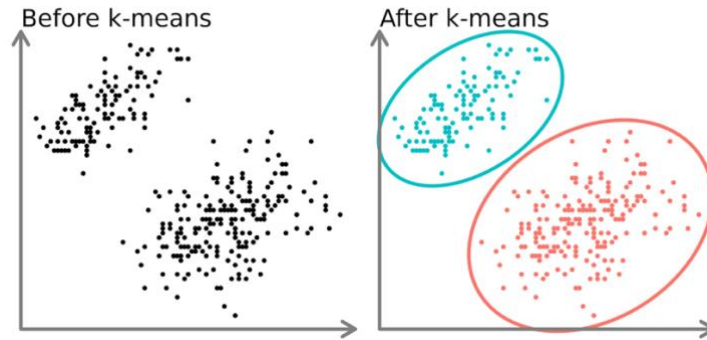
Figure 5. Visual representation of k-means algorithm. The left plot depicts two distinct sets of points unlabelled and coloured as similar data points. Fitting a k-means model to this data reveals two distinct groups, right plot. [39]

2.3.6 AI-BASED SEGMENTATION

Techniques based on artificial intelligence (AI) have currently been widely considered for some image segmentation processes. Convolutional neural networks, evolutionary computation and fuzzy logic are some examples. Artificial intelligence aims to perform tasks similarly to humans based on a previous learning and human reasoning. Those models should be combined among themselves, or other conventional segmentation techniques, as a means of improving the performance of the segmentation [40].

Artificial neural networks (ANNs), computing systems which loosely model brain neurons, have been applied to segment skin lesion images. ANNs are comprised of node layers distributed among an input layer, one or more hidden layers and an output layer. Each node connects to another and it has a certain weight and threshold. If the output of a node is above the threshold value, its data is sent to the next node layer. Otherwise, no information is passed along [40].

Convolutional neural networks (CNNS) are a type of deep neural networks used in deep learning and machine learning primarily for image recognition and processing due to its great ability to identify image patterns. CNN consists of three layers: a convolutional layer, a pooling layer and a fully connected layer. The first layer is the so-called convolutional layer and it is considered the main building block of a CNN. It computes the convolution operation of the input image using filters, parameters of which are to be learned during the training of the algorithm. Then, the pooling layers are added after the convolutional layers in order to downsample the data matrix and reduce the number of parameters to learn. Finally, the fully connected layers, placed at the end of the CNN, are employed to take the end result from the convolution and pooling processes so a classification decision is reached [41].

U-Net is a convolutional neural network (CNN) developed for biomedical image segmentation. It has been revolutionary in performance refinement compared to prior state-of-the-art approaches. It is widely used in image biomedical segmentation tasks, such as brain or liver segmentation, but it is not limited to this single application and the model can still solve more complex problems. U-Net architecture has broadly been used in object detection, image denoising or image registration.

It broadly highlights because of its ability to handle high-resolution images [42]. U-Net is the most widespread image segmentation architecture due to its flexibility, culmination in all biomedical image modalities, and well-adjustment to many challenging real-world applications.

Among other artificial intelligence approaches, many are the existent algorithms developed nowadays. It is a must to mention the role other machine learning architectures play in the medical segmentation field. For example, a similar algorithm to U-Net is SegNet architecture [43], which consists of an encoder and decoder network, followed by a final pixelwise classification layer. Another case is DeepLab [44], a state-of-the-art semantic segmentation model designed and open-sourced by Google that employs dilated convolutions for image segmentation.

But not only AI-algorithms are implemented just to segment images, but to classify them among different categories. AlexNet [45] is a deep convolutional neural network which objective is to push the boundaries of image classification accuracy and demonstrate the powerful tool AI can be. At last, VGG16, also known as ConvNet [46], is a convolution neural net architecture that is used for image recognition. VGG16 purpose is object detection and classification; it is able to classify 1,000 images of 1,000 different categories with a high rate accuracy. It is one of the most popular algorithms for image classification.

2.4 STATE OF THE ART

Segmentation is a challenging and critical process in the workflow of skin lesion analysis. For skin lesions, the task is binary, separating the lesion from the background. Manual delineation is a laborious task that suffers from significant inter- and intra- variability. The development of a fast and efficient segmentation algorithm is thus an indispensable step in computational diagnosis. Automated skin lesion segmentation, however, is impeded by brightness and contrast issues, different skin colours, ambiguous boundaries, artifacts such as hair or bubbles, noise and variability within image acquisition methods, so segmentation gets difficulted. Besides, the lack of datasets containing ground-truth segmentation masks, hinders a reliable evaluation of the model. This has led to various studies being carried out in the development of algorithms for better diagnosis and prediction of skin diseases in the latest times [47].

Next a detailed state-of-the-art is presented in which some algorithms and techniques for performing skin lesion segmentation are discussed. *Table 1* depicts the performance comparison of different algorithms reviewed that have been employed to segment images of skin lesions, which are mostly accomplished in an automatic manner. The table indicates the number of images used and their origin (name of the database), the skin disease studied, the colour spaces and channels used and the pre-processing, segmentation technique and post-processing steps applied.

Ref.	Year	Number of images	Database	Skin disease	Pre-processing steps	Colour space	Segmentation method	Technique	Post-processing steps
[48]	2011	320	Department of Dermatology, Health Waikato New Zealand	Melanoma	Image rescaling + artifact reduction (hair and air bubbles removal)	Grayscale	Active contour-based	Chan–Vese (CV) segmentation	-
[24]	2011	100	DRA Interactive Atlas of Dermoscopy	Non-melanocytic and melanocytic lesions	Colour normalisation and contrast enhancement	RGB and grayscale	AI-based	Neural network	-
[49]	2011	60	Non-specified	Non-melanocytic and melanocytic lesions	-	RGB	Region-based	Iterative stochastic region-merging	-
[50]	2011	85	Melbourne Royal Hospital	Skin lesions	Hair removal + noise reduction + contrast enhancement	XYZ, RGB and grayscale	Thresholding-based	Hybrid border detection	-
[51]	2012	175	Non-specified	Skin cancer tumours	Illumination correction + contrast enhancement + hair removal	JCh	Active contour-based	Non-specified	-
[52]	2012	426	Department of Dermatology, Keio University School of Medicine and University of Naples and Graz	Non-melanocytic and melanocytic lesions	Smoothing + illumination correction	RGB	Thresholding-based	Otsu's thresholding	-
[53]	2013	100	DRA Interactive Atlas of Dermoscopy	Non-melanocytic and melanocytic lesions	-	Grayscale	Active contour-based	Mean shift based gradient vector flow	-
[54]	2013	152	Dermnet dataset	Non-melanocytic and melanocytic lesions	Illumination correction	RGB	Thresholding-based + active contour-based	Otsu's thresholding + Chan-Vese (CV) segmentation	Morphological operations
[55]	2014	68	-	Melanocytic lesions	-	RGB	Active contour-based	Level set method	-
[56]	2016	200	PH^2 dataset	Nevus and melanocytic lesions	Colour enhancement + hair removal	RGB, L*a*b and blue channel	Other	Discrete Wavelet Transform	Boarder Smoothing Region Filling
[57]	2016	200	PH^2 dataset	Skin lesions	Artifact removal + image equalization	RGB	Region-based	ASLM	-
[58]	2018	Non-specified	ISIC-2017 and PH^2 datasets	Skin lesions	Image normalisation + illumination correction + hair removal	RGB	Region-based + Clustering-based	GrabCut algorithm + k-means	Image upscaling to original size
[59]	2020	800	ISIC-2017 and PH^2 datasets	Non-specified	-	RGB	AI-based	Convolutional neural network	-
[60]	2021	1200	ISIC dataset and PH^2 dataset	Skin lesions	Image resizing + hair removal	RGB	Active contour-based	K-mean clustering with optimized firefly algorithm	-
[61]	2022	400	PH^2 dataset and ISIC-2018 dataset	Skin lesions	Hair removal + image enhancement	RGB	Region-based	GrabCut algorithm	-

Table 1. Literature review regarding skin lesion segmentation. A set of papers are depicted within the table and the main characteristics of their segmentation models proposed are described.

The accuracy of the segmentation depends on the model and the procedures used to solve the problem. Thresholding-based methods are widely employed because of its simplicity, good performance and computational efficiency.

Garg *et al.* [60] proposed a model comprised of two main steps which are the pre-processing of the image and the segmentation stage. In the pre-processing stage, the unwanted artifacts such as hair or illumination, are diminished using a threshold and morphological operation. The segmentation of the skin lesion using k-means algorithm with optimized firefly algorithm (FFA) is used to achieve the highest accuracy. The images employed for the segmentation step are taken from two different public datasets; International Skin Imaging Collaboration (ISIC) dataset and dermatology service of Hospital Pedro Hispano (PH$_2$) dataset. The outcomes are measured with respect to different parameters like the specificity, sensitivity, accuracy, dice coefficient, or error. The authors report that the method provides an accuracy of 99.1% using ISIC database and 98.9% using PH$_2$. Moreover, it performs better than existing techniques such as k-means.

Garnavi *et al.* [50] proposes an automatic border detection based on color space analysis and clustering-based histogram thresholding. The methodology is tested on 85 dermoscopy images using a wide set of metrics including accuracy, precision, sensitivity, specificity, and border error. The images are firstly preprocessed in order to remove hair, reduce noise and enhance the contrast. The method developed is comprised of two stages; the first one aims to determine the most effective and discriminative colour channels. The second stage purpose is to enhance the sensitivity of the segmentation: the image undergoes a hybrid clustering-based histogram thresholding algorithm. It is shown to be a highly competitive method and potentially fast computing.

Zhou *et al.* [53] presents a new type of dynamic energy force algorithm combining Gradient Vector Flow (GVF) with a mean shift based on Snakes. The methodology is developed so that when the contour reaches the equilibrium state, the energy forces are balanced. The experimental outcomes achieved from publicly accessible datasets demonstrate that this approach detects the borders of the lesion effectively. The paper then compares the method presented with other approaches such as classical GVF or level sets. The authors finally determine that the mean shift GVF algorithm has a significantly better performance with sensitivity of 86% while the others achieve a sensitivity lower than 81%.

Pennisi *et al.* [57] developed a fully automatic algorithm for the segmentation of skin lesions. It is based on Delaunay Triangulation, which is employed to extract the binary mask of the lesion region of interest. In the first step, the outliers of the image are eliminated by morphological closing and the image contrast is enhanced by equalization. Then, two segmentations are carried out in parallel. One segmentation detects the skin region and filters it while the second segmentation is created by implementing edge detection and Delaunay Triangulation. Both images are then merged to extract the final lesion area. The results given by the authors demonstrate that the approach is

highly accurate when segmenting benign lesions, but this decreases when dealing with malignancies.

Then, even though it is not explicitly depicted in *Table 1*, the beforehand commented in-house application developed is a must to finally introduce. The MATLAB App is called MedicalSeg [62] and it was conceived by Dr. Christian Mata, current supervisor of the project. It consists of a graphical user-interface with the aim of segmenting all type of medical images in order to equip physicians a clinician tool for their respective screenings. A more exhaustive review regarding this tool is further commented.

2.5 STATE OF THE SITUATION

As it can be depicted from the state-of-the-art section, a wide set of articles have put their efforts into developing softwares about skin lesion segmentation. Different methods are employed not only regarding segmentation but classification of the segmented image into a certain category, which could be both benign or malignant. The decision of exploiting a segmentation-based methodology or other may deal with the objectives of the research lines and the milestones themselves.

Regarding the aim of this study and after a laborious an exhaustive search about segmentation methods field, the decision taken by the group established at Hospital Sant Joan de Déu was to consider under study those segmentation approaches which could be reason of comparison because of the computer properties variation under which a pixel is classified as foreground or background. That is, pixels can be labeled according to the information from the gradient of the image, such as edge-based segmentation approaches, or the similarity degree between neighboring pixels intensity values.

The situation in which the project is framed considers that the theme of the study has to involve melanocytic lesions. As exhibited in *Table 1*, most of the reports which have been developed in the last years deal with the segmentation of melanocytic lesions. Both benign an malignant melanoma lesions have lately comprised an enlarged number of cases reported worldwide, so the efforts in evaluating the status of these lesions is of major importance. The identification whether the finding could be harmful for the patient or not must be determined as soon as possible due to high morbidity rates malignant melanomas represent.

The following sections will explain in more detail which are the methodologies aimed to assess their efficacy and sensitivity regarding the segmentation of melanocytic lesion images so the comparative study of the project frame can be successfully fulfilled.

3. MARKET ANALYSIS

3.1 MARKET TARGET

A in depth characterization of the customers is primordially required to define the project. The market segment is composed of those patients suffering from any kind of skin lesion which are aimed to be analysed in order to assess a valid diagnosis. It is necessary to highlight that the ones who are going to make use of the product/service are the physicians, rather than the patients, but the major aiming is to validate the performance of the three different pipelines proposed by means of a quantitative analysis. All in all, the market segment in which the hypothetical software application is framed is undoubtedly robust and even expected to grow globally in the near future due to the increased prevalence of unhealthy lifestyles, among other reasons. The attractiveness of the possible implementation of the project thought of is high.

3.2 MARKET EVOLUTION

Skin lesion analysis is an emerging research field which aims to alleviate the burden and the cost of skin lesion screening. Accurate segmentation plays a crucial role in the examination of dermatological conditions, such as melanocytic lesions. The market for skin lesion segmentation has been witnessing a significant expansion due to the reasons depicted next.

The prevalence of dermatological disorders has lately been rising globally. The development of new segmentation methodologies able to perform an accurate diagnosis of the disease is of vital importance as a means of aiding in timely treatment. Advancements in medical imaging technologies have also made possible acquiring more detailed images of the affected skin region and more sophisticated algorithms solutions have also been launched. In addition, artificial intelligence and machine learning algorithms have shown promising results in this field, but that is going to be more in deep addressed in the following section.

Current researchers have an increased interest in developing computer-aided diagnosis (CAD) systems to provide physicians an efficient tool to reduce the challenges encountered during manual examination [63]. A CAD approach is a system that assists clinicians in the interpretation of medical images by highlighting conspicuous regions in order to support clinical decisions. The primary goal of CAD is to increase the detection of the disease while reducing the false negative ratio and diagnostic errors resulting from the subjectivity of the visual interpretation. CAD techniques comprise steps such as image acquisition, pre-processing, segmentation and feature extraction and classification [64]. The rising demand for CAD systems incorporating skin lesion segmentation systems is revolutionising the market sector.

As it has been reflected in *Section 2.4*, the papers dealing with segmentation algorithms for medical purposes are enormous, and an a considerable amount of articles continue being published yet. The outcomes reported are not always resolved by a single research group since some

partnerships among different healthcare institutions or research organisations have been carried out as a means of promoting the innovation and market growth within this field.

It is important to denoted that the evolution of the market sector is an ongoing process and many further advancements are yet to be developed. Market trends, future disease prevalence, and new emerging applications differ across regions and healthcare systems so the evolution of the field is yet to be seen.

3.3 COMPETITION

As mentioned earlier, dermatological disorders segmentation is a growing field which is expected to provide incalculable remuneration due to its already high demand. That is the reason why a search regarding the viability of launching the hypothetical product is a must in order to avoid the failure of the project. Analysing which are the strengths and the weaknesses of the competitors are essential just to provide the best solution to the possible consumers. In fact, developing any software tool which has been previously that much exploited is a tough task since the ultimate desire is to bring the market a non-innovative solution or an already existing service. For this reason, it has been thought of to perform an exhaustive investigation about those products that are already in the field market and those which are yet to come and are still in the research stage.

Despite, the documentation found focuses more on the diagnosis and the support to the clinician rather to the segmentation procedure itself. This is, the software tools commercialised aim to assess whether the lesion presented in the skin could be potentially harmful for the health of the patient or not. Segmentation, then, takes part of the pipeline for the detection of anomalous conditions.

3.3.1 PRODUCTS ALREADY IN THE MARKET

After the mentioned research, different software tools have been found of to be employing algorithms for skin purposes. Some of the different solutions engineered recently are going to be presented in the following lines.

The company MetaOptima is a fast-growing digital company based in Canadas that endeavours to empower people with the most innovative solutions to lead a healthy life. I was not until 2016 that they developed their most powerful intelligent tool called DermEngine [65], a software for the imaging, documentation and analysis of skin disorders such as melanomas. Its main purpose is to assist clinicians in diagnosing skin problems while evaluating the different treatment options. Some of the features regarding the app include evolution tracker, which allows users to compare images of a lesion taken over time to check whether there has been any change. It can also detect artefacts such as hair and remove it.

The software company named Skin Analytics [66] provides an AI-supported teledermatology service to analyse high quality images of suspected cancer lesions and support dermatologists in

their evaluation. The medical device receives the name DERM (Deep Ensemble for the Recognition of Malignancy), and it is designed to accelerate patient's diagnosis by ensuring that those more prevalent cases are identified as early as possible.

Furthermore, it ought to be mentioned that other existing segmentation tools not developed for medical purposes could be, in fact, adopted for the aim it is presented in this project. For example, the Segment Anything model (SAM) [67] recently developed by Meta AI, which is an approach to build a promptable image segmentation model involving as minimum as possible the human being. The survey and application of these types of models should be deliberated even though they are not highly aimed at any health discipline.

3.3.2 PRODUCTS YET TO REACH THE MARKET

Lots of papers have been recently published aiming to assess the efficacy of their latest application innovation for supporting dermatologists in real-time skin analysis. Some of those reports are highlighted next.

A current project carried out by Francese *et al.* [68] studied the viability of an augmented reality application for the detection of skin lesions. The tool has dermatologists as target and it only provides a decision support. The mobile application affords the analysis of skin lesions by means of exploiting deep learning classification techniques and adopts augmented reality for the visualization of those features which are generally evaluated when assessing a diagnosis.

Another example is the AI-based app development performed by Krohling *et al*. [69]. The app's purpose is to assist those clinicians who have no experience regarding dermatology or do not have any access to a dermoscope. By employing this tool, clinicians may prioritise patients with possible skin cancer. After the acquisition of the image, the application sends it to the server, which performs a convolutional neural network (CNN) model and returns a diagnosis prediction.

3.4 FUTURE MARKET PROSPECTIVE

Skin lesion analysis is a fast-growing investigation area that requires a more in-depth study to help researchers strategize future solutions. The aforementioned field is critical in early detection stage and skin cancer diagnosis. The increasing incidence of melanomas worldwide makes it of vital importance to develop more accurate and efficient methodologies for a proper segmentation and a later valid diagnosis. That is why the future of skin lesion segmentation promises several emerging trends and advancements.

Deep learning has raised as a promising technology in recent years. With the ongoing progression in the field of deep learning, artificial intelligence has made great breakthroughs in the medicine field. Deep learning models, such as convolutional neural networks (CNNs) show brilliant performance in several image-based tasks such as skin lesion segmentation. Before deep learning (DL) models revolution, segmentation was highly based on classical image processing and

machine learning methodologies such as active contours, region-based or unsupervised clustering [70]. These approaches mainly depend on hand-crafted features difficult to engineer. For that reason, they do not always perform as good as wished on complex datasets. In contrast, DL integrates feature extraction and task-specific decision, and actually has a need for larger datasets. Moreover, these models automatically learn complex features of the image in order to enhance the performance of the model and improve the accuracy of the segmentation methodology.

Another recent trend which should be addressed in depth is the use of multi-modal imaging for the segmentation of skin lesions. What multi-modal imaging refers to is the combination and fusing of different imaging techniques, such as dermoscopy, reflectance confocal microscopy and optical coherence tomography, to enhance those characteristics of the image that may be essential in a later diagnosis by the physician. Different imaging techniques capture different lesion aspects: color, texture, depth, etc. By combining the information provided by each modality, the precision of the segmented result is strongly amended so a more complete understanding of the lesion is acquired. The combination of the acquisition approaches can help in overcoming the limitations that each technique causes [71].

Overall, the future of skin lesion segmentation aims to address those limitations that still need to be resolved. This field is likely to continue evolving and improving due to the wider integration of new emerging technologies, for instance, of computer-aided diagnosis systems. Moreover, two other technologies which are aimed to be studied in a near future are augmented and virtual reality. They can further improve the reliability of the segmentation by providing a three-dimensional visualisation of the skin lesion while providing the dermatologist an immersive and interactive experience. These advancements commented are yet in initial development stages but are key in the future revolution of skin cancer screening and amelioration of patient outcome.

4. CONCEPTION ENGINEERING

In this section, it is going to be discussed the different approaches considered for the development of the project itself. The options contemplated for the conception of the work are described and studied so posterior selection of the most appropriate solution can be decided on. So, it can be ensured that the outcome of the project would be as accurate and valuable as it is aimed to be.

4.1 SOLUTIONS CONCEIVED FOR THE PROJECT DEVELOPMENT

4.1.1 POSSIBLE DERMOSCOPIC IMAGE DATASETS

Clinical images employed for the development and validation of the segmentation algorithms are usually those acquired through non-invasive techniques such as dermoscopy, photography, reflectance scanning laser microscopy, ultrasound imaging and spectroscopic imaging. A key point to automatise skin lesion segmentation is the employment of a wide variety and quantity of dermatological disorder images of a given database. Public datasets for skin lesion analysis comprise International Skin Imaging Collaboration (ISIC), PH^2, Interactive Atlas of Dermoscopy (IAD), MED-NODE, etc.

The International Skin Imaging Collaboration (ISIC) dataset includes five different versions for five consecutive years: ISIC-16 [72], ISIC-17 [73], ISIC-18 [74], ISIC-19 [75], and ISIC-20 [76] respectively for the years from 2016 to 2020. ISIC datasets have become the primary repository for researchers for skin lesion analysis, especially in the malignancy assessment. All five years include images for training but ISIC-16 also contains test images and ISIC-2017 validation and test ones. The, ISIC-18 and ISIC-19 not only differentiate nevus and melanoma classes but include images for several classes of lesions. All of the ISIC datasets, both training, validation and testing, contain ground truth images.

ISIC-2016 contains 900 training and 379 testing images but no validation images. It distinguishes benign lesions from melanoma using images with resolutions from 556 × 679 to 2,848 × 4,828 pixels. ISIC-2017 comprises 2,750 images for training, validation and testing with resolutions from 453 × 679 to 4,499 × 6,748 pixels. The ISIC-2018 images are derived from another public dataset named HAM1000 [77]. It comprises 12,500 training images but it does not contain validation or test images. All of them have resolutions of 450 × 600 pixels. Moreover, it comprises seven classes; images are class-wised as nevus, Seborrheic Keratosis (SK), Basal Cell Carcinoma (BCC), Actinic Keratosis (AK), Dermatofibroma (DF), Vascular Lesion (VL), and melanoma. The following year, in ISIC-2019, Squamous Cell Carcinoma (SCC) was added. This year's dataset contains 25,331 images from multiple sources such as HAM10000 or BCN20000 and the images have resolutions of 600 × 450 to 1,024 × 1,024 pixels. Finally, ISIC-2020 contains 33,126 dermoscopic images with the resolutions of 1,024 × 1,024 pixels. Similar to ISIC-16 and ISIC-2017, this dataset only aims to classify rather the image consists of a nevus or melanoma. [63]

The PH$_2$ [78] is a database built up through a joint research collaboration between two different entities, the Dermatology Service of Hospital Pedro Hispano in Matosinhos Portugal, and the *Universidade do Porto*. They are 8-bit RGB colour images with a resolution of 768 x 560 pixels. The image database contains 200 images of melanocytic lesions; 160 are nevus (80 common nevi and 80 atypical nevus) and 40 are melanomas. The PH² database contains medical annotation of all the images, clinical and histological diagnosis and the evaluation of some dermoscopic criteria such as the asymmetry, the presence or absence of streaks, the different colours present in the lesion image, etc.

The IAD [79] is another type of skin image dataset containing 100 clinical cases, each with at least two images of the lesion: a close-up clinical image and a dermoscopic image. It has 700 × 447 pixels of images but does not have segmentation masks. MED-NODE [80] is a dataset consisting of 70 melanoma and 100 nevus images from the digital image archive of Department of Dermatology of the University Medical Center Groningen (UMCG) used for the testing of skin cancer detection system. Finally, SKINL2 [81] database consists of a total of 376 images acquired all of them under similar conditions. It classifies the images using eight different categories according to the type of skin lesion (melanoma, dermatofibroma, psoriasis, hemangioma, etc.). A summary of these datasets presented and their sample distribution are presented in *Table 2*.

DATASETS		NUMBER OF IMAGES FOR VARIOUS TASKS	
		Dermoscopy images	Ground truth masks
ISIC-2016	Training	900	900
	Testing	379	379
ISIC-2017	Training	2000	200
	Validation	150	150
	Testing	600	600
ISIC-2018	Training	12500	12500
ISIC-2019	Training	25331	25331
ISIC-2020	Training	33126	33126
PH$_2$	Training	200	200
IAD	Training	100	100
MED-NODE	Training	170	170
SKINI2	Training	376	376

Table 2. Publicly available datasets of dermoscopy images with variable amount of image samples.

Different images examples from some of the databases talked about beforehand are seen in *Figure 6*. It can be appreciated the similarity among the images within each dataset.

Figure 6. Sample images from the datasets: horizontally Row 1 shows some images from the ISIC-2016 database [72]. Row 2 consists of sample images from PH² [78] dataset and Row 3 from MED-NODE dataset [80].

4.1.2 POSSIBLE PROGRAMMING SOFTWARES

For the segmentation of the skin lesion images, many are the programming softwares available for the performance of the main objective of the project. Next, some options are proposed considering previous skills and knowledge alongside the easiness of the management of the platform.

MATLAB [82] is a programming platform, developed by MathWorks, intended primarily for numeric computing, development of algorithms, data analysis, manipulation and visualisation. MATLAB also provides additional toolboxes which may result interesting for concrete tasks such as neural networks, symbolic computations, image processing, and statistics. Regarding those toolboxes, there are some apps which are a must to mention: Image Segmenter and MedicalSeg. These applications are next commented in *Section 4.1.3*.

Pyhton [83] is a free programming language often used to build softwares, automate tasks and conduct data analysis. Python is a general-purpose language, this is, it can be used to create a wide variety of programs so it is not specialized for any concrete problem. This versatility, along with its easy-to-handle nature and beginner-friendliness, has made Python one of the most-use programming softwares. A huge list of packages and modules for image processing are supported by Python such as *OpenCV*, *PIL*, *SciPy* or *Scikit-Image*. As commented beforehand, these libraries facilitate the processing of large amount of data rapidly and efficiently so data scientists have recently put all their efforts on those kind of tools.

ImageJ [84] is a public domain open source software for image processing. This software allows users to display, annotate, edit, measure, analise, process and save image data by means of a friendly-user interface and effortless to use. It is mainly adopted in many peer-reviewed scientific publications, and in diverse fields such as life sciences, astronomy or physics. In the life sciences, it is used for the processing of medical images to aid in applications ranging from skin analysis to

26

neuroscience. ImageJ contains a plugging for the segmentation of images through an auto-threshold method in order to obtain a binary mask. Furthermore, the image can previously be preprocessed to remove artifacts or reduce noise, and lately analised by some numerical study.

Last of all, another solution thought of was the use of a new AI model developed by Meta AI named Segment Anything [67]. It has been depicted earlier in this project the usefulness of this algorithm to segment any region of interest within an image. It works by simply clicking or interactively selecting different points to include or exclude regions from the object, or by drawing bounding boxes. The model is designed and trained to be promptable.

4.1.3 POSSIBLE SEGMENTATION PIPELINES

After the intensive research work performed and literature review about skin lesion segmentation state-of-the-art, some proposes have been studied in order to implement them in this current project. Each of them deals with different pre-processing steps, programming softwares, segmentation-based methods, etc. The set of options considered must be cautiously examined and reviewed so the outcomes reached are meant to be accurate and crucial conclusions can be extracted.

The first one is Image Segmenter [85], a MATLAB application that provides access to some different ways to segment an input image. It consists of an iterative process in which a variety of segmentation options can be tried. The segmentation techniques employed behave differently depending on the image type so the output outcomes may slightly vary. Image Segmenter allows the creation of a binary mask by the employment of automatic algorithms, such as flood fill, semi-automatic, such as graph cut, or manual methods such as drawing regions of interest.

The second tool is the app named MedicalSeg [62], an in-house developed user interface application for the management of medical images based on their pre-processing and segmentation. It is intended to provide a tool to visualise, assess and compare the results of different segmentation methods according to specific needs, so the generation of a vast ground-truth database can be attained. The interface provides a series of segmentation algorithms such as Canny edge segmentation, intensity segmentation, ThreshMAX percentage segmentation or Otsu segmentation. The application has been developed in order to be installed in the main app panel of MATLAB workstation. *Figure 7* consists of a screenshot of the user's interface in order to visually comprehend how the application is established; it depicts some sections such as the one for uploading the images, the different segmentation algorithm possibilities to make use for, the image processing tools available, such as Gaussian filter or erode filter, etc.
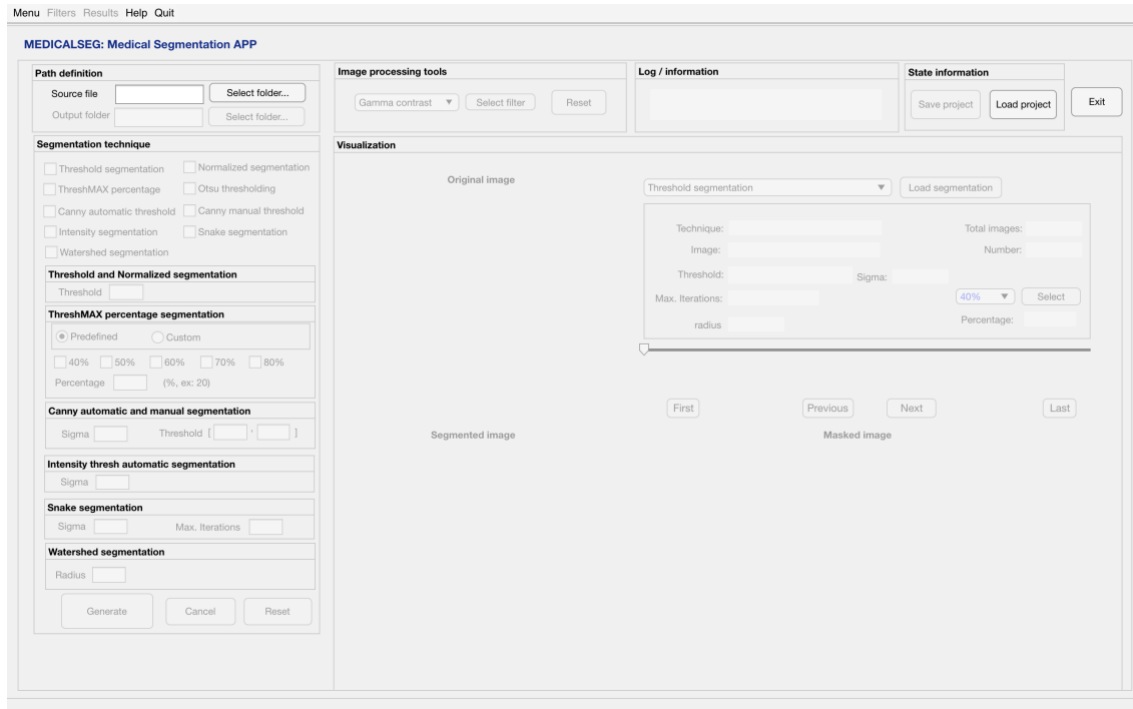
Figure 7. MedicalSeg MATLAB application user's interface [62].

From the different papers cited in *Section 2.4*, some have been the options contemplated in order to validate the main purposes of the present study. Foremost, the method proposed by Khalid *et al.* [8] was thought to be an interesting pipeline to implement, which lies on the basis of wavelet transform along with morphological operations. The authors' method states an accuracy rate of 93.87% on dermoscopic images of PH_2 dataset. Its easiness of implementation and high success performance ratio were considered an appealing reason to develop this approach.

The paper proposed by Ma Z. and Tavares J. [55], defined a computer-aided diagnosis system for skin cancer segmentation of dermoscopy images by the utilisation of the level set method. The authors remark the robustness of the algorithm against the influence of noise, hair and skin textures and highlight the methodology employed as a flexible segmentation tool. The idea behind the method is to represent a boundary as the zero level set of a higher-dimensional function, known as the level set function. *Figure 8* depicts an illustrative representation of the curve evolution.
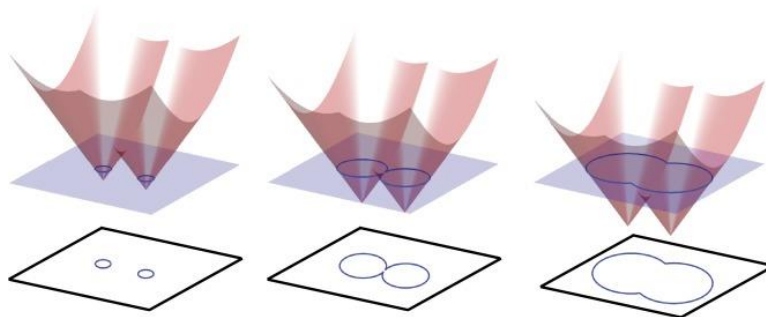
Figure 8. Level set method representation of a boundary, blue line in two dimensions. The level set function, in three dimensions, is depicted in red. [86]

28

The beforehand commented article proposed by Garnavi *et al.* [50] became an appealing choice too since the authors proposed a novel border detection method providing high accuracy, precision and sensitivity rates. Additionally, the method intended to determine the best colour space in which the segmentation could be performed since colour information plays an exceptional role in dermoscopy image analysis. The uniqueness of the color space review in this report, was an indisputable point to highlight when validating the pipelines to select.

Jaisakthi *et* al. [58] proposed an automatic skin lesion segmentation method used as a preliminary stage for the classification of the lesion. The method consists of two major steps; firstly, a preprocessing of the image, and secondly, the segmentation itself. The preprocessing step comprises illumination correction and hair removal by filtering techniques. Then, for the segmentation step, a region-based segmentation tool is employed, the GrabCut algorithm. The k-means clustering is then used to improve the boundaries of the segmented lesion.

4.1.4 POSSIBLE STATISTICAL TOOLS FOR SEGMENTATION EVALUATION

Once the binary masks are obtained after the segmentation processes, it is essential to determine whether or not the results are successful. The main goal is to score which is the similarity between the predicted segmented image and its ground truth. For this reason, the evaluation of the model performance must be carried out. In this section, an overview of the most appropriate evaluation metrics is provided, and their interpretation and implementation are described.

The metrics [87] presented next are all based on the computation of a confusion matrix for a binary segmentation mask, which consists of a number of true positive (TP), false positive (FP), true negative (TN), false negative (FN), with all values ranging between 0 and 1. But, what do all this parameters mean? TP accounts for the number of pixels that haven correctly been classified as region of interest (ROI), FP represents those background pixels misclassified as ROI, TN stands for the number of background pixels properly classified as background and FN consists of those pixels being misclassified as background. Then, some metrics can be derived taking into account this four parameters explained. Those are precision, sensitivity, accuracy and the F-scores.

Precision (*Equation 1*) can be defined as the number of true positive results divided by the number of all positive results while sensitivity (*Equation 2*) stands for the number of true positive results divided by the number of pixels that should have been identified as positive. Accuracy (*Equation 3*) is the number of correct predictions divided by the total number of predictions.

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

Equation 1. Precision formula in terms of TP and FP.

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

Equation 2. Sensitivity formula in terms of TP and FN.

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

Equation 3. Accuracy formula in terms of TP, FP, TN and FN.

F-score is a measure of a test's accuracy. It is mostly used to evaluate binary segmentation systems. The F-score is a way of combining the precision and sensitivity of a prediction. Based on this F-score, there are two popular evaluation metrics in medical image segmentation, the Dice similarity coefficient and the Jaccard similarity coefficient.

The first evaluation metric is the so called Dice similarity coefficient (DSC), *Equation 4*, one of the most widespread scores for performance measuring in medical image segmentation. The Dice coefficient, also known as the Sørensen–Dice index, is a statistical tool that measures the similarity between two datasets, this is, it calculates the overlap between two geometrical objects, defined by the following equation:

$$Dice(A, B) = \frac{2|A \cap B|}{|A| + |B|}$$

Equation 4. Dice coefficient formula.

where A and B are the sets of pixels of those objects which are aimed to be compared. The Dice coefficient ranges between 0 and 1, where 1 corresponds to perfectly overlapping objects while 0 means a configuration where the intersection is zero. This index is commonly used in the field of image segmentation to evaluate how algorithms fulfilled the generation of the binary segmentations.

The second metric is the known Jaccard similarity coefficient (JSC), *Equation 5,* or Intersection-over-Union (IoU), and it is formally defined as the size of the intersection divided by the size of the union of the sample sets. The mathematical representation of this coefficient is written as follows:

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

Equation 5. Jaccard coefficient formula.

where A and B are the sets of pixels of those objects which are aimed to be compared. The index also ranges from 0 to 1, where a value of 1 indicates that both sets are identical or are completely overlapped while 0 indicates that the objects have no common elements, this is, no intersection between A and B is present.

Taking into account the parameters first described, the Dice index and the Jaccard index, *Equation 6* and *Equation 7* respectively*,* can also be described in terms of TP, FP, TN, FN. The expressions are depicted next.

$$Dice = \frac{2TP}{2TP + FP + FN}$$

Equation 6. Dice coefficient in terms of TP, FP and FN.

$$Jaccard = \frac{TP}{TP + FP + FN}$$

Equation 7. Jaccard coefficient in terms of TP, FP and FN.

Both DSC and JSC are often used to measure the similarity or dissimilarity between two sets of data, particularly when dealing with binary data. Nonetheless, the difference between the two metrics is that the Jaccard index penalizes under- and over-segmentation far more than the Dice index.

Finally, the last statistical metric considered for the evaluation of the following project is the Average Hausdorff distance [88], a widely used performance measure to calculate the distance between two point sets. In medical image segmentation, it is employed to compare ground truth images with the segmented results. This metric is specifically recommended for segmentation tasks with complex boundaries and small thin segments. Compared to other measures such as Dice index, average Hausdorff distance has the advantage that takes voxel localization into account.

In the field of medical segmentation image, much more metrics exist and can be applied in order to ensure the validity of the segmented results. Nevertheless, this work is focused on the most suitable metrics to establish a standardised segmentation evaluation and to increase its reproducibility.

4.2 PROPOSED SOLUTION

The above mentioned options were meticulously analised and discussed from the point of view of the necessities of the current project, so the most suited candidates have been selected for the course and development of the work.

Regarding the datasets explored, it has been deliberated that the most appropriated one is the ISIC-2016 challenge dataset. As previously depicted in *Section 4.1.1*, the database comprises 900 training images and 379 test images, so it is constructed by a total of 1,279. The training data is labeled and consists of 172 melanomas and 728 benign lesions. In addition, the similarity among this dataset and the images acquired daily in Hospital Sant Joan de Déu has promoted the selection of ISIC-2016.

Among the different datasets within the ISIC challenge database, the 2016 year has been selected because the size of the dataset is not that large. Since it is not being considered the implementation of any artificial intelligence tool, it is not necessary to employ that much data so a limited quantity of images is enough for the correct development of the project. Another important aspect to mention

is that, since the project is not intended to train any artificial tool, the fact that ISIC-2016 does not comprise validation images is irrelevant to the course of the project. In addition, the scope of this work does not consider any classification stage of the skin lesion segmented so considering that the images are not labelled into a class, as happens in the that specific database, is also unimportant. But the main decision factor that has led to decide ISIC-2016 as the selected option has not only been these reasons but its high-quality dermoscopic images and their clinical relevance, since it includes real-world clinical scenarios encountered in dermatology practice.

As regards to the other databases mentioned, those were discarded because during the course of bibliographic research, few articles were found exploiting a database which was not from the ISIC challenge. That made think of their clinical irrelevance among skin lesion segmentation field.

The core of the project regards to the segmentation of a set of skin lesion images, so the decision whether a segmentation pipeline is selected or not is crucial for the obtaining of reliable outcomes. From the proposed methods depicted in *Table 1*, some have been considered due to its high performance rates reported by the authors and segmentation success denoted. It was to our concerning attempting to implement three different pipelines which were based on different segmentation basis in order to make a comparison among them and determine which could apport more accurate results for the given segmentation of melanocytic lesions.

From the encountered papers and bibliographic research, it was finally decided the implementation of two of the options. Firstly, the paper proposed by Ma Z. and Tavares J. [55], regarding the level set method. The authors report good results and accurateness of the segmentation method when compared to ground truth. The numerical results illustrate the effectiveness and robustness of the algorithm. What is more, the color information of the image is efficiently used in order to perform the segmentation, which attracted even more its application. The employment of this characteristic is a novel approach and has not already been exploited that much.

The second paper picked out is the one proposed by Jaisakthi *et al.* [58], regarding the GrabCut algorithm and k-means clustering. The implementation of not only the segmentation of the skin lesion itself by the application of some different pre-processing steps appealed its validation within the course of this given project due to its complexity. Different pre-processing steps are made use of in the input image in order to enhance the performance of the pipeline. Moreover, the results reported and concluded by the authors in the paper demonstrate high accuracy, sensitivity and specificity rates not only when considering the pipeline excluding the preprocessing steps but also when incorporating them.

Then, the pipeline that most attracted to introduce within the project has been the MedicalSeg application. It was thought to be an encouraging solution to take advantage of a graphical interface to make more complete the resolution of the project. In fact, it was concluded between the research group members that it could be an appealing proposal to better study how their in-house developed

algorithm behaves. Among the different technique options available in the interface of the app, the one chosen to apply to the melanocytic skin lesions has been intensity thresh automatic segmentation. The reason of this segmentation method selection has been because of the author's report of low time-computing, and high Dice coefficient values when compared to ground truth images. Segmentation methods such as snake segmentation or Canny threshold were thought of but active contour-based and region-based methods are already the basis of the previous two pipelines chosen. Others such as watershed segmentation or normalised segmentation were tried but the resulting binary images obtained were discarded on account of their inadequate and inaccurate execution.

The alternative papers have not been tried to implement since the segmentation pipelines described used a programming software never used so it would become more difficult to replicate the pipeline. What's more, the mathematical basis behind the development of the segmentation algorithm described was too complex and beyond our academic intellectual reach. In addition, the images they employ to segment regarded all type of skin lesions while the objective of the project is mainly to process with melanocytic lesions as the papers selected do. Other option exhaustively studied, which was the imaging tool named ImageJ, was also discarded since MedicalSeg had a similar aimed and was more focused on medical image segmentation.

In the following section (*Section 5*), a more accurate explanation of the methodology of these three pipelines is commented and described so a higher understanding and comprehension of the solutions implemented is accomplished.

Concerning the programming software of preference, not only one of the earlier explained language tools was picked out, but a few of them were chosen. The first two algorithms selected, level set method and the one combining GrabCut and k-means, had been engineer in Python programming language. Then, MedicalSeg application has been developed to be currently installed in the main app panel of MATLAB workstation. For this reason, the second programming software to consider has been MATLAB. This interface was used to developed most of the project; approximately 65% of the work was carried out in MATLAB since it was also employed as the language tool for the conduction of the assessment and evaluation of the segmentation results. No inconvenient was found during the resolution of the project, since both platforms have been widely used in the course of the four years of the biomedical engineering degree. The experience gathered together with their ease of usage and apprenticeship made those two the best options. What is more, the implementation of their image processing libraries was familiar to us due to their previous usage along the degree.

After all, the segmented images must be analised so the performance of the three different methodologies can be assessed. To do so, the statistical metrics chosen have been the precision, sensitivity, accuracy, Dice coefficient and Jaccard coefficient. The reason behind the selection of the indicated parameters has been since the literature review done during the bibliographic

research led to lots of papers making use of these five metrics for the evaluation of their own segmentation processes. That is, most of the researchers used them for evaluation purposes in the context of skin lesion segmentation. The mean of the different parameters is aimed to compute so the results of the segmentation process are validated.

5. DETAILED ENGINEERING

This section describes in detail the proposed solution as well as all the different steps carried out in order to implement the project. Those are depicted and minutely defined next.

5.1 EVALUATION DATASET

5.1.1 ISIC-2016 DATASET

The usage of ISIC (International Skin Imaging Collaboration) datasets is broad for implementing and validating the majority of tasks focusing on segmentation and classification. They have widely become a driving force for research into melanocytic lesion segmentation. The dataset consists of biopsy-proven digital and high resolution dermoscopic images that contain annotations from clinicians all around the world. The main aim of the construction of this database was to promote research in the medical segmentation field in order to develop more powerful and automated CAD systems for the diagnosis of skin disorders. The ISIC community yearly organises skin lesion challenges to attract the participation of new research groups who are looking forward to enhance the diagnosis performance of those CAD algorithms. They also purpose to spread awareness of the current challenge the condition represents.

As beforehand explained, the year dataset selected for the evaluation of the current study is 2016. The ISIC-2016 database contains 900 training images and 379 test images in JPEG format, so it is formed by a total of 1,279 images. Ground truth data, that is manual segmentations performed by experts, is also provided for both training and test sets in PNG format. The image distribution within the training dataset consists of 172 melanoma images and 728 benign cases. This is, the ISIC-2016 dataset only involves data regarding melanocytic lesions, so it does not differentiate among different skin disorders but it is centered on whether the lesion images are a melanoma or they are simply a benign tumour. However, for the development of the ongoing project, this should not be taken as a disadvantage, considering that the scope not encompasses the classification step of the lesion whether it is, for example, melanoma or congenital nevus, but only the segmentation of the image.

5.1.2 DATA PREPARATION AND PRE-PROCESSING

ISIC-2016 dataset provides RGB images with resolutions ranging from 556 × 679 to 2,848 × 4,828 pixels. All images were resampled and their size was down by 25% using a MATLAB code (see *Appendix*). This step was executed as a means of improving the computational time of the algorithms' performance.

An additional step before the implementation of the different chosen pipelines, was the selection of a set of images from the 1,279 images data bench. As earlier specified in the scope's section of the project, this research does not embrace rather the implementation or evaluation of any specific

machine learning tool so a large amount of data is not fundamental for the correct evaluation of the presented algorithms.

So, as a way of maximizing their performance capabilities and minimizing the computational time one more time, only 48 images were selected from the ISIC-2016 database training images folder. These images were conscientiously selected in order to complete a varied dataset to evaluate the results of the proposed methods. This was executed in this manner since melanocytic lesions can be found in innumerable types of shapes, colours, textures, sizes, etc. Constructing a dataset with similar lesions characteristics would have led to non-precise results; this is, even though the segmentation algorithm performs well for similar melanomas sizes does not mean that for enlarged tumours the pipeline will work as well.

5.2 IMAGE SEGMENTATION PIPELINES

The implementation of the three pipelines has taken place by means of two different programming language tools; MATLAB and Python. Level set method application, throughout Python, has been achieved due to its free publication on a *Kaggle* repository [89], as well as through its corresponding paper describing the segmentation process [55]. Then, the algorithm combining GrabCut and k-means methods, implemented in Python too, has been possible to put into practice by means of manually replicating the pipeline described and proposed by Jaisakthi *et al.* [58], since the code itself is not public available. The different processing steps described within the paper for the completion of the total segmentation pipeline have been extracted from different *GitHub* repositories and informatic web pages. Then, the last algorithm analised, intensity threshold automatic segmentation, has been applied by means of the medical user interface named MedicalSeg, a self-contained MATLAB App. The code developed for the application of the algorithm lies integrated within the software application itself, so no public code research has been looked through for this first case.

5.2.1 LEVEL SET METHOD

Recently, there are many image segmentation algorithms that have been based on the level set method. It has been employed for a wide range of different applications, and a large number of researchers are still putting all their efforts into improving and enhancing the efficiency and effectiveness of the algorithm.

The level set method was firstly developed in order to track curve evolution in computational physics but it has been recently applied in other areas of image processing such as the medical one. The method makes it very easy to follow shapes that change topology such as cases when a shape gets divided in two or develops holes. Recent publications have enhanced the main capabilities of this algorithm when applied to the segmentation of dermoscopic images; it is less sensitive to the influence of noise compared to other methods. In this section, the methodology of

the level set method is explained. [55] The workflow of the proposed method is depicted in *Figure 9.*
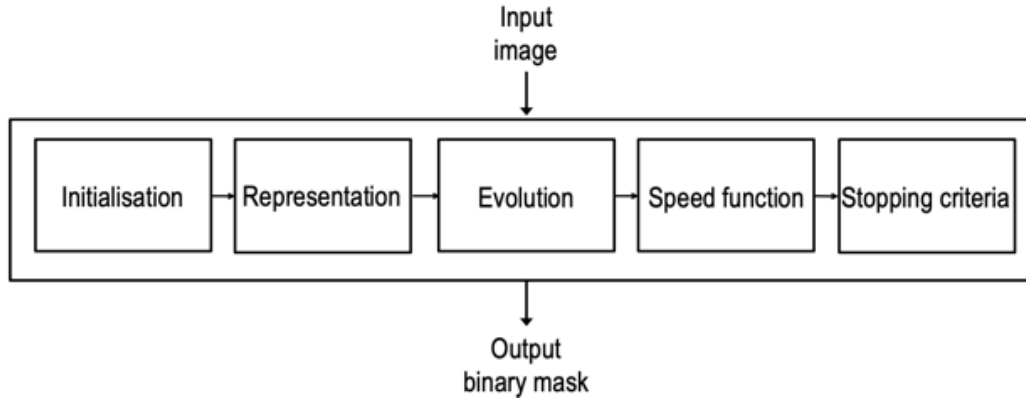
Figure 9. Workflow of the level set method [own source].

The level set method was proposed to resolve topological changes during curve evolution. The key idea behind this method is to represent curves as the zero level set of a higher dimensional function. The function is typically referred to as the level set function or the signed distance function. This approach not only provides accurate numerical implementations, but it can handle topological changes effortlessly. What it basically means is that the evolving curve in a two-dimensional plane x-y is embedded into a continuous surface of a three-dimensional level set function $\phi(x, y, t)$ as its zero level set, that is, the evolution is followed by finding the zero level set of the function $\phi(x, y, t)$ at time t. The motion equation of the level set method is depicted in *Equation 8.* [55]

$$\frac{\partial \phi}{\partial t} + F|\nabla \phi| = 0$$

Equation 8. Motion equation of level set method.

where $\phi(x, y, t)$ is the level set function, and F is the speed function, which is a function related to the evolving surface and image. It must be defined so the curve can reach the object boundary and accomplish and stable status there, this is, its ideal value ought to be zero on the edge of the object. The main purpose behind the level set method is to model segmentation as a curve evolution procedure. [55]

The distribution of colour in a dermoscopic skin lesion is habitually non-homogeneous. If the curve evolves inside the region of the skin lesion, the resulting segmentation could be wrongly performed since the curve can be easier attracted to the inner boundaries. This is the reason why this algorithm constrains the evolution of the curve in normal skin area. The initial curves, then, cover the entire region of the skin lesion. With the curve as the boundary, the values of the level set function $\phi(x, y, t)$ at $t = 0$ are defined as the signed distance function, which assigns positive values to points outside the surface and negative values inside. [55]

The proposed method was implemented using Python & *OpenCV.* The pipeline starts by initialising an initial contour enclosing the object of interest, the skin lesion. Next, evolution of the surface comes about over time by using partial differential equations (PDEs) such as the mean curvature flow or the geometric flow. It involves the updating of the signed distance function over time to track the deformation of the curve. Then, the speed function is defined so it controls the evolution of the surface. This function can be based on image intensity, gradient information, and it is employed to guide the curve towards the boundaries of the lesion. Finally, a stopping criterion must be established to determine whether the surface has evolved enough or not. That can stand on the magnitude of the surface velocity or the change in surface shape. Finally, the final segmentation is extracted by thresholding the signed distance function. The overall process is iterated as many times as desired until the aimed convergence criteria is met. [55] For the case of this proposed algorithm, the number of iterations was of 15.

In order to illustrate the method proposed, *Figure 10* shows the evolution of the curve towards the boundary of the lesion. The process starts with the initialisation of the rectangle which encloses the region of interest. Then, the curve evolves; it moves and changes shape until it correctly separates the lesion of the background. Once the curve has evolved enough, the stopping criterion determines the end of the segmentation process. The resulting binary mask consists of black pixels conforming the background and white pixels conforming the lesion region. Finally, the mask is upscaled to its original size, so it can be further compared to its ground truth.
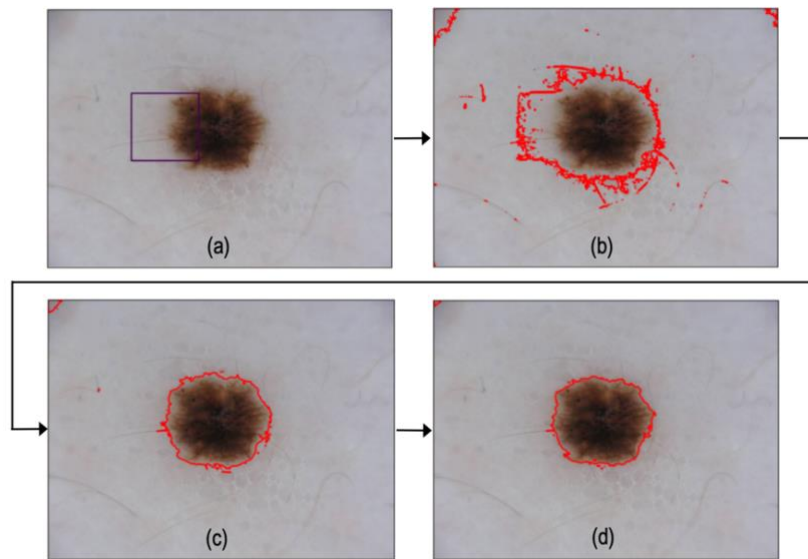
Figure 10. Visual workflow of the level set method. (a) Initialisation of the rectangle, (b) first iteration, (c) seventh iteration, (d) fifteenth and last iteration (final segmentation result). Images from ISIC-2016 database [72].

Some other examples regarding the level set method are depicted in *Figure 11*. Due to its great stability, the method consists of a powerful and versatile tool for image segmentation which can result in accurate results in a wide range of applications. Some of its most remarkable benefits include its flexibility when dealing with complex shapes and the preservation of topology.

Nonetheless, it is necessary to remember that the method can become computationally intensive and that may require fine-tuning of parameters. It is an appropriate choice when handling with well-defined boundaries. Despite all, it remains a valuable tool for medical image segmentation, and it persists to be widely researched in other many fields.
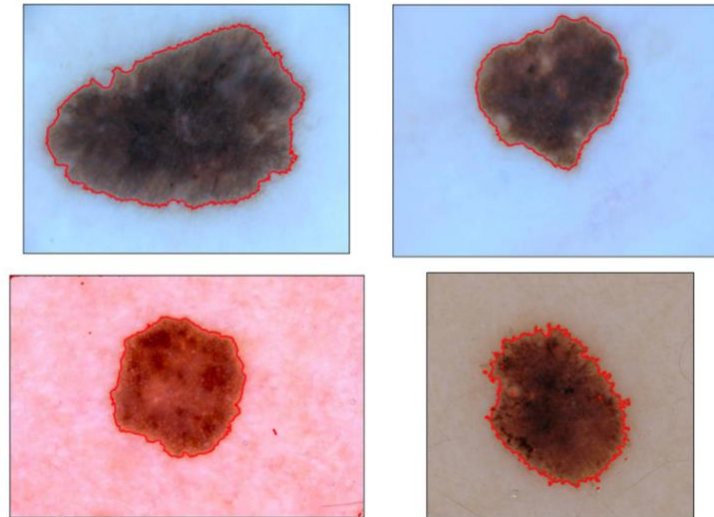
Figure 11. Segmentation results using level set method for ISIC dataset images from [72].

5.2.2 GRABCUT AND K-MEANS SEGMENTATION ALGORITHMS COMBINATION

The current segmentation model consists of a semi-supervised learning technique to automatically segment melanocytic lesions of a given image. The methodology is based mainly in two main steps; the first one regards to the appropriate pre-processing of the image in order to filter out the artefacts present in the image, such as illumination imbalances or hair, and the second step which is the segmentation itself. The segmentation presented in this pipeline is done by the combination of two different approaches: GrabCut, a region-based algorithm, and k-means, a clustering-based algorithm. [58] The workflow of the proposed method is depicted in *Figure 12.*
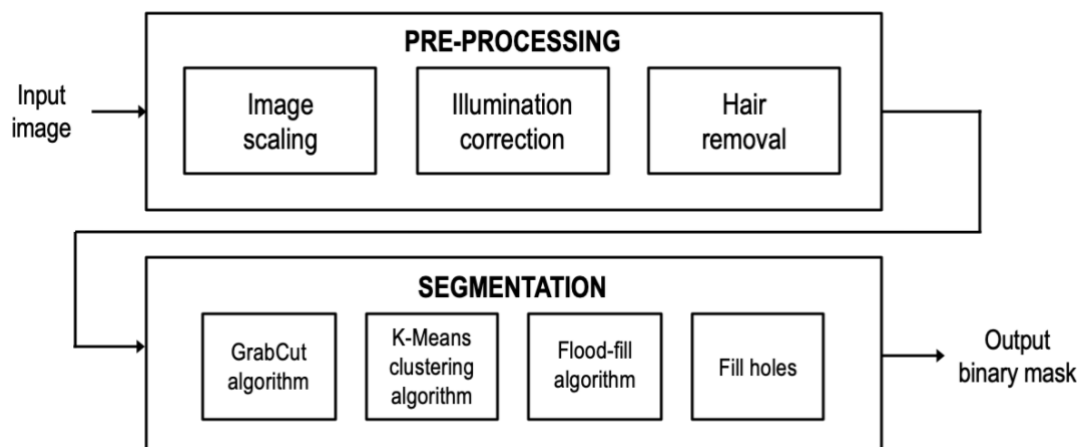
Figure 12. Workflow of the combination of GrabCut and k-means segmentation algorithms [own source].

The proposed method was implemented using Python & *OpenCV*. The self-made code developed is attached in the *Appendix*. The first step of the process involves the pre-processing. The structure of a skin lesion image widely varies among skin condition, ethnicity, image acquisition modality, etc. That is the reason why artefacts such as hair, air bubbles, ruler images or contrast variations ought to be removed so the segmentation process is facilitated. So, to improve the performance of the lesion segmentation, it is a must to take into consideration the pre-processing step.

The workflow described by the paper begins by scaling the image. However, the 48 images have been previously downscaled when preparing the data so this specific step does not have to be re-applied again.

The following stage comprises the correction of the illumination in order to eliminate uneven illumination of the image. The RGB image is converted into Lab colour space and the Contrast Limited Adaptive Histogram Equalisation (CLAHE) algorithm is applied to L channel. CLAHE is an upgraded version of Adaptive Histogram Equalisation (AHE) algorithm. This technique is used to enhance the local contrast of the image and to improve the definition of edges in each region of an image. The contrast enhanced L channel is then merged to form the Lab image and converted into the RGB space again. The code performing CLAHE algorithm has been extracted from [90].

The last step of the pre-processing is the hair removal. Dermoscopic images often present hair, which makes the segmentation process even harder because the hair pixels impede important characteristics of the image such as boundaries. In the proposed method by Jaisakthi *et al.*, the Frangi vesselness algorithm along with Fast March Method (FMM) is used to remove the hair. However, for the development of the given project, the combination of these two algorithm has not been contemplated and the BlackHat algorithm has been put into practice due to its easiness of implementation. The BlackHat algorithm is a morphological operator used in image processing to enhance dark structures within an image. So, by highlighting those darker features and intensifying the artefacts contours, the hair is detected, and the image can be easily inpainted. Code is publicly available in [91].

Now the segmentation step takes place. The proposed method involves first the GrabCut algorithm to localise the approximate lesion region and then the k-means clustering is employed to confine the exact lesion region. The detailed description of both methods is explained next.

The GrabCut segmentation algorithm uses both the boundary and the information of the region to segment the foreground image in contrast with many algorithms which employ the information of either the edges or the region. This makes a key difference for the selection of GrabCut since skin lesions do not have defined edges. Code is publicly available in [92].

The GrabCut algorithm starts by defining a rectangle that encases the main object. Every pixel outside the rectangle is considered as background pixel and the pixels inside are still unknown. Then, based on the given input, a Gaussian Mixture Model (GMM) is used for modeling foreground

and background. GMM generates labels for the unknown pixels and each of them is clustered in terms of color statistics. A weighted graph is generated where the pixels in the images are regarded as nodes and two additional nodes, Source node and Sink node, are added. *Figure 13* represents visually the scheme described. The foreground pixels are connected to the Source node and the background ones to the Sink. Then, the algorithm iteratively updates the labeling of the pixels based on an energy function minimisation and estimates the probability of belonging to a class or another by means of the GMM. When the energy function reaches a minimum, the pixels connected to the Source node are classified as foreground and the ones connected to the Sink as background. [92]
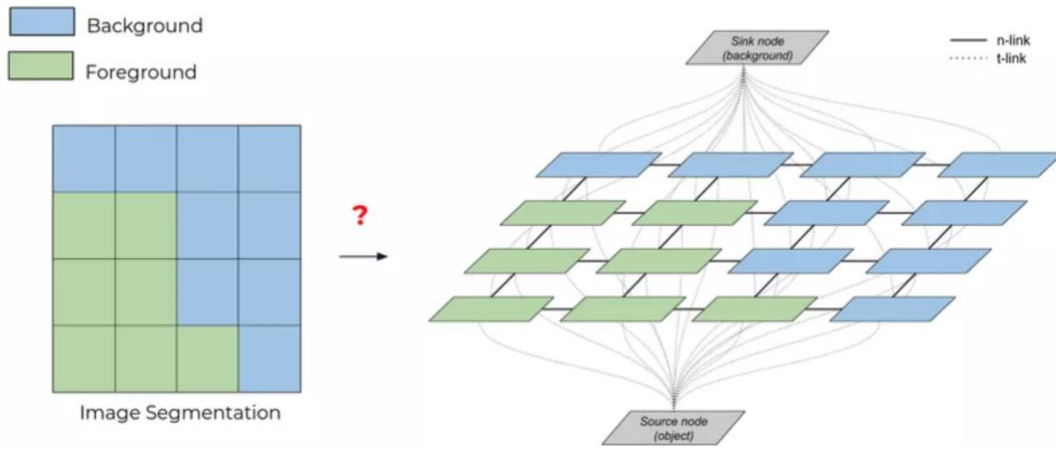
Figure 13. Visual representation of the graph generated in the GrabCut algorithm. [93]

The segmented image obtained from the GrabCut algorithm is further pre-processed by means of k-means clustering algorithm of open access from [94]. It has been employed in order to group the pixels of the foreground in RGB colour space. The method picks a first cluster centre randomly from input pixels. Then, each cluster is distinguished from the rest by means of a probability proportional to the squared distance between the pixel and the closest cluster centre. The value of k chosen has been of four, so the different groups to which the pixels can belong to are (i) interior skin lesion, (ii) lesion boundary, (iii) lesion background (normal skin) and (iv) image background. Image *(e)* from *Figure 14* illustrates the labels: dark brown corresponds to the interior of the skin lesion, brown is the lesion boundary, light brown consists of the lesion background and black is image background.

From the four clusters defined, the aimed one to choose is that with enclose the region of interest. So, from the remaining clusters, the lesion boundary region is identified using the flood-fill algorithm, an image processing algorithm used to fill connected regions with a specific colour. [58] Then, the binary mask is upscaled to its original size, so it can be further compared to its ground truth.

Finally, the lesion region is extracted but this may contain holes and these holes are filled by means of morphological operators. The image is converted into a binary mask; black pixels are background

and white pixels the lesion region. Only this final step is conducted in MATLAB, that takes part of the post-processing of the image (see *Appendix*).
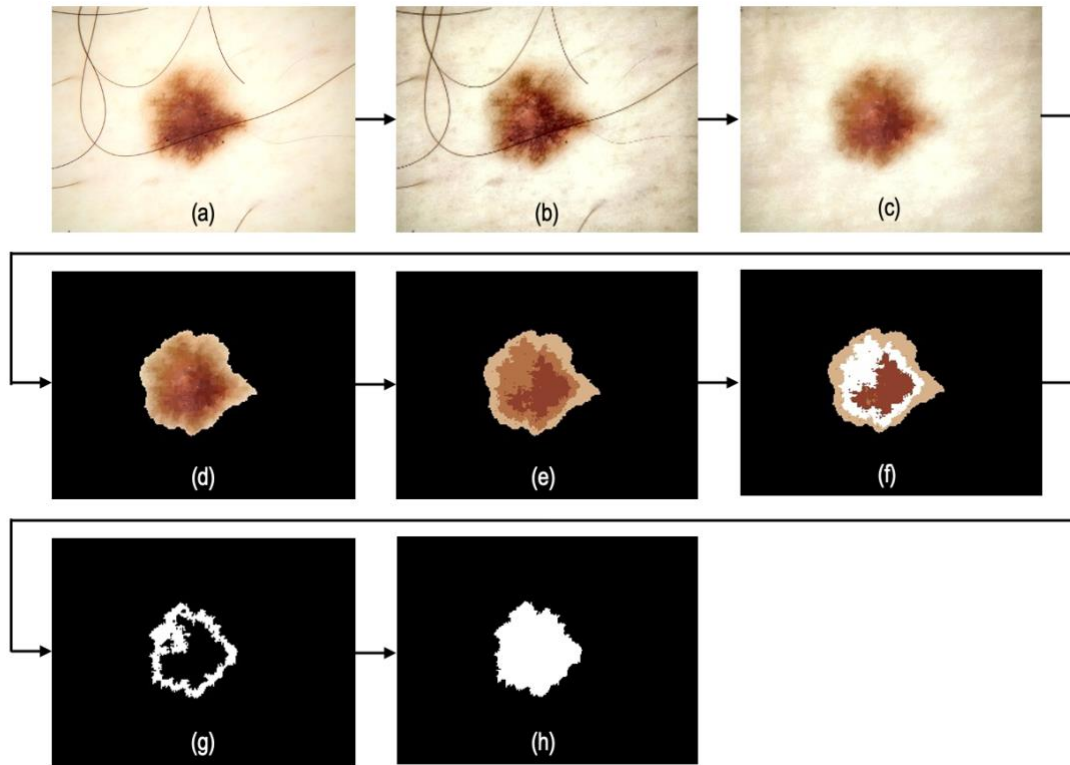
Figure 14. (a) Original image [72], (b) Illumination corrected image, (c) Hair inpainted image, (d) Foreground extraction image by GrabCut algorithm, (e) Formation of colour clusters by k-means algorithm, (f) Boundary region selected after flood-fill algorithm and (g) extraction, (h) Segmented binary image after fill holes [own source].

GrabCut algorithm offers several advantages in skin lesion segmentation field. Firstly, it is an interactive refinement, that is, it enables users to have a fine control over the segmentation process since they are able to select an initial seed to designate foreground. In addition, GrabCut algorithm is known for its accuracy when delineating boundaries. By iteratively optimizing the segmentation, the algorithm can easier capture refined details and irregular shapes. GrabCut has demonstrated good generalisation capabilities across different skin lesion types and it can effectively adapt to colour, texture and shape variations. While GrabCut offers those significant advantages, it may still have some limitations, such as complex or highly textured lesions. This is why, by the combination with k-means algorithm, the outcomes obtained seem to be promising. [92]

The incorporation of k-means to the pipeline enhances the strengths of each methodology independently and it can convert the overall process in an approach able to overcome many limitations. K-means helps in modelling colour information and it can capture easier its variations so a more accurate differentiation between foreground and background is performed. Then, since k-means not only considers colour characteristics but additional texture features, the segmentation can be benefited with a higher comprehensive representation of the lesion.

5.2.3 MEDICALSEG

The in-house developed MedicalSeg application has as main objective to create a platform for comparing different segmentation methods and to generate segmented images purposed as artificial intelligence tools. It has been developed in order to provide the medical community a graphical user interface (GUI) to manage big amount of images and to create ground truth data for different objectives. [62] Even though the graphical interface of the MedicalSeg application supplies a variety of segmentation approaches, only one of them is implemented along the execution of the project. The method finally selected is intensity threshold automatic segmentation. But not only the medical application is intended for segmentation, but it allows the pre-processing of the image. In the present case, the images firstly undergo a Gaussian filter provided by the segmentation process itself. The workflow of the proposed method is depicted in *Figure 15.*
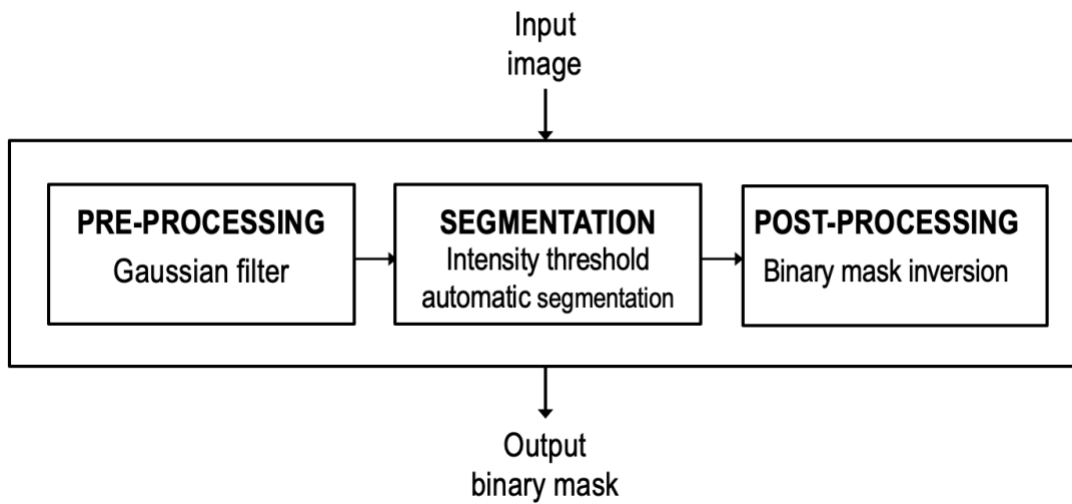
Figure 15. Workflow of the intensity thresh automatic algorithm [own source].

Herein, a detailed description of the process carried out is explained. The first step is to define a path definition workspace. User must select an input folder containing the images aimed to be segmented, which will then be loaded. Next, an output folder must also be selected in order to save the results generated by the segmentation process.

Once the images have been successfully loaded, they are processed by means of a Gaussian filter. This filter is available in a tools panel section present in the interface of the application. However, the selected segmentation algorithm has within its pipeline the Gaussian filter by default so the pre-processing step is not manually performed. The filter is applied to smooth the image and to remove noise and artefacts. A Gaussian filter is a low-pass spatial filter employed for reducing those high-frequency components present within an image. The filter uses a kernel that has the shape of a Gaussian distribution function. [62]

As already stated, the segmentation approach for evaluation is intensity thresh automatic segmentation. The algorithm is based on Otsu's thresholding and uses a standard deviation σ for

the Gaussian filter. It consists of separating regions of interest from the background based on the intensity values of the image pixels. It takes a specific threshold value that minimizes the interclass variance of the black and white pixels; then, it separates the pixels with intensities above or below the threshold [62]. The method is selected by simply ticking the box in the corresponding segmentation technique's panel. The sigma value chosen has a value of 0.25 and the threshold value is an automatic thresholding from a global threshold of the grayscale image.

Then, the segmentation is automatically generated and the results are showed within the interface of the MedicalSeg application; the original image, the segmented image and the masked image are depicted. The output folder previously chosen now includes not only the segmented and the binary masked images, but the corresponding image matrices. As an example, the outcomes of the pipeline are pictured in *Figure 16.*

Figure 16. Original image, segmented image for a melanocytic lesion given in the ISIC-2016 dataset [72] and its binary mask [own source].

The binary mask obtained represents the lesion as black and the background as white. However, the desired outcome must be the complement, the lesion region coloured as white while the background remains black. It is for this reason that the final image must be inverted before the evaluation stage. Moreover, the mask must be upscaled again to its original size, so it can be further compared to its ground truth. The resizing and inversion is not performed within the interface of the application but by a simple coding in MATLAB.

Then main advantage regarding the usage of segmentation algorithms based on pixels intensity is their straightforward and mainly the easiness of implementation of the methods. What's more, the low computation cost they provide must also be enhanced. Intensity-based segmentation algorithms are often computationally efficient when compared to other more complex methods since they typically involve a simple thresholding or basic mathematical operations. Then, the minimal requirements the method demands are of such importance: not extensive training data is needed, so it makes them suitable for scenarios where data may be limited. Last but not least, the applicability to different lesion types has to be remarked on account of the fact that it is effective for disorders such as melanoma, basal cell carcinoma or other conditions. [33]

Even though the several advantages it offers, it may not capture complex details or handle complicated cases when boundaries overlap. In those situations, the combination of the intensity-

based technique with other advanced methods ought to be considered to improve segmentation accuracy. [33]

5.3 SEGMENTATION RESULTS

5.3.1 FINAL RESULTS

As mentioned beforehand, to evaluate the performance of the three proposed methods, 48 images of the dataset ISIC-2016 training folder are used. The ground truth file for each image consists of the binary mask image with black pixels representing the background and white pixels representing the lesion region. Level set method algorithm and the combination of GrabCut and k-means algorithms have been implemented in Python, throughout the help of the library *OpenCV*, while intensity thresh automatic segmentation has been implemented in MATLAB. The predicted binary masks obtained are all of them in JPEG format.

Then, validation of the results has been executed using MATLAB. After loading both predicted binary masks and ground truth masks, the proposed methods are evaluated using performance indices: precision, sensitivity, accuracy, Dice and Jaccard coefficient are calculated.

In order to visually characterise the performance of the three different approaches, *Figure 17* depicts five examples. Each row is a lesion case and it consists of the original image, and the binary masks obtained throughout the level set method, the combination of the GrabCut and k-means algorithms and the intensity thresh automatic segmentation approach. The ground truth image is also depicted in the last column, so a better interpretation of the segmentation outcomes can be assessed.

At first glance, the images may seem similar among them. They show good delimitation of the lesion and no artefacts of the original image seem to disrupt the outcomes. Besides, not only the similarity is found among the three methods but with the ground truth image. Ground truth appears to consider more pixels as skin lesion since it visually can seem to have a larger area. Nonetheless, the objective of the project purposes to validate in a more accurate way how the performance of each approach is. For this reason, *Table 3* depicts the different performance metrics previously commented.
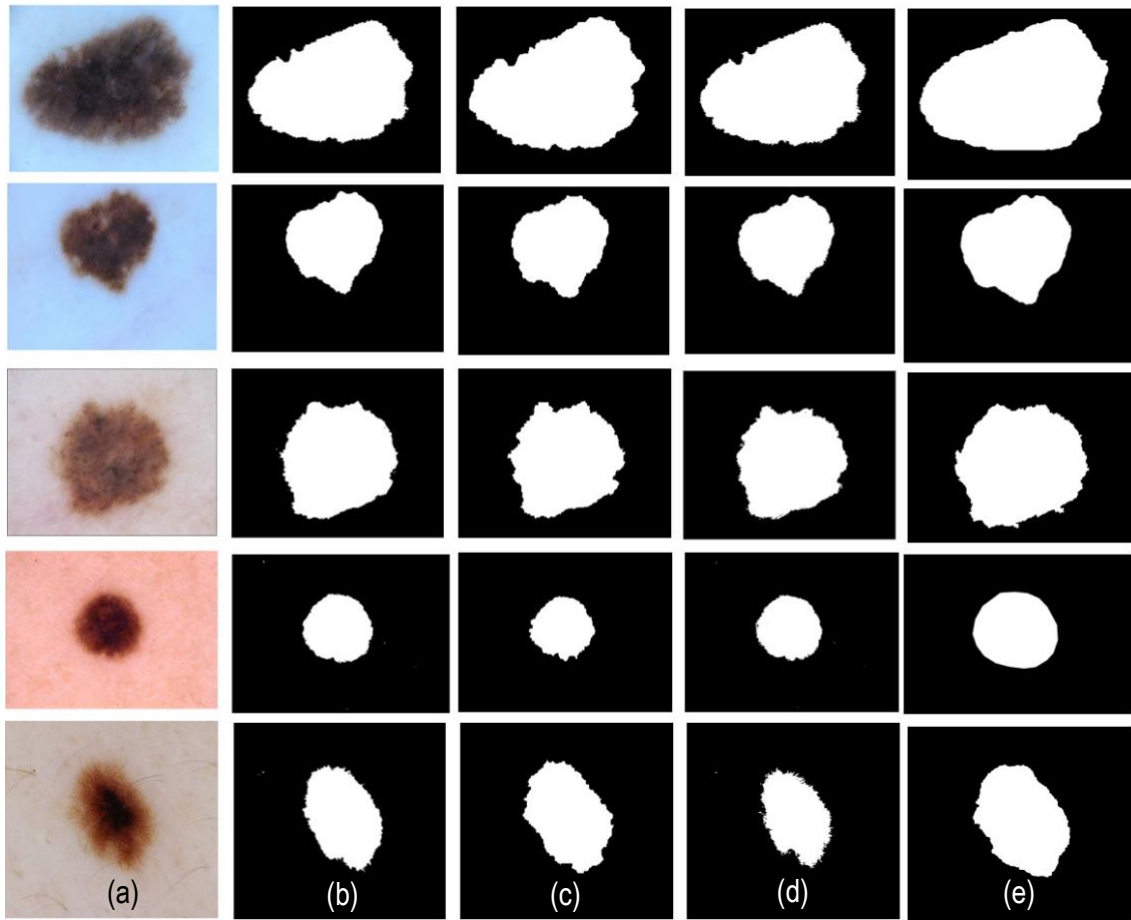
Figure 17. Binary masks obtained for each of the three methods proposed. (a) Original image [72], (b), binary mask obtained throughout level set method, (c) binary mask obtained throughout the combination of GrabCut and k-means algorithms, (d) binary mask obtained throughout MedicalSeg in-house developed algorithm (intensity thresh automatic) and (e) ground truth image [own source].

Table 3 involves the precision, sensitivity, accuracy and Dice and Jaccard indices for each of the five cases shown in *Figure 17*. The values presented are all but diverse; some images depict high precision values when compared to ground truth while its sensitivity is lower. The MATLAB code implemented for the calculation of these parameters is attached in the *Appendix*. In the coming sections, the interpretation and discussion of what these different values mean is going to take place.

	Performance metric	Level set method	GrabCut and k-means	Intensity thresh automatic
	Precision	1	0.9850	0.9999
	Sensitivity	0.8754	0.9520	0.8881
Image 1	**Accuracy**	0.9420	0.9709	0.9479
	Dice coefficient	0.9336	0.9682	0.9407
	Jaccard coefficient	0.8754	0.9384	0.8880
	Precision	0.9995	0.9810	0.9992
	Sensitivity	0.9275	0.9609	0.9208
Image 2	**Accuracy**	0.9850	0.9881	0.9835
	Dice coefficient	0.9621	0.9708	0.9584
	Jaccard coefficient	0.9271	0.9433	0.9200
	Precision	0.9683	0.9998	0.9471
	Sensitivity	0.9005	0.8754	0.8255
Image 3	**Accuracy**	0. 9593	0.9606	0.9304
	Dice coefficient	0.9332	0.9335	0.8821
	Jaccard coefficient	0.8747	0.8753	0.7891
	Precision	0.9966	1	0.9963
	Sensitivity	0.7510	0.6625	0.7054
Image 4	**Accuracy**	0.9646	0.9525	0.9582
	Dice coefficient	0.8566	0.7970	0.8260
	Jaccard coefficient	0.7491	0.6625	0.7036
	Precision	0.9983	0.9641	0.9989
	Sensitivity	0.8228	0.9633	0.6957
Image 5	**Accuracy**	0.9700	0.9878	0.9488
	Dice coefficient	0.9021	0.9637	0.8202
	Jaccard coefficient	0.8217	0.9299	0.6952

Table 3. Precision, sensitivity, accuracy and Dice and Jaccard coefficients for the images depicted in *Figure 17*.

5.3.2 MODELS VALIDATION

But the depiction of five different examples is not enough to validate the final performance of the models. The following section aims to present the mean value of the performance metrics regarding each one of the methods. The information got by the evaluation metrics is used for the performance checking and a result's comparison is then given.

The performance indices used are pixel level estimators. Precision quantifies the proportion of correctly labeled as lesion pixels among all the pixels identified as lesion. It indicates how reliable the segmentation results are in terms of minimizing false positives. Sensitivity refers to the amount of lesion pixels that are predicted as lesion pixels. Accuracy of the segmentation is the estimation of how many pixels out of all are correctly labeled when compared to the ground truth image. Dice and Jaccard indices evaluate which is the degree of overlapping of the ground truth image and the segmented lesion region. The performance indices are calculated for each of the 48 pair of ground

truth image and predicted binary mask image. Then, the mean is computed for each performance index and method. *Table 4* shows the average performance of the three algorithms studied, as well as the standard deviation of each index.

	Level set method	GrabCut and k-means	Intensity thresh automatic
Precision	0.9331 ± 0.1155	0.9842 ± 0.0251	0.9689 ± 0.0530
Sensitivity	0.8720 ± 0.1045	0.8629 ± 0.1125	0.7830 ± 0.1386
Accuracy	0.9542 ± 0.0342	0.9643 ± 0.0288	0.9390 ± 0.0459
Dice coefficient	0.8931 ± 0.0798	0.9145 ± 0.0637	0.8586 ± 0.0954
Jaccard coefficient	0.8151 ± 0.1186	0.8483 ± 0.1012	0.7632 ± 0.1348

Table 4. Results of the proposed methods using ISIC-2016 dataset.

High precision rates indicate low number of false positive pixels, which is a crucial aspect when aiming to identify accurately the boundaries of the skin lesion. Then, sensitivity achieved for each algorithm widely varies among the intensity thresh automatic segmentation and both level set method and the one combining GrabCut and k-means approaches. The first one mentioned has provided with a low sensitivity value compared to the other two, which have come to a higher similar result. A high sensitivity rate indicates a low number of false negative pixels. That is, the rates demonstrate that the techniques have identified at least some part of the lesion region.

The choice whether precision or sensitivity may suggest a greater importance lies in the specific requirements of the algorithms implementation. If the main concern is to rigorously identify skin lesion pixels, precision should be taken more into account. But, if the main concern is to capture as many skin lesion pixels as possible, even if it accentuates the detection of false positives, sensitivity may be prioritised. However, a balance between both indices should be accomplished. This can be achieved throughout the usage of the Dice or Jaccard coefficients.

Then, accuracy values achieved are the highest rate values regarding the segmentation performance indices. The values obtained are remarkable in all the three cases, which implies reliability within the methods. So, the results denote that the majority of the image pixels are accurately classified, whether they are part of the skin lesion or they are normal skin.

Considering the Dice and Jaccard coefficients acquired, the outcome values represent a high degree of overlap between the segmented region and the ground truth, giving insight again into the accuracy of the segmentation algorithms. However, the method combining GrabCut and k-means algorithm achieves the highest indices. Considering that it is the only one methodology in which the preprocessing steps are taken in consideration, it can be shown the importance of making the images undergo some pre-processing tools in order to achieve as good as possible segmentation results. Both metrics contribute with valuable insights into the accuracy and precision of the segmented skin lesion.

5.3.3 DISCUSSION OF THE RESULTS

After the exhaustive literature review and exploration among the different segmentation algorithms presented in the current study, the outcomes achieved must be notoriously emphasized because of their good results. The three of them end in precisely segmented melanocytic lesions and provide good accuracy so they might be think of to be implemented in clinical practice. From Hospital Sant Joan de Déu experts project team, the validation of the binary masks predicted has been considered as a success and they have ascertained their utility in routinely investigation.

Automatic delineation of skin lesion is a greatly demanded step for CAD systems for a later disorder classification. This section reviews the segmentation performance of the proposed methods using 48 images from the ISIC-2016 training folder dataset. Quantitatively, *Table 4* summarises the segmentation performance results of the level set method, the algorithm combining GrabCut and k-means pipelines and the intensity thresh automatic algorithm. The performance is shown in terms of precision, sensitivity, accuracy, dice, and Jaccard indices.

When validating the results accomplished, many are the aspects which can be discussed. One of the most important ones when assessing the performance of a segmentation algorithm is, at our concern, accuracy. The quantitative results in *Table 4* show the accuracy of the lesion segmentation using any of the three pipelines for ISIC-2016 dataset and demonstrates their robustness. The values obtained are 0.9542 for level set method, 0. 9643 for the method combining GrabCut and k-means algorithms and 0.9390 for intensity thresh automatic algorithm. From these values, it can be highlighted the ability to reliably identify the boundaries of the skin lesion as a larger proportion of the pixels are correctly assigned to their respective class. That means that each algorithm has a strong capability to differentiate among lesion and normal skin.

Regarding precision and sensitivity, much more cannot be said as their values are within the same ranges that the accuracy ones, even though sensitivity values are a little bit lower. Since precision focuses on minimising false positives, it can be extracted from the values achieved that the algorithms produce a low number of false positives, that is, it robustly identifies lesion pixels without misclassifying too many normal skin as part of the lesion. Then, sensitivity focuses on reducing false negatives and identifying positive pixels out of all the actual ones present in the image.

Then, the values for precision and sensitivity respectively are 0.9331 and 0.8720 for level set method, 0.9842 and 0.8629 for the method combining GrabCut and k-means algorithms and 0.9689 and 0.7830 for intensity thresh automatic method. What it is tried to demonstrate with these values is that the three pipelines perform successfully since they do not mislabel a normal skin pixel as part of the region of interest, while its complexity gets worsened when having to identify lesion pixels out of all the actual ones.

Last but not least, the resulting values of the Dice and Jaccard indices are a must to be stated. The values for Dice and Jaccard coefficients respectively are 0.8931 and 0.8151 for level set method,

0.9145 and 0.8483 for the one combining GrabCut and k-means algorithms and 0.8586 and 0.7632 for intensity thresh automatic method. It is then necessary to emphasize a strong overlap between the predicted binary masks and the ground truth images. A high value for these indices implies that the methods accurately delineate the lesion region while aligning satisfactory with ground truth. So it proves the ability of the pipelines to merely diminish false positives and negatives at the same time.

In order to have a better comprehension whether the Dice and Jaccard indices outcomes are significant, a comparative analysis with other segmentations methods reported is depicted now. Garg *et al*. [60] demonstrated that their methodology proposed, active contour-based, produced a Dice coefficient of 1, while other researchers and authors such as Rehman *et al*. [61], region-based, or Khalid *et al*. [56], wavelet-based, reported a Dice value of 0.84 or 0.92, respectively. From this, it can be extracted that our values are not that far away from the current outcomes accomplished by other recognized researchers. Moreover, a founding of interest to mention is that the segmentation method proposed by Hasan *et al*. [59], which regards artificial neural networks, concludes a Dice coefficient of approximately 0.955 and Jaccard coefficient of 0.914. Then, remarking again the importance of the future evolution of machine learning is essential.

5.4 LIMITATIONS OF THE STUDY AND DIRECTIONS FOR FUTURE RESEARCH

The previous binary masks illustrated in *Figure 17* are only a bit taste of all the images segmented for the development of the project. That is, not all the images employed for the segmentation stage have been segmented that successfully. In this present section, what is purposed to show is what have been the different limitations this current study has been dealing to face. *Figure 18* aims to show some of the images in which the algorithms have mostly failed. The figure consists of the original image, the predicted binary masks obtained after the implementation of the level set method, the method combining GrabCut and k-means algorithms and the intensity thresh automatic algorithm, respectively, and the ground truth image.
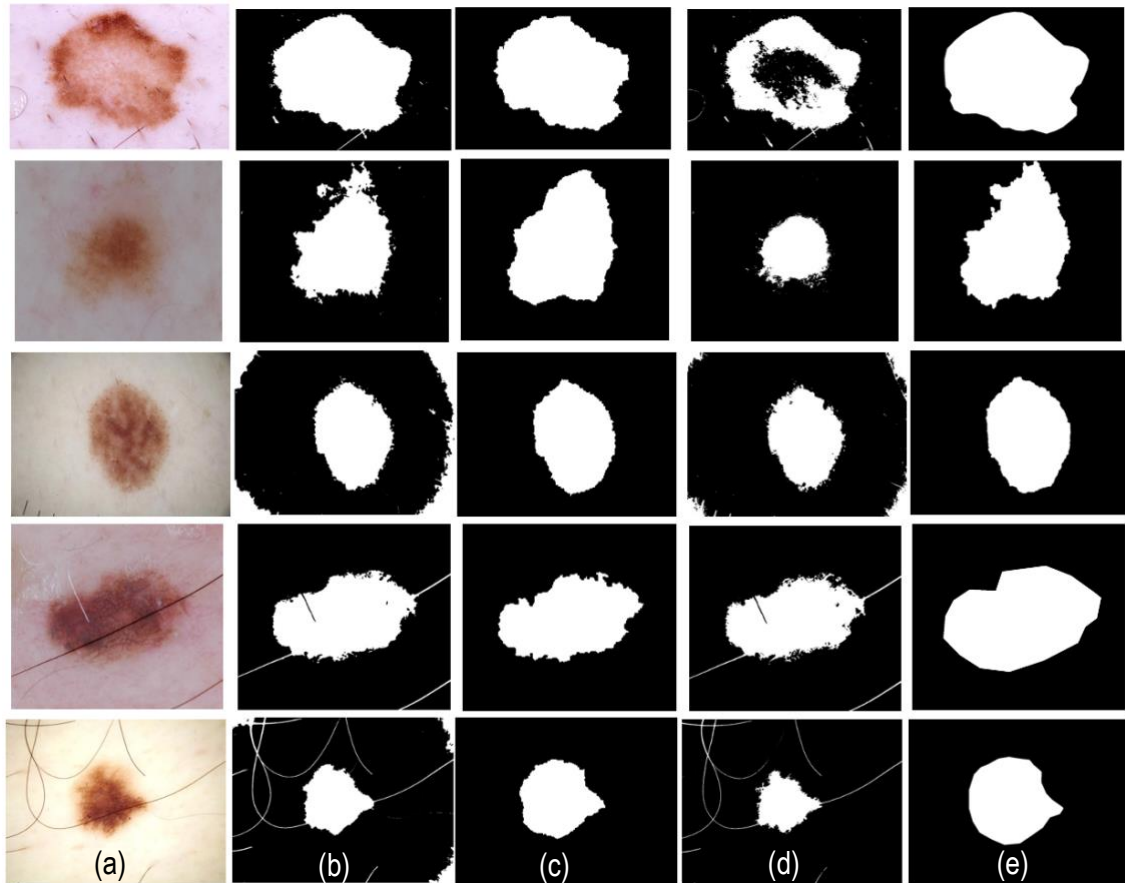
Figure 18. Binary masks obtained for each of the three methods proposed. (a) Original image [72], (b), binary mask obtained throughout level set method, (c) binary mask obtained throughout the combination of GrabCut and k-means algorithms, (d) binary mask obtained throughout MedicalSeg in-house developed algorithm (intensity thresh automatic) and (e) ground truth image [own source].

Some key aspects can be concluded from each of the five image cases presented in *Figure 18*. First of all, from Image 1 some differences among the three methods can be resolved; case (d) has failed in the segmentation stage since the core of the lesion is not detected, only the boundary pixels are considered as lesion region. What is more, case (b) and (d) again show some artefacts from the background that have also been labeled as lesion, which results in a misclassification.

Image 2 consists of a case of incorrect detection of the lesion boundaries. By simply contemplating at the image, it is effortlessly to fall within an erroneous result due to the similarity among the lesion colour and the normal skin one. Case (d) only detects the core of the lesion itself while case (b) comprises a wider area of the lesion. Finally, case (c) depicts the more similar result when comparing to ground truth. This fact could be explained due to the pre-processing step that the images undergo in case (c); the method combining GrabCut and k-means algorithms involves an illumination correction which enhances the colours, so the implementation of the algorithm is leveraged with more advantages when segmenting. Image 3 handles the same problem; the corners of the image are darker than the rest of it. Since case (c) is the only one that considers a correction of the illumination, the final binary image does not detect erroneously these corners as skin lesion region.

Image 4 is a greatly interesting case. Here it is presented an artifact dealt with formerly, hair. The original image presents the apparition of hair, which crosses the lesion entirely. When looking at the segmented images, it can be without a doubt mentioned the importance again of the pre-processing steps. Cases (b) and (d) show how the hair remains after the segmentation of the image, since it has not been removed. Then, when comparing these binary masks to the one who undergone hair removal, case (c), the necessity of pre-processing steps implementation creates a stronger sense. By examining visually this case, the final results beforehand depicted can be more easily comprehended. This same problem can be extrapolated to Image 5, which deals with the same obstacle, the hair. Case (b) and (d) again demonstrate how their performance is neither accurate nor reliable when having to face with those types of artifacts, while case (c) performs without any repercussion.

From these failed results, some different points can be concluded regarding the main limitations of skin lesion segmentation. First of all, it is key to mention the wide variability among lesions; each presents different colours, textures, shapes, etc. This variability makes it challenging to develop a single segmentation algorithm able to precisely and accurately handle any lesion type. Additionally, the same lesion at the same time could be acquired in so many different ways by the image acquisition technique; external factors such as room illumination, angle position of the dermoscope or whether there is hair presence may result in different dermoscopy images. Nevertheless, not only the lesion itself is the one causing an altered result but the lack of generalisation to different ethnicities might impact the algorithm's performance.

A second limit is the ambiguity in the boundary definition. The boundaries of a skin lesion may not be as well-defined as desired from the surrounding normal skin. This, in fact, can difficult the manual segmentation task of the physicians so the definition of the ground truth might not be rigorously determined. Linked to this point, another restriction which may challenge the correct validation of the segmentation results is the observer variability among dermatologists. Each one may interpret differently the boundaries of the lesion, which leads to dissimilarities in their manual annotations; a certain specialist may resolve a ground truth image, while others may come to a very distinct one.

Then, as has been visually assessed in *Figure 18*, it is a must to enhance the importance of the pre-processing steps. Skin lesions can be partially occluded by hair, clothes or any other structures that may introduce shadows or glare to the image. The binary masks acquired not only vary depending on the segmentation method under which they are being processed, but the previous stages they are subjected to.

Addressing these limitations demands ongoing research and future development, as the incorporation of more variated datasets (including wider differences among lesions and ethnicities), more robust algorithms and exhaustive clinical validation tools.

As a future research project, it could be demanding to perform the implementation of artificial intelligence tools as segmentation methods rather than the conventional ones. All the methodology described during the current study could be simplified and a greater efficiency and automatic manner might be accomplished. As prior mentioned in *Section 3.3.1*, the apparition of the recently developed Segment Anything artificial intelligence tool was tried to execute so its challenging performance could have been tested. In spite of this, due to the lack of validation of this particular tool by the experts and the lack of time, the idea has not been possible to successfully apply. For this reason, it might be considered a great research line to follow and future work continuation.

6. EXECUTION CRONOGRAM

The following section consists of the definition of the project tasks and milestones for the purpose of keeping track of the different phases and accomplishing the initial objectives in time. The proposed tasks have been broken down hierarchically by means of a Work Breakdown Structure (WBS) scheme. Furthermore, a precedence analysis and a GANTT diagram is presented as a way to clarify the project perspective and manage its timing.

6.1 WORK BREAKDOWN STRUCTURE

A Work Breakdown Structure (WBS) is a visual and hierarchical tool that deconstructs the project into smaller tasks in order to organise the team's work into manageable phases. It is a tree structure which shows consecutive subdivisions required to achieve the major aim of the project. WBS is developed by starting with the finale objective and successively dividing it into smaller assignments in terms of duration and responsibility [95].

For the case of the current project, the hierarchical decomposition of the total scope is presented in *Figure 19*.
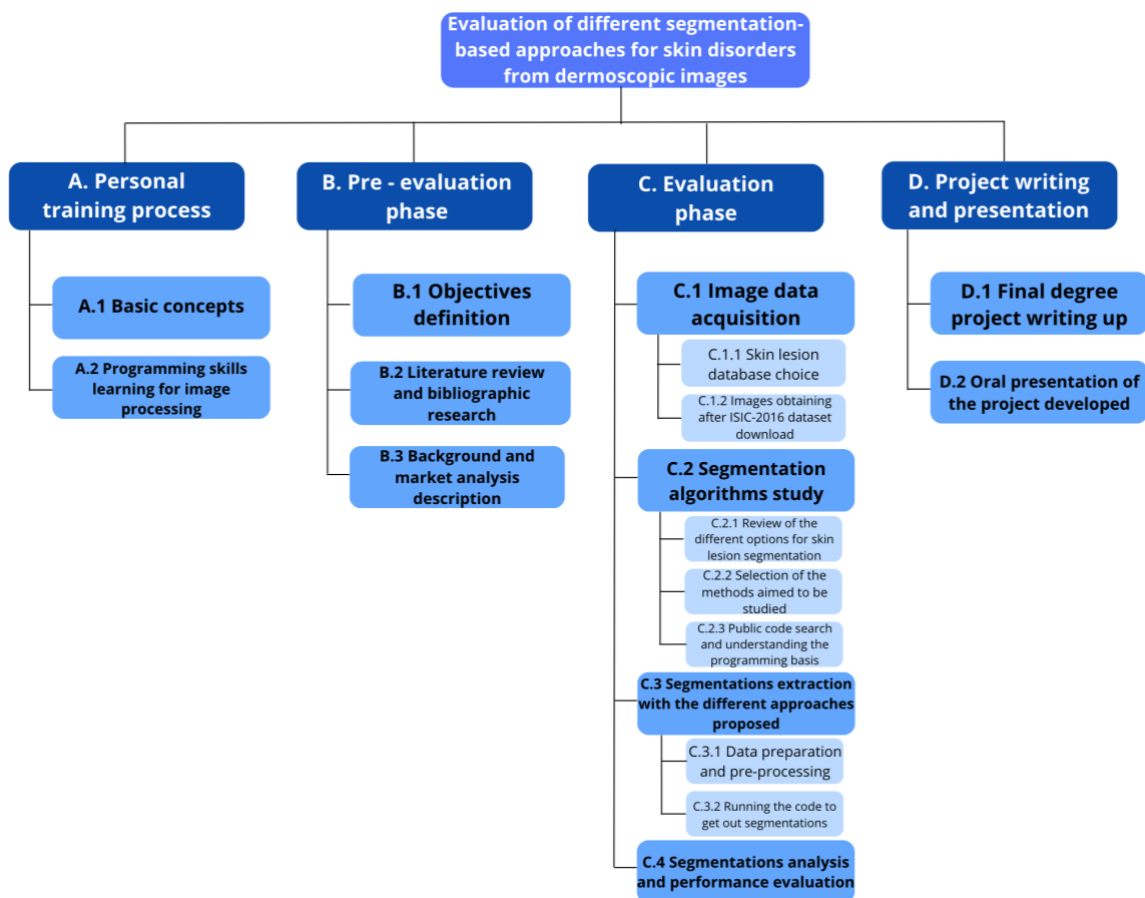
Figure 19. Work Breakdown Structure of the current project.

6.1.1 WORK BREAKDOWN STRUCTURE DICTIONARY

The following WBS dictionary includes a brief definition about the different tasks that must be conducted for the proper development and completion of the study.

A. Personal training process: the first task depicted is understanding the framework in which the project is located in order to address the problem as suited as possible.

> **A.1** Basic concepts: introduction to the main topic of the work, skin lesion segmentation, as a means of getting familiarised with basic notions.
>
> **A.2** Programming skills learning for image processing: introduction to the medical segmentation field through an intensive search on different software programmes and their libraries for image processing.

B. Pre - evaluation phase: the second phase of the project comprises those activities which are aimed to ease the familiarisation with the biomedical image processing field of study.

> **B.1** Objectives definition: description of the different objectives intended to be achieved at the end of the project.
>
> **B.2** Literature review and bibliographic research: investigation on different medical segmentation algorithms papers via free full-text archive of biomedical journal literature such as PubMed.
>
> **B.3** Background and market analysis description: analysis of the background of the project and the environment from which is going to be constructed.

C. Evaluation phase: this stage corresponds to the conduction of the project purposed to be evaluated. It consists in several tasks which are depicted next.

> **C.1** Image data acquisition: obtention of the database that is going to be employed for the extraction of the segmentations.
>
>> **C.1.1** Skin lesion database choice: the different options presented are studied and a final database is chosen after assessing its advantages and disadvantages compared to the other alternatives.
>>
>> **C.1.2** Images obtaining after *ISIC-2016* public dataset download: once the dataset is selected, the images are downloaded from the public repository.
>
> **C.2** Segmentation algorithms study: similar as in *Task C.1*, the different approaches reviewed in *Task B.2*, are considered and a final set of them is picked out in order to conduct the project development.

C.2.1 Review of the different options for skin lesion segmentation: analyse the different options found during the literature review and bibliographic research so the most appropriate algorithms are selected as a means of depicting significant results.

C.2.2 Selection of the methods aimed to be studied: three different pipelines are selected in order to conduct the project. Each algorithm is based on different segmentation approaches (intensity-based, active contours-based and combined region-based and clustering-based) in order to properly determine which could be more properly for dermoscopic images. depending on the final one chosen, the usage of whether one programming language or another is also ascertained.

C.2.3 Public code search and understanding the programming basis: the algorithms proposed are searched in public repositories and a comprehensive study of the methodology followed is fulfilled.

C.3 Segmentations extraction with the different approaches proposed: the binary masks are acquired after following the next tasks.

C.3.1 Data preparation and pre-processing: a bit taste of the images downloaded from the database ISIC-2016 is selected since the archive contains up to 900 images. In order to facilitate the management of the data, 48 images are taken.

C.3.2 Running the code to get out segmentations: the segmentation process itself is executed so the masks of the skin lesion are finally obtained for a posteriori evaluation of the algorithm performance.

C.4 Segmentations analysis and performance evaluation: evaluation on how the segmented image and the ground truth image are alike by means of the determination and employment of different performance indices.

D. Project writing and presentation: the final task conducted consists of the development of the writing of the study and the completion of a brief presentation.

D.1 Final degree project writing up: writing of the final project report.

D.2 Oral presentation of the project developed: preparation and realisation of an oral presentation to depict all the work carried out during months.

6.2 PRECEDENCE ANALYSIS

In this section, *Table 5* presents the dependencies between the tasks mentioned in *Section 7.1.1*. An approximate duration of the tasks has also been included in the table.

ACTIVITY	NAME	ANTECEDENT	DURATION
A.1	Basic concepts	-	7 days
A.2	Programming skills learning for image processing	-	14 days
B.1	Objectives definition	-	3 days
B.2	Literature review and bibliographic research	-	21 days
B.3	Background and market analysis description	-	14 days
C.1.1	Skin lesion database choice	B.2	2 days
C.1.2	Images obtaining after public dataset download	C.1.1	1 day
C.2.1	Skin lesion segmentation options review	B.2	7 days
C.2.2	Selection of the methods aimed to study	C.2.1	1 days
C.2.3	Code search and understanding the programming basis	C.2.2	21 days
C.3.1	Data preparation and pre-processing	C.1.2	2 days
C.3.2	Running the code to get out segmentations	C.2.3/C.3.1	20 days
C.4	Segmentations analysis and performance evaluation	C.3.2	21 days
D.1	Final degree project writing up	A/B/C	4 months
D.2	Oral presentation of the project developed	D.1	1 day

Table 5. Precedence analysis of the current project.

6.3 GANTT DIAGRAM

Once all the tasks and sub-tasks have been defined for the achievement of the main objective, a GANTT diagram has been created to assess the deadline for each phase. GANTT diagram is a visual chart that illustrates the timing of the project tasks. *Figure 20* depicts the GANTT chart of the project.

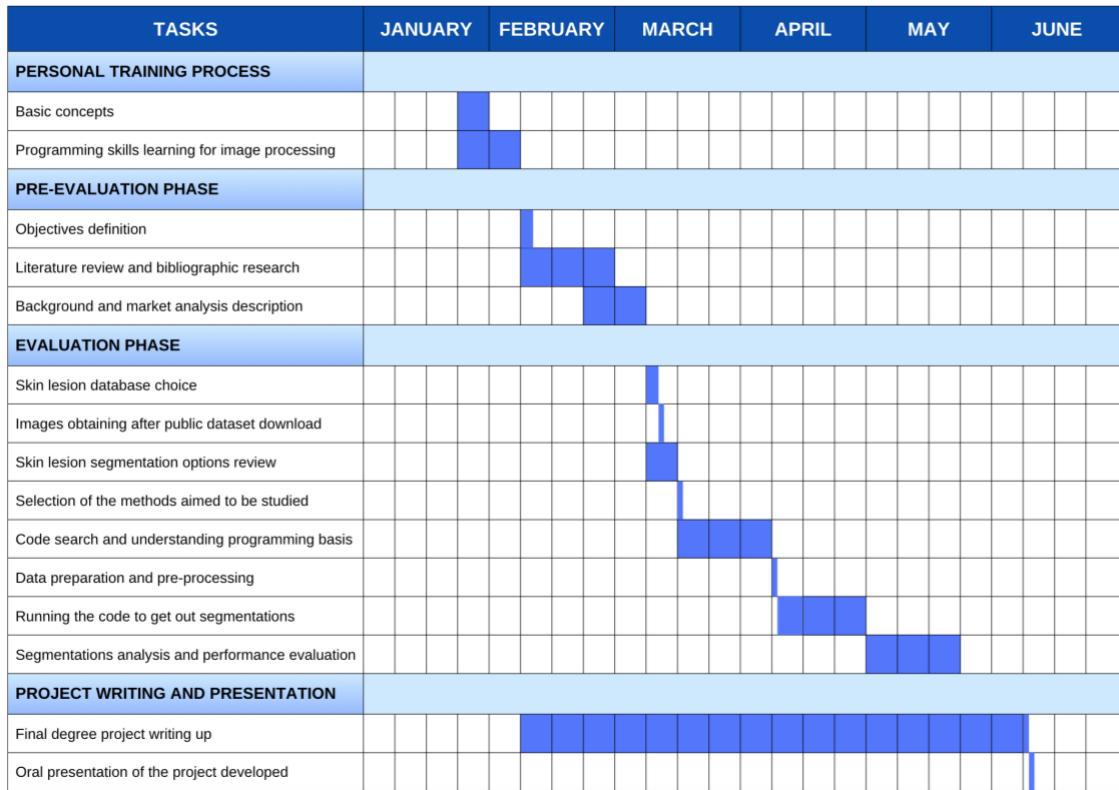| TASKS	JANUARY	FEBRUARY	MARCH	APRIL	MAY	JUNE
PERSONAL TRAINING PROCESS						
Basic concepts	▓					
Programming skills learning for image processing	▓					
PRE-EVALUATION PHASE						
Objectives definition		▓				
Literature review and bibliographic research		▓▓				
Background and market analysis description		▓▓				
EVALUATION PHASE						
Skin lesion database choice			▓			
Images obtaining after public dataset download			▓			
Skin lesion segmentation options review			▓			
Selection of the methods aimed to be studied			▓			
Code search and understanding programming basis			▓▓			
Data preparation and pre-processing				▓		
Running the code to get out segmentations				▓▓		
Segmentations analysis and performance evaluation					▓▓	
PROJECT WRITING AND PRESENTATION						
Final degree project writing up		▓▓▓▓	▓▓▓▓	▓▓▓▓	▓▓▓▓	
Oral presentation of the project developed						▓

Figure 20. GANTT diagram for the current project.

7. TECHNICAL VIABILITY

A detailed analysis of the project technical viability is carried out in the present section. Its feasibility can be assessed through a SWOT analysis, a strategic planning technique for evaluating four different aspects: strengths, weaknesses, opportunities and threats. It is designed to identify from a positive and negative point of view which are the major factors that affect the work. Bearing in mind the status of the project can help in the planning of a further strategy. *Table 6* presents de SWOT analysis performed.

STRENGTHS	WEAKNESSES
Internal characteristics	
- Previous open-source softwares knowledge and experience. - Potential emerging sector. - Non-invasive approach. - Versatility.	- Dataset variability. - Consensus lack on evaluation metrics. - No previous experience with segmentation models.
OPPORTUNITIES	THREATS
External characteristics	
- Development of medical applications for diagnosis purposes.	- Ethical concerns. - Overreliance.

Table 6. SWOT analysis of the project.

7.1 STRENGHTS

When it comes to the strengths of the project, it could be advantageously taken the fact that there is a previous knowledge over the Python and MATLAB softwares employed. Those programming languages have been worked on over the past three degree courses so an intermediate understanding of the basic functions were a priori known. This fact has permitted to reach the aim of the project presented in *Section 1*. If there was no previous learning whether programming or artificial intelligence field, the development of the project could have been affected.

The future market prospective for skin lesion segmentation, as depicted in *Section 3*, is presented as an in emerging field which may contribute and promote several innovative advancements. It is obvious that what lies ahead in the medical field is an enormous progression in health technologies and machine learning algorithms in order to assist clinicians regarding clinical examination and diagnosis performance. Along the same line, segmentation tools not only have helped to develop efficient and feasible diagnosis algorithms but they are non-invasive, which ensures patient's safety and reduces risk of complications.

Last of all, it is without a doubt ensured that the three methodologies described in this project are a small taste compared to the enormous quantity of medical segmentation algorithms developed

over the last years. Depending on the input images got, a segmentation-based approach or another can be implemented to perform the obtention of the binary masks. If the input skin lesion image has, for example, overlapping boundaries, an intensity segmentation method may not be the most properly. In summary, skin segmentation algorithms can easily adapt and face possible weaknesses by means of constant changes.

7.2 WEAKNESESS

Regarding the weaknesses the project must face, the most challenging task in medical image analysis is the variability among the dataset, this is, the variability in image resolution, lighting conditions, image acquisition methodology, etc. Not all images are captured following the same protocol, so it should not be somewhat unexpected to be facing those issues. What is more, skin lesions appear to be different in colour, size, texture and shape, what makes it even more difficult for the algorithm to accurately segment the region of interest. Not only this problem might be faced, but the fact that there is no a standardised concord respect to evaluation metrics for skin lesion segmentation is the cause of the main limitation when trying to compare the segmentation outcomes from different approaches.

Lastly, prior to the development of the project, the student had slightly been in contact with segmentation approaches so the knowledge regarding this field has had to be constructed during several weeks for the resolution of the different tasks.

7.3 OPORTUNITIES

The main opportunity that could be extracted from this project is its potential application in the near future in clinical examination and diagnosis. It would be a powerful tool an artificial intelligence application which might be able to automatically perform the segmentation of a patient's skin lesion and its classification, whether benign or malign. The aim is to develop a tool to assist clinicians in order to facilitate diagnosis. It could be interesting not only to consider the visual examination, such as the assessment of the size or shape, but to determine a treatment plan when accurately segmenting the region of interest. Dermatologists might have more facilities when treating, once they know which is the target area, or when excising in a surgery. In addition to these two medical applications, the fact that the evolution of the skin lesion could be monitored along time is another robust purpose for which a deeper investigation should be set out about when developing the app. Finally, the education and research fields regard to confront the resolution of new datasets and machine learning algorithms as a means of building up more accurate and efficient automated diagnostic approaches.

7.4 THREATS

The most critical constraint about the threats of the project is the ethical concerns regarding the patients privacy. Medical images include sensitive information related to patient's clinical health

information so the potential risk of its spreading could conclude in terrible consequences for the patient such as discrimination. This is the reason why a public dataset has been employed for the segmentation methods assessment as a means of avoiding confidentiality concerns. Any paediatric image from Hospital Sant Joan de Déu was taken in order to protect children from clinical invasion.

To conclude, it might be interesting to comment not as seen as should limitation of skin lesion segmentation. The existing risk of overreliance can lead to neglect other important aspects present in the image while carrying out the diagnosis. Recently, it has been expected that artificial intelligence will perform more accurately than humans, which results in greater reliance on computational algorithms so the opinion clinicians or dermatologists may not be as contemplated as should. What this issue originates are missed opportunities and incorrect conclusions in the long run.

8. ECONOMIC VIABILITY

Evaluating which is the resulting economic impact of the project may allow its correct managing and completion successfully. An estimation of the hypothetical total cost and budget associated with its development is next depicted in the above section. The budget is divided into two categories: material costs and human costs.

8.1 SOFTWARE

The project conducted was accomplished by the implementation of two open-source softwares—Python and MATLAB. The image processing and manipulation as well as the final evaluation of the three different methodologies selected did not entail any sort of economic investment since Python's software is freely downloaded while MATLAB's access has been accomplished due to a personal academic license provided by *Universitat de Barcelona*. Moreover, regarding the ISIC-2016 database utilised for the skin lesion segmentation validation, the acquisition of the data did not take any additional expense since it is a public dataset available for whom interested in medical imaging. Last but not least, Microsoft Office® 365 services were used for the writing of the final degree project and its oral presentation, specifically Microsoft Word and Microsoft PowerPoint respectively.

8.2 HARDWARE

No hardware material was employed or bought specifically for the resolution of the project. So this cost is neglected in the final economic evaluation. Only one expense may be considered for this section, which is the needed of a computer for the carry-on of the literature review, segmentation and validation performance, and final writing of the report. For this purpose, a laptop has been employed during the 17 weeks duration of the project because of practical reasons since it has not been carried out in a fixed position.

8.3 HUMAN RESOURCES

When it comes to human resources, the cost associated with the biomedical engineer student to complete the tasks previously depicted for the project's development is presented now. In order to estimate the total human cost, it was considered an approximate time inverted by the student and an estimated salary per hour for its work. The average salary weighted for the undergraduate is calculated to be about 10 € per hour. Assuming a workday of 5 hours, excluding weekends, and a project lasting 17 weeks approximately, the total amount of work hours invested in the elaboration of the research is estimated to be 425 hours. This results in a total cost of 4,250 €.

8.4 TOTAL COST OF THE PROJECT

The different costs of the project previously described can be assessed in *Table 7*, and presents an estimation of the overall budget of the resolution and implementation of the segmentation project.

	UNIT/HOUR COST (€)	UNIT/HOURS	TOTAL
HARWARE			
MacBook Pro11,1	1,299 €	1	1,299 €
SUBTOTAL			**1,299 €**
SOFTWARE			
Python software	-	1	-
MATLAB software	-	1	-
Microsoft Office® 365	-	1	-
SUBTOTAL			-
HUMAN RESOURCES			
Biomedical engineering student	10	425	4,250 €
SUBTOTAL			**4,250 €**
TOTAL			**5,549 €**

Table 7. Total project cost table.

9. REGULATION AND LEGAL ASPECTS

The data used in this project comes from a public dataset in order to avoid those issues related to privacy and legal considerations. It was aimed to evaluate how different segmentation algorithms performed to segment dermatological disorders so it was considered to employ images which have may not end up having any juridical issues. Although the problem is not present in this work, it is fundamental to assess which are the regulatory aspects one must consider if the work were transformed as a commercially available tool for clinicians.

That is, if hypothetically any kind of medical application was launch to the market in a way to assist clinicians work, some regulatory and legal aspects might have to be taken into account. In this section, it is going to be explained which would be the different steps that must be followed up in order to convert this segmentation study into a market product within European Union regulation.

The United States Food and Drug Administration (FDA) [96] is responsible for protecting the public health by regulating food, drugs, cosmetics, medical devices, biological and blood products. It aims to ensure safety, efficacy and security of human beings. Nonetheless, the regulation and legislation aspects that the application would need to meet up are the ones under the European Union laws. The EU Medical Device Regulation (EU MDR) [97] replaced the European Union Medical Devices Directive (EU MDD) in a way to impose stricter regulatory requirements to be fulfilled by the medical devices.

EU MDR intent is to ensure a high standard of safety and quality for medical devices that are produced in, or supplied to, any country of the European Union. The regulatory framework is meant to ensure health and safety while encouraging innovation. Medical devices that incorporate algorithms must also be subjected to regulatory and legislation rules such as other devices, and they should undergo any kind of evaluation procedure to ensure they meet the requirements. That procedure has to involve a detailed assessment of the algorithm's performance and safety.

EU MDR legislation is the most stringent legal document regarding Medical Device Software (MDSW), but this is not the only law or guidance document that might be considered when launching any medical device. Since the device collects and processes personal data, the European Union's General Data Protection Regulation (GDPR) [98] must be cautiously considered and examined. The GDPR main objective is to enhance individuals rights over their personal data.

Various guidance documents have been released to provide clarity and guidance on the implementation of medical device regulations in the European Union (EU). Listed now, the main documents which must be kept in mind are depicted.

- Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices.
- MDCG 2019-11: Guidance on Qualification and Classification of Software (MDSW).

- MDCG 2020-1: Guidance on Clinical Evaluation (MDR) / Performance Evaluation (IVDR) of Medical Device Software.
- MDCG 2019-16: Guidance on Cybersecurity for medical devices.
- Regulation (EU) 2017/746 of the European Parliament and of the Council of 5 April 2017 on in vitro diagnostic medical devices and repealing (IVDR).
- Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on General Data Protection Regulation (GDPR).

The incoming Regulation (EU) 2017/745 presents some challenges for medical device manufacturers. Regarding software medical devices, changes to classification have recently been creating some concerns; nearly all software devices are likely to be class IIa or III (the most stringent categories), while previously they were classified as class I. The cause of this up-classification is the apparition of Rule 11. Its drafting allocates some MDSW to a higher class. The aim of this rule is to manage the information given by the device to infuse a clinical decision. Rule 11 states [99]:

Software intended to provide information which is used to take decisions with diagnosis or therapeutic purposes is classified as class IIa, except if such decisions have an impact that may cause:

- *death or an irreversible deterioration of a person's state of health, in which case it is in class III; or*
- *a serious deterioration of a person's state of health or a surgical intervention, in which case it is classified as class IIb.*

Software intended to monitor physiological processes is classified as class IIa, except if it is intended for monitoring of vital physiological parameters, where the nature of variations of those parameters is such that it could result in immediate danger to the patient, in which case it is classified as class IIb.

All other software is classified as class I.

Most software medical devices recently developed are intended to provide information used to take decisions with a diagnostic or therapeutic purpose. Software can perform an innocuous function, but if the final decision can result in death or harm to the patient, the device is classified as class III.

10. CONCLUSION AND FUTURE LINES

Segmenting skin lesions still remains on the major issues to tackle when dealing with dermatologic screenings. In the current project, three segmentation algorithms have been tested in order to characterise their execution. The main objectives established when addressing the problem laid out have been successfully accomplished; after an intensive bibliographic review, the most challenging algorithms were selected, so they could be implemented and validated with the aim of ascertaining their main capabilities when facing skin lesions presented in a wide variety of shapes and sizes.

After assessing the quantitative analysis stage, by means of the performance indices, some key points were determined: firstly, it is a must to highlight the importance of the pre-processing step. It has been without a doubt witnessed that the previous step to the segmentation process is broadly an indispensable require so the most rigorous results are achieved. The fact that skin lesion image acquisition takes commonly place via clinical photography or through a dermoscope, relies on the necessity to cautiously consider illumination correction. Further to this, these kind of lesions are not uncommon to present hair on account of where they are located. It has been demonstrated the efficacy of hair removal algorithms which facilitate the detection of the ROI.

One additional point it would be liked to remark is the importance of a further study and deep in investigation about the current filed. Trying to implement this kind of tools must be with due precautions done since they aim to deal with populations health. That is, the work not simply tries to validate how different algorithms work under different circumstances but it is thought of to be used for clinical training to assist physicians. Accounting to the fact that the tool might handle severe cases, the segmentation results obtained must be as exact as possible so the best treatment options can be facilitated to the patient.

Regarding the three segmentation methods themselves, it has been clearly demonstrated the usefulness of each case in whether one application or other. Nevertheless, a few images were not that smoothly segmented. As mentioned several times, skin lesions are presented in a wide range of aspects; even though an image can exhibit the boundaries of the lesion perfectly delimited, some other disorders are difficult to be differentiated from normal skin, including for clinicians. No specific segmentation-based method will under any circumstances perform better than others since it depends on so many image factors. That is the reason why, for example, methods based on regions may segment greater when facing images with clear colour differences since they group pixels with similar characteristics. Another example yet is the case of intensity-based; it might provide outperformance results but when handling with inconsistent illumination conditions, the method will probably end in struggling.

All in all, it could be said that conventional segmentation techniques are a promising tool that should not be underestimated taken and, when combined with other approaches, the predicted binary

masks achieved could probably end in with even more precise and accurate outcomes. Mentioning once again the importance of considering artificial tools as a following line to this project is highlighted. The presented ANNs for medical image segmentation in the state-of-the-art section could not be make usage for due to computational problems. The lack of validation by the clinical experts from Hospital Sant Joan de Déu has difficulted the task of implementing machine learning tools in the current study so a wider and exhaustive review regarding them has unfortunately not been accomplished. Therefore, it is considerable to think of as a significant future work which aims to employ AI tools to fulfil the process in a more efficient and automatic way.

Last but not least, it is quite essential to talk about my personal experience throughout the carry-on of the project. It has been extremely fulfilling for my growth; considering first the academic side, I have been able to learn and extend my knowledge regarding image segmentation tools, a topic not widely dealt with during the four degree courses. It has been advantageously and favourable for me to understand the main basis upon which computational algorithms lay, a promising field within biomedical engineering future. On the personal side, I would like to mention that I have been able to successfully handle the project's pressure and I have concluded in meaning results which could further be employed in future clinical applications. What is more, I have been capable of scheduling myself well and accomplishing the goals established every week. The satisfaction level got after the fulfilment of the project is beyond all expectations. In spite of that, if I had to comment any unfavourable aspect, I would talk about my personal preference to work over a clinical database of Hospital Sant Joan de Déu rather than a public dataset.

11. BIBLIOGRAPHY

[1] Richard MA, Paul C, Nijsten T, Gisondi P, Salavastru C, Taieb C, Trakatelli M, Puig L, Stratigos A; EADV burden of skin diseases project team. Prevalence of most common skin diseases in Europe: a population-based study. J Eur Acad Dermatol Venereol. 2022 Jul;36(7):1088-1096. doi: 10.1111/jdv.18050. Epub 2022 Mar 22. PMID: 35274366; PMCID: PMC9415115.

[2] *Home page: Journal of the American academy of dermatology.* Jaad.org. https://www.jaad.org/

[3] Goceri E. Deep learning based classification of facial dermatological disorders. Comput Biol Med. 2021 Jan;128:104118. doi: 10.1016/j.compbiomed.2020.104118. Epub 2020 Nov 13. PMID: 33221639.

[4] Abbas Q, Garcia IF, Emre Celebi M, Ahmad W, Mushtaq Q. A perceptually oriented method for contrast enhancement and segmentation of dermoscopy images. Skin Res Technol. 2013 Feb;19(1):e490-7. doi: 10.1111/j.1600-0846.2012.00670.x. Epub 2012 Aug 13. PMID: 22882675.

[5] Liu, Y., Jain, A., Eng, C. *et al.* A deep learning system for differential diagnosis of skin diseases. *Nat Med* 26, 900–908 (2020). https://doi.org/10.1038/s41591-020-0842-3

[6] Richard, M. A., Paul, C., Nijsten, T., Gisondi, P., Salavastru, C., Taieb, C., Trakatelli, M., Puig, L., & Stratigos, A. (2022). Prevalence of most common skin diseases in Europe: a population-based study. *Journal of the European Academy of Dermatology and Venereology*, *36*(7), 1088–1096. https://doi.org/10.1111/JDV.18050

[7] Linares MA, Zakaria A, Nizran P. Skin Cancer. Prim Care. 2015 Dec;42(4):645-59. doi: 10.1016/j.pop.2015.07.006. PMID: 26612377.

[8] Langley RG, Krueger GG, Griffiths CE. Psoriasis: epidemiology, clinical features, and quality of life. Ann Rheum Dis. 2005 Mar;64 Suppl 2(Suppl 2):ii18-23; discussion ii24-5. doi: 10.1136/ard.2004.033217. PMID: 15708928; PMCID: PMC1766861.

[9] Berke R, Singh A, Guralnick M. Atopic dermatitis: an overview. Am Fam Physician. 2012 Jul 1;86(1):35-42. PMID: 22962911.

[10] Frazier W, Bhardwaj N. Atopic Dermatitis: Diagnosis and Treatment. Am Fam Physician. 2020 May 15;101(10):590-598. PMID: 32412211.

[11] Bareiro Paniagua, Laura Raquel, Leguizamón Correa, Deysi Natalia, Pinto-Roa, Diego P., Vázquez Noguera, José Luis, & Salgueiro Toledo, Lizza A. (2016). Computerized Medical Diagnosis of Melanocytic Lesions based on the ABCD approach. *CLEI Electronic Journal*, *19*(2), 6.

[12] Davis LE, Shalin SC, Tackett AJ. Current state of melanoma diagnosis and treatment. Cancer Biol Ther. 2019;20(11):1366-1379. doi: 10.1080/15384047.2019.1640032. Epub 2019 Aug 1. PMID: 31366280; PMCID: PMC6804807.

[13] Tímár J, Ladányi A. Molecular Pathology of Skin Melanoma: Epidemiology, Differential Diagnostics, Prognosis and Therapy Prediction. Int J Mol Sci. 2022 May 11;23(10):5384. doi: 10.3390/ijms23105384. PMID: 35628196; PMCID: PMC9140388.

[14] Melanoma skin cancer - Treatment. Nhs.uk. https://www.nhs.uk/conditions/melanoma-skin-cancer/treatment/

[15] Dermoscopy Atlas. http://www.dermoscopyatlas.com/ (2007)

[16] Bodman MA, Al Aboud AM. Melanocytic Nevi. [Updated 2022 Jul 18]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2023 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK470451/

[17] Ferrara G, Gianotti R, Cavicchini S, Salviato T, Zalaudek I, Argenziano G. Spitz nevus, Spitz tumor, and spitzoid melanoma: a comprehensive clinicopathologic overview. Dermatol Clin. 2013 Oct;31(4):589-98, viii. doi: 10.1016/j.det.2013.06.012. PMID: 24075547.

[18] Krakowski AC, Hafeez F, Westheim A, Pan EY, Wilson M. Advanced basal cell carcinoma: What dermatologists need to know about diagnosis. J Am Acad Dermatol. 2022 Jun;86(6S):S1-S13. doi: 10.1016/j.jaad.2022.03.023. PMID: 35577405.

[19] Wong CS, Strange RC, Lear JT. Basal cell carcinoma. BMJ. 2003 Oct 4;327(7418):794-8. doi: 10.1136/bmj.327.7418.794. PMID: 14525881; PMCID: PMC214105.

[20] Waldman A, Schmults C. Cutaneous Squamous Cell Carcinoma. Hematol Oncol Clin North Am. 2019 Feb;33(1):1-12. doi: 10.1016/j.hoc.2018.08.001. PMID: 30497667.

[21] Karadag AS, Parish LC. The status of the seborrheic keratosis. Clin Dermatol. 2018 Mar-Apr;36(2):275-277. doi: 10.1016/j.clindermatol.2017.09.011. Epub 2017 Sep 8. PMID: 29566932.

[22] So N, Waldman R, Waldman S. Professionalism of clinical photography in the pediatric setting. Curr Probl Pediatr Adolesc Health Care. 2019 Apr;49(4):74-78. doi: 10.1016/j.cppeds.2019.03.007. Epub 2019 Apr 11. PMID: 30981457.

[23] Barut C, Ertilav H. Guidelines for standard photography in gross and clinical anatomy. Anat Sci Educ. 2011 Nov-Dec;4(6):348-56. doi: 10.1002/ase.247. Epub 2011 Jul 27. PMID: 21796798.

[24] Schaefer G, Rajab MI, Celebi ME, Iyatomi H. Colour and contrast enhancement for improved skin lesion segmentation. Comput Med Imaging Graph. 2011 Mar;35(2):99-104. doi: 10.1016/j.compmedimag.2010.08.004. Epub 2010 Oct 28. PMID: 21035303.

[25] Ko RF, Smidt AC, Durkin JR. Reflectance confocal microscopy in pediatric dermatology: A state-of-the-art review. Pediatr Dermatol. 2021 Nov;38(6):1488-1499. doi: 10.1111/pde.14837. Epub 2021 Oct 14. PMID: 34651341.

[26] Mazzeo M, Manfreda V, Botti E, Bianchi L. The Role of Reflectance Confocal Microscopy on Diagnosis of Melanoacanthoma. Actas Dermosifiliogr (Engl Ed). 2021 Apr 14:S0001-7310(21)00137-X. English, Spanish. doi: 10.1016/j.ad.2019.07.025. Epub ahead of print. PMID: 33864795.

[27] Catalano O, Roldán FA, Varelli C, Bard R, Corvino A, Wortsman X. Skin cancer: findings and role of high-resolution ultrasound. J Ultrasound. 2019 Dec;22(4):423-431. doi: 10.1007/s40477-019-00379-0. Epub 2019 May 8. PMID: 31069756; PMCID: PMC6838298.

[28] Umaa Mageswari, S., Sridevi, M., & Mala, C. (2013). An Experimental Study and Analysis of Different Image Segmentation Techniques. *Procedia Engineering*, *64*, 36–45. https://doi.org/10.1016/J.PROENG.2013.09.074

[29] Bong, C. W., Liew, C. C., & Lam, H. Y. (2016). Ground-glass opacity nodules detection and segmentation using the snake model. *Bio-Inspired Computation and Applications in Image Processing*, 87–104. https://doi.org/10.1016/B978-0-12-804536-7.00005-3

[30] Ooi AZH, Embong Z, Abd Hamid AI, Zainon R, Wang SL, Ng TF, Hamzah RA, Teoh SS, Ibrahim H. Interactive Blood Vessel Segmentation from Retinal Fundus Image Based on Canny Edge Detector. Sensors (Basel). 2021 Sep 24;21(19):6380. doi: 10.3390/s21196380. PMID: 34640698; PMCID: PMC8512020.

[31] Comparing Edge Detection Methods. Retrieved from https://medium.com/@nikatsanka/comparing-edge-detection-methods-638a2919476e

[32] Rogowska, J. (2009). Overview and fundamentals of medical image segmentation. Handbook of Medical Image Processing and Analysis, 73–90. https://doi.org/10.1016/B978-012373904-9.50013-1

[33] Niu, Zuodong & Li, Handong. (2019). Research and analysis of threshold segmentation algorithms in image processing. Journal of Physics: Conference Series. 1237. 022122. 10.1088/1742-6596/1237/2/022122.

[34] Giraldi, G. A., Marturelli, L. S. GRADIENT VECTOR FLOW MODELS FOR BOUNDARY EXTRACTION IN 2D IMAGES.

[35] Nascimento, J. C., Marques, J. S. (2005). Adaptive Snakes Using the EM Algorithm. IEEE TRANSACTIONS ON IMAGE PROCESSING. https://doi.org/10.1109/TIP.2005.857252

[36] Bora, Dibya & Gupta, Anil. (2015). A Novel Approach Towards Clustering Based Image Segmentation.

[37] Ahmed, M.; Seraj, R.; Islam, S.M.S. The k-means Algorithm: A Comprehensive Survey and Performance Evaluation. Electronics 2020, 9, 1295. https://doi.org/10.3390/electronics9081295

[38] James C. Bezdek, Robert Ehrlich, William Full, FCM: The fuzzy c-means clustering algorithm, Computers & Geosciences, Volume 10, Issues 2–3, 1984, Pages 191-203, ISSN 0098-3004, https://doi.org/10.1016/0098-3004(84)90020-7

[39] Anello, E. (2023). K-Means Clustering in R Tutorial. https://www.datacamp.com/tutorial/k-means-clustering-r

[40] Oliveira, R. B., Filho, M. E., Ma, Z., Papa, J. P., Pereira, A. S., & Tavares, J. M. R. S. (2016). Computational methods for the image segmentation of pigmented skin lesions: A review. *Computer Methods and Programs in Biomedicine*, *131*, 127–141. https://doi.org/10.1016/J.CMPB.2016.03.032

[41] Anwar SM, Majid M, Qayyum A, Awais M, Alnowami M, Khan MK. Medical Image Analysis using Convolutional Neural Networks: A Review. J Med Syst. 2018 Oct 8;42(11):226. doi: 10.1007/s10916-018-1088-1. PMID: 30298337.

[42] Xiao-Xia Yin, Le Sun, Yuhan Fu, Ruiliang Lu, Yanchun Zhang, "U-Net-Based Medical Image Segmentation", *Journal of Healthcare Engineering*, vol. 2022, Article ID 4189781, 16 pages, 2022. https://doi.org/10.1155/2022/4189781

[43] Badrinarayanan, V., Kendall, A.; Cipolla, R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. http://mi.eng.cam.ac.uk/projects/segnet/

[44] Chen, L., Papandreou, G., Kokkinos, I., Murphy, K.P., & Yuille, A.L. (2016). DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. IEEE Transactions on Pattern Analysis and Machine Intelligence, 40, 834-848.

[45] Krizhevsky, A., Sutskever, I., Hinton, G. E. ImageNet Classification with Deep Convolutional Neural Networks.

[46] Simonyan, K., & Zisserman, A. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition. CoRR, abs/1409.1556.

[47] Adeyinka, A.A., Viriri, S. (2018). Skin Lesion Images Segmentation: A Survey of the State-of-the-Art. In: Groza, A., Prasath, R. (eds) Mining Intelligence and Knowledge Exploration. MIKE 2018. Lecture Notes in Computer Science(), vol 11308. Springer, Cham. https://doi.org/10.1007/978-3-030-05918-7_29

[48] Abbas Q, Fondón I, Rashid M. Unsupervised skin lesions border detection via two-dimensional image analysis. Comput Methods Programs Biomed. 2011 Dec;104(3):e1-15. doi: 10.1016/j.cmpb.2010.06.016. Epub 2010 Jul 21. PMID: 20663582.

[49] Wong, A., Scharcanski, J., Member, S., & Fieguth, P. (2011). Automatic Skin Lesion Segmentation via Iterative Stochastic Region Merging.

[50] R. Garnavi, M. Aldeen, M.E. Celebi, G. Varigos, S. Finch, Border detection in dermoscopy images using hybrid thresholding on optimized color channels, Comput. Med. Imaging Graph. 35 (2011) 105–115.

[51] Abbas, Q., Celebi, M. E., & García, I. F. (2012). A novel perceptually-oriented approach for skin tumor segmentation. International Journal of Innovative Computing, Information and Control, 8(3 A), 1837–1848.

[52] Norton, K. A., Iyatomi, H., Celebi, M. E., Ishizaki, S., Sawada, M., Suzaki, R., Kobayashi, K., Tanaka, M., & Ogawa, K. (2012). Three-phase general border detection method for dermoscopy images using non-uniform illumination correction. Skin Research and Technology, 18(3), 290–300. https://doi.org/10.1111/J.1600-0846.2011.00569.X

[53] Huiyu Zhou, Xuelong Li, Gerald Schaefer, M. Emre Celebi, Paul Miller, Mean shift based gradient vector flow for image segmentation, Computer Vision and Image Understanding, Volume 117, Issue 9, 2013, Pages 1004-1016, ISSN 1077-3142, https://doi.org/10.1016/j.cviu.2012.11.015.

[54] Cavalcanti PG, Scharcanski J. A coarse-to-fine approach for segmenting melanocytic skin lesions in standard camera images. Comput Methods Programs Biomed. 2013 Dec;112(3):684-93. doi: 10.1016/j.cmpb.2013.08.010. Epub 2013 Aug 27. PMID: 24075079.

[55] Ma, Zhen & Tavares, Joao. (2014). Segmentation of Skin Lesions Using Level Set Method. 228-233. 10.1007/978-3-319-09994-1_20.

[56] Khalid S, Jamil U, Saleem K, Akram MU, Manzoor W, Ahmed W, Sohail A. Segmentation of skin lesion using Cohen-Daubechies-Feauveau biorthogonal wavelet. Springerplus. 2016 Sep 19;5(1):1603. doi: 10.1186/s40064-016-3211-4. PMID: 27652176; PMCID: PMC5028360.

[57] Andrea Pennisi, Domenico D. Bloisi, Daniele Nardi, Anna Rita Giampetruzzi, Chiara Mondino, Antonio Facchiano, Skin lesion image segmentation using Delaunay Triangulation for melanoma detection, Computerized Medical Imaging and Graphics, Volume 52, 2016, Pages 89-103, ISSN 0895-6111, https://doi.org/10.1016/j.compmedimag.2016.05.002

[58] Sakthi, S.M.Jai & Palaniappan, Mirualini & Aravindan, Chandrabose. (2018). Automated Skin Lesion Segmentation of Dermoscopic Images using GrabCut and K-Means Algorithms. IET Computer Vision. 12. 10.1049/iet-cvi.2018.5289.

[59] Hasan MK, Dahal L, Samarakoon PN, Tushar FI, Martí R. DSNet: Automatic dermoscopic skin lesion segmentation. Comput Biol Med. 2020 May;120:103738. doi: 10.1016/j.compbiomed.2020.103738. Epub 2020 Apr 2. PMID: 32421644.

[60] Garg, S., Jindal, B. Skin lesion segmentation using k-mean and optimized fire fly algorithm. Multimed Tools Appl 80, 7397–7410 (2021). https://doi.org/10.1007/s11042-020-10064-8

[61] Rehman M, Ali M, Obayya M, Asghar J, Hussain L, K Nour M, Negm N, Mustafa Hilal A. Machine learning based skin lesion segmentation method with novel borders and hair removal techniques. PLoS One. 2022 Nov 10;17(11):e0275781. doi: 10.1371/journal.pone.0275781. PMID: 36355845; PMCID: PMC9648757.

[62] Mata C, Munuera J, Lalande A, Ochoa-Ruiz G, Benitez R. MedicalSeg: A Medical GUI Application for Image Segmentation Management. Algorithms. 2022; 15(6):200. https://doi.org/10.3390/a15060200

[63] Md. Kamrul Hasan, Md. Asif Ahamad, Choon Hwai Yap, Guang Yang, A survey, review, and future trends of skin lesion segmentation and classification, Computers in Biology and Medicine, Volume 155, 2023, 106624, ISSN 0010-4825, https://doi.org/10.1016/j.compbiomed.2023.106624.

[64] Castellino RA. Computer aided detection (CAD): an overview. Cancer Imaging. 2005 Aug 23;5(1):17-9. doi: 10.1102/1470-7330.2005.0018. PMID: 16154813; PMCID: PMC1665219.

[65] Sonka, M., Hlavac, V., & Boyle, R. (1993). Image pre-processing. *Image Processing, Analysis and Machine Vision*, 56–111. https://doi.org/10.1007/978-1-4899-3216-7_4

[66] Skin Analytics. *Artificial Intelligence - Skin Analytics*. https://skin-analytics.com/artificial-intelligence/

[67] Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A. C., Lo, W.-Y., Dollár, P., & Girshick, R. (2023). Segment Anything. https://segment-anything.com/

[68] Francese, R., Frasca, M., Risi, M., & Tortora, G. (2021). *A mobile augmented reality application for supporting real-time skin lesion analysis based on deep learning*. *18*, 1247–1259. https://doi.org/10.1007/s11554-021-01109-8

[69] Krohling, B., Castro, P. B. C., Pacheco, A. G. C., & Krohling, R. A. (n.d.). A Smartphone based Application for Skin Cancer Classification Using Deep Learning with Clinical Images and Lesion Information.

[70] Mirikharaji, Z., Barata, C., Abhishek, K., Bissoto, A., Avila, S., Valle, E., Celebi, M.E., & Hamarneh, G. (2022). A Survey on Deep Learning for Skin Lesion Segmentation. *ArXiv, abs/2206.00356*.

[71] Zhou, T., Ruan, S., & Canu, S. (2019). A review: Deep learning for medical image segmentation using multi-modality fusion. *Array*, *3–4*. https://doi.org/10.1016/J.ARRAY.2019.100004

[72] *ISIC Challenge*. https://challenge.isic-archive.com/data/#2016

[73] *ISIC Challenge*. https://challenge.isic-archive.com/data/#2017

[74] *ISIC Challenge*. https://challenge.isic-archive.com/data/#2018

[75] *ISIC Challenge*. https://challenge.isic-archive.com/data/#2019

[76] *ISIC Challenge*. https://challenge.isic-archive.com/data/#2020

[77] P. Tschandl, C. Rosendahl, H. Kittler, The ham10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions, Scientific data 5 (2018) 1–9.

[78] Teresa Mendonça, Pedro M. Ferreira, Jorge Marques, Andre R. S. Marcal, Jorge Rozeira. PH² - A dermoscopic image database for research and benchmarking, 35th International Conference of the IEEE Engineering in Medicine and Biology Society, July 3-7, 2013, Osaka, Japan.

[79] Argenziano, Giuseppe, Soyer, H. Peter, De Giorgio, Vincenzo, Piccolo, Domenico, Carli, Paolo, Delfino, Mario, Ferrari, Angela, Hofmann-Wellenhof, Rainer, Massi, Daniela, Mazzocchetti, Giampero, Scalvenzi, Massimiliano, and Wolf, Ingrid H. (2000). Interactive atlas of dermoscopy. Milan, Italy: Edra Medical Publishing & New Media.

[80] I. Giotis, N. Molders, S. Land, M. Biehl, M.F. Jonkman and N. Petkov: "MED-NODE: A computer-assisted melanoma diagnosis system using non-dermoscopic images", Expert Systems with Applications, 42 (2015), 6578-6585

[81] S. M. M. de Faria *et al.*, "Light Field Image Dataset of Skin Lesions," *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Berlin, Germany, 2019, pp. 3905-3908. DOI: 10.1109/EMBC.2019.8856578

[82] MATLAB. MathWorks ®. https://www.mathworks.com/products/matlab.html

[83] Python. Python™. https://www.python.org/

[84] ImageJ. https://imagej.net/ij/index.html

[85] *Getting Started with Image Segmenter - MATLAB & Simulink - MathWorks España*. (s. f.). https://es.mathworks.com/help/images/image-segmentation-using-the-image-segmenter-app.html

[86] Frederic Gibou, Ronald Fedkiw, Stanley Osher, A review of level-set methods and some recent applications, Journal of Computational Physics, Volume 353, 2018, Pages 82-109, ISSN 0021-9991, https://doi.org/10.1016/j.jcp.2017.10.006

[87] Müller D, Soto-Rey I, Kramer F. Towards a guideline for evaluation metrics in medical image segmentation. BMC Res Notes. 2022 Jun 20;15(1):210. doi: 10.1186/s13104-022-06096-y. PMID: 35725483; PMCID: PMC9208116.

[88] Aydin OU, Taha AA, Hilbert A, Khalil AA, Galinovic I, Fiebach JB, Frey D, Madai VI. On the usage of average Hausdorff distance for segmentation performance assessment: hidden error when used for ranking. Eur Radiol Exp. 2021 Jan 21;5(1):4. doi: 10.1186/s41747-020-00200-2. Erratum in: Eur Radiol Exp. 2022 Oct 31;6(1):56. PMID: 33474675; PMCID: PMC7817746.

[89] Naim Mhedhbi. September 2020. Active-Contour-Model-Python. Version 2. Retrieved from https://www.kaggle.com/code/naim99/active-contour-model-python/notebook

[90] Marimuthu, P. (2022). Image Contrast Enhancement Using CLAHE. Analytics Vidhya. https://www.analyticsvidhya.com/blog/2022/08/image-contrast-enhancement-using-clahe/

[91] Basak, Hritam and Kundu, Rohit and Sarkar, Ram. MFSNet: A Multi Focus Segmentation Network for Skin Lesion Segmentation. (2022). GitHub repository. https://github.com/Rohit-Kundu/MFSNet/blob/main/inpaint.py

[92] OpenCV: Interactive Foreground Extraction using GrabCut Algorithm. https://docs.opencv.org/3.4/d8/d83/tutorial_py_grabcut.html

[93] GrabCut for Automatic Image Segmentation [OpenCV Tutorial]. (n.d.). https://www.sicara.fr/blog-technique/grabcut-for-automatic-image-segmentation-opencv-tutorial

[94] Introduction to Image Segmentation with K-Means clustering. (2020, September 20). The AI Dream. https://www.theaidream.com/post/introduction-to-image-segmentation-with-k-means-clustering

[95] Burghate, N. (2018). Work Breakdown Structure: Simplifying Project Management. *International Journal of Commerce and Management Studies (IJCAMS)*, *3*(2). www.ijcams.com

[96] U.S. Food and Drug Administration. U.S. Food and Drug Administration. https://www.fda.gov/

[97] "EU MDR – Regulation (EU) 2017/745." [Online]. https://eumdr.com/

[98] "General Data Protection Regulation (GDPR) – Official Legal Text." [Online]. https://gdpr-info.eu/

[99] "European Commission." [Online]. https://ec.europa.eu/docsroom/documents/37581

APPENDIX

DATA PREPARATION

```matlab
path = '/Users/Maria/Desktop/SegmentationFolder/InputImages'
jpeg_files = dir(fullfile(path,'*.jpg'));
for cnt = 1 : 48;
    input{cnt} = imread(fullfile(path, jpeg_files(cnt).name));
end
%%
for i = 1 : 48;
    mat = cell2mat(input(i:i));
    image = uint8(mat); %convert image into uint8 type
    resize = imresize(image, 0.25); %downscale the image
    file_name = sprintf('BinaryISIC_0000%03d.jpg', i);
    imwrite(resize,file_name,'jpg'); %save the downscaled input images
    % in a given folder determined
end
```

LEVEL SET METHOD

The code for the implementation of this algorithm has been extracted from the Kaggle repository beforehand cited. [89]

METHOD COMBINING GRABCUT AND K-MEANS CLUSTERING ALGORITHMS

Code implementation for a given skin lesion image.

```python
# import the needed libraries for the segmentation
import numpy as np
import cv2
from matplotlib import pyplot as plt
import os

    ## PRE-PROCESSING

bgr = cv2.imread('ISIC_0010014.jpg') # read the image
rgb = cv2.cvtColor(bgr, cv2.COLOR_BGR2RGB) # convert to RGB space


# 1st step: ILLUMINATION CORRECTION

# convert to LAB color space
lab = cv2.cvtColor(rgb, cv2.COLOR_RGB2LAB)

# split channel into L, A and B
lab_channels = cv2.split(lab)

# apply CLAHE algorithm to L channel in order to equalise the histogram
clahe = cv2.createCLAHE(clipLimit = 2.0, tileGridSize = (8,8))
lab_channels[0] = clahe.apply(lab_channels[0])

# merge the three channels and convert image to RGB colour space again
lab = cv2.merge(lab_channels)
equalised = cv2.cvtColor(lab, cv2.COLOR_LAB2RGB)
```

```python
# 2nd step: HAIR REMOVAL, BlackHat algorithm

# convert the original image to grayscale
grayScale = cv2.cvtColor(equalised, cv2.COLOR_RGB2GRAY)

# kernel for the morphological filtering
kernel = cv2.getStructuringElement(1, (17,17))

# perform the blackHat filtering on the image to find hair countours
blackhat = cv2.morphologyEx(grayScale, cv2.MORPH_BLACKHAT, kernel)

# intensify the hair countours
ret, thresh2 = cv2.threshold(blackhat, 10, 255, cv2.THRESH_BINARY)

# inpaint the original image depending on the mask
dst = cv2.inpaint(equalised, thresh2, 1, cv2.INPAINT_TELEA)

    ## SEGMENTATION

# GRABCUT ALGORITHM

def grabcut(image): #define the function performing GrabCut algorithm
    mask = np.zeros(image.shape[:2], np.uint8)
    bgdModel = np.zeros((1, 65), np.float64)
    fgdModel = np.zeros((1, 65), np.float64)
    # initial rectangle for foreground
    rect = (25, 25, image.shape[1]-25, image.shape[0]-25)

    cv2.grabCut(image, mask, rect, bgdModel, fgdModel, 5,
cv2.GC_INIT_WITH_RECT)

    # create a mask with foreground and probable foreground pixels as 1
    mask2 = np.where((mask == 2) | (mask == 0), 0, 1).astype('uint8')

    # apply the mask to the input image
    result = image * mask2[:, :, np.newaxis]

    return result

output_image = grabcut(dst) # apply GrabCut algorithm


# K-MEANS CLUSTERING ALGORITHM

# define the function for k-means clustering algorithm
def kmeans_image_segmentation(image, num_clusters):
    # reshape the image into a 2D array
    pixels = image.reshape(-1, 3).astype(np.float32)

    # k-means clustering
    criteria = (cv2.TERM_CRITERIA_EPS + cv2.TERM_CRITERIA_MAX_ITER, 10,
1.0)

    _, labels, centers = cv2.kmeans(pixels, num_clusters, None, criteria,
10, cv2.KMEANS_RANDOM_CENTERS)

    # convert to uint8 type
```

```python
    centers = np.uint8(centers)

    # assign each pixel to its corresponding cluster
    segmented_image = centers[labels.flatten()]

    # reshape the segmented image back to its original shape
    segmented_image = segmented_image.reshape(image.shape)

    return segmented_image

num_clusters = 4 # define the number of clusters

# apply the function to the image
segmented_image = kmeans_image_segmentation(output_image, num_clusters)




# FLOOD FILL ALGORITHM

# create the function defining the flood fill algorithm
def flood_fill_segmentation(image, seed_point, new_color):
    # create a mask of zeros with the same shape as image
    mask = np.zeros((image.shape[0] + 2, image.shape[1] + 2), dtype =
np.uint8)

    cv2.floodFill(image, mask, seed_point, new_color) # perform flood fill
on the image

    return image

seed_point = (100, 75)  # starting point for flood fill
new_color = (255, 255, 255)  # segmentation colour chosen is white

segmented_image2 = flood_fill_segmentation(segmented_image, seed_point,
new_color)




# EXTRACT THE MASK

def extract_intensity_mask(image, intensity):
    mask = np.array(image) == intensity
    return mask.astype(np.uint8) * 255

# the mask extracted is white
extracted = extract_intensity_mask(segmented_image2, 255)




# RESIZE THE IMAGE TO ITS ORIGINAL SHAPE
height = len(rgb)
width = len(rgb[0])
resized = cv2.resize(extracted, (width, height),
interpolation=cv2.INTER_LINEAR)




# SAVE THE IMAGES

# save the image in the folder indicated, from which the predicted
```

```
# binary masks are going to be selected for the further evaluation step

os.chdir("/Users/Maria/Desktop/SegmentationFolder/GrabCut+KMeans/Predicted
Seg")

cv2.imwrite('BinaryPredictedImage.jpg', resized[:,:,0])
```

MEDICALSEG

The medical application code is of private access. [62]

PERFORMANCE INDICES EVALUATION

1. FILL HOLES OF PREDICTED BINARY MASKS FROM THE METHOD COMBINING GRABCUT AND K-MEANS ALGORITHM

```
path = ('/Users/Maria/Desktop/SegmentationFolder/GrabCut+KMeans/PredictedSeg')
jpeg_files = dir(fullfile(path,'*.jpg'));
for cnt = 1 : 48;
    I{cnt} = imread(fullfile(path, jpeg_files(cnt).name));
end

for i = 1 : 48;
    mat = uint8(cell2mat(I(i:i))); %convert image into uint8 type
    image = imbinarize(mat); %binarise the image
    fill = imfill(image); %fill the holes present in the image
    file_name = sprintf('BinaryISIC_0000%03d.jpg', i);
    imwrite(fill,file_name,'jpg'); %save the filled predicted binary images in
the desired folder
end
```

2. INVERSION AND RESIZING OF PREDICTED BINARY MASKS FROM INTENSITY THRESH ALGORITHM

```
path='/Users/Maria/Desktop/SegmentationFolder/MedicalSeg/PredictedSeg';
jpeg_files = dir(fullfile(path,'*.jpg'));
for cnt = 1 : 48;
    I{cnt} = imread(fullfile(path, jpeg_files(cnt).name));
end

% Inversion and resizing of the predicted binary images since ground truth has
black background and white lesion region
for i = 1 : 48;
    mat = cell2mat(I(i:i));
    image = uint8(mat); %convert image into uint8 type
    resize = imresize(image, 4);
    compli = imcomplement(resize);
    file_name = sprintf('BinaryISIC_0000%03d.jpg', i);
    imwrite(compli, file_name,'jpg'); % save the inverted predicted binary
images in the folder called InvertedPredictedSeg
end
```

3. IMAGES IMPORTATION
3.1 Predicted segmented images importation

```matlab
path='/Users/Maria/Desktop/SegmentationFolder/LevelSetMethod/PredictedSeg'
%path for predicted binary images from Level set method
path='/Users/Maria/Desktop/SegmentationFolder/GrabCut+KMeans/PredictedSeg'
%path for predicted binary images from GrabCut and k-means algorithms combination
path='/Users/Maria/Desktop/SegmentationFolder/MedicalSeg/InvertedPredictedSeg;
%path for predicted binary images from intensity thresh automatic segmentation
algorithm


% read the images and import them into the Workspace
jpeg_files = dir(fullfile(path,'*.jpg'));
for cnt = 1 : 48;
    segIm{cnt} = imread(fullfile(path, jpeg_files(cnt).name));
end
```

3.2 Ground truth images importation

```matlab
path='/Users/Maria/Desktop/SegmentationFolder/GroundTruth'; %same path for each
case


% read the images and import them into the Workspace
png_files = dir(fullfile(path,'*.png'));
for cnt = 1 : 48;
    truth{cnt} = imread(fullfile(path, png_files(cnt).name));
end
```

4. PERFORMANCE METRICS CALCULATION

```matlab
function [Precision, Sensitivity, Accuracy, Dice, Jaccard] =
StatisticalTools(GroundTruth, PredictedMask)
% define the function to calculate each performance index
    TP = sum(PredictedMask(:) & GroundTruth(:)); %define true positives
    TN = sum(~PredictedMask(:) & ~GroundTruth(:)); %define true negatives
    FP = sum(PredictedMask(:) & ~GroundTruth(:)); %define false positives
    FN = sum(~PredictedMask(:) & GroundTruth(:)); %define false negatives

    % compute the given formula for each metric
    Precision = TP / (TP + FP);
    Sensitivity = TP / (TP + FN);
    Accuracy = (TP + TN) / (TP + TN + FP + FN);
    Dice = (2 * TP) / (2 * TP + FP + FN);
    Jaccard = (TP) / (TP + FP + FN);
end

for i = 1 : 48
    GroundTruth = uint8(cell2mat(truth(i:i))); % read ground truth image
    PredictedMask = uint8(cell2mat(segIm(i:i))); % read segmented image
    [Precision{i}, Sensitivity{i}, Accuracy{i}, Dice{i}, Jaccard{i}] =
StatisticalTools(GroundTruth, PredictedMask);
end
```

5. MEAN AND STANDARD DEVIATION CALCULATION

```matlab
% convert cell into proper data type (numeric) to assess mean value
SegPrecision = cell2mat(Precision);
SegSensitivity = cell2mat(Sensitivity);
SegAccuracy = cell2mat(Accuracy);
SegDice = cell2mat(Dice);
SegJaccard = cell2mat(Jaccard);

% mean value calculation
meanPrecision = mean(SegPrecision);
meanSensitivity = mean(SegSensitivity);
meanAccuracy = mean(SegAccuracy);
meanDice = mean(SegDice);
meanJaccard = mean(SegJaccard);

% standard deviation calculation
stdPrecision = std(SegPrecision);
stdSensitivity = std(SegSensitivity);
stdAccuracy = std(SegAccuracy);
stdDice = std(SegDice);
stdJaccard = std(SegJaccard);
```