Towards The Early Detection Of Melanoma By Automating The Measurement Of Asymmetry, Border Irregularity, Color Variegation, And Diameter In Dermoscopy Images

Abder-Rahman Ali

A thesis submitted in fulfillment of the requirements for the degree of Doctor of Philosophy

Department of Computer Science and Mathematics
School of Natural Sciences
University of Stirling

September, 2020

Abstract

The incidence of melanoma, the most aggressive form of skin cancer, has increased more than many other cancers in recent years. The aim of this thesis is to develop objective measures and automated methods to evaluate the ABCD (Asymmetry, Border irregularity, Color variegation, and Diameter) rule features in dermoscopy images, a popular method that provides a simple means for appraisal of pigmented lesions that might require further investigation by a specialist. However, research gaps in evaluating those features have been encountered in literature. To extract skin lesions, two segmentation approaches that are robust to inherent dermoscopic image problems have been proposed, and showed to outperform other approaches used in literature. Measures for finding asymmetry and border irregularity have been developed. The asymmetry measure describes invariant features, provides a compactness representation of the image, and captures discriminative properties of skin lesions. The border irregularity measure, which is preceded by a border detection step carried out by a novel edge detection algorithm that represents the image in terms of fuzzy concepts, is rotation invariant, characterizes the complexity of the shape associated with the border, and robust to noise. To automate the measures, classification methods that are based on ensemble learning and which take the ambiguity of data into consideration have been proposed. Color variegation was evaluated by determining the suspicious colors of melanoma from a generated color palette for the image, and the diameter of the skin lesion was measured using a shape descriptor that was eventually represented in millimeters. The work developed in the thesis reflects the automatic dermoscopic image analysis standard pipeline, and a computer-aided diagnosis system (CAD) for the automatic detection and objective evaluation of the ABCD rule features. It can be used as an objective bedside tool serving as a diagnostic adjunct in the clinical assessment of skin lesions.

Publications

Journal Papers

1. **Abder-Rahman Ali**, Jingpeng Li, Guang Yang, Sally Jane O'Shea, "*A Machine Learning Approach to Automatic Detection of Irregularity in Skin Lesion Border Using Dermoscopic Images*," PeerJ Computer Science, 2020 (IF: 3.09).

2. **Abder-Rahman Ali**, Jingpeng Li, Guang Yang, "*Automating the ABCD Rule for Melanoma Detection: A Survey*," IEEE Access, 2020 (IF: 3.745).

3. **Abder-Rahman Ali**, Jingpeng Li, Summrina Kanwal, Guang Yang, Amir Hussain, Sally Jane O'Shea, "*A Novel Fuzzy Multilayer Perceptron (F-MLP) for the Detection of Irregularity in Skin Lesion Border Using Dermoscopic Images*," Frontiers in Medicine, 2020 (IF: 3.9).

4. **Abder-Rahman Ali**, Jingpeng Li, Sally Jane O'Shea,, "*Towards the Automatic Detection of Skin Lesion Shape Asymmetry, Color Variegation and Diameter in Dermoscopic Images*," PLoS ONE, 2020 (IF: 2.740).

Conference Papers

1. **Abder-Rahman Ali**, Jingpeng Li, Sally Jane O'Shea, Guang Yang, Thomas Trappenberg, Xujiong Ye, "*A Deep Learning Based Approach to Skin Lesion Border Extraction With a Novel Edge Detector in Dermoscopy Images*," IEEE International Joint Conference on Neural Networks (IJCNN), Budapest, Hungary, July 14-19, 2019.

2. **Abder-Rahman Ali**, Jingpeng Li, Thomas Trappenberg, "*Supervised Versus Unsupervised Deep Learning Based Methods for Skin Lesion Segmentation in Dermoscopy Images*," 32nd Canadian Conference on Artificial Intelligence (CanadianAI 2019), Kingston, Ontario, Canada, May 28-31, 2019.

Posters

1. **Abder-Rahman Ali**, Sally Jane O'Shea, Jingpeng Li, "*Are Neural Networks Effective in Detecting Melanoma Using Genomic Data?*," 16th Annual Joint GenoMel, BioGenoMel, MELGEN Scientific Meeting, Athens, Greece, April 8-10, 2019.

2. **Abder-Rahman Ali**, Sally Jane O'Shea, Jingpeng Li, Lisa Roche, "*A Deep Learning Approach to Deblurring Skin Cancer Images*," British Association of Dermatologists, 99th Annual Meeting, Liverpool, UK, July 2-4, 2019.

Acknowledgments

I dedicate this thesis to my mother's soul, Khawla, who always prayed for me and was a source of tenderness, love, kindness, and purity throughout all her life. I would also like to thank my father, Dr. Hasan, for his inspiration and everlasting by all means support. Special thanks also go to my wife, Ola, for her motivation and patience throughout my studies. To my son Hasan and daughter Mariyya who are a source of joy and activity in my life. Thanks also go to my dear brothers and sisters for their warm community. It was a real pleasure working under the supervision of Professor Jingpeng Li, and would like to sincerely thank him for his guidance, feedback, recommendations, and support that helped me bring out the thesis as it is today.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

Skin cancer is a disease where cancerous cells form in the tissues of the skin, the body's largest organ. The skin is composed of several layers, but the two main layers are the epidermis (upper/outer layer) and the dermis (lower/inner layer); it is in the epidermis where skin cancer starts [1]. The incidence of skin cancer is rapidly increasing throughout the world and is becoming one of the deadliest forms of cancers, especially in countries with a large Caucasian population [2, 3]. Skin cancer is one of the most common forms of cancer in the UK, with recent figures suggesting an annual incidence of at least 70,000 cases, accounting for over 20% of all cancer cases. This is however considered an underestimate as registration of non-melanoma skin cancer is incomplete; the actual number of skin cancer cases in the UK each year is estimated at 100,000 cases. More than 2,500 deaths occur each year from skin cancer in the UK, of which around three-quarters are attributed to malignant melanoma [223].

Malignant melanoma is the third most frequent type of skin cancer [9]. The vast majority of melanomas are caused by the sun (one UK study demonstrated that about 86% of melanomas can be attributed to the exposure to ultraviolet (UV) radiation from the sun (Fig.1.1) [4]), and its incidence has increased more than many other cancers in recent years. In 1935 in the U.S., melanoma incidence was 1 in 1500. It is projected to be 1 in 90 for those born in 1990 and as high as 1 in 75 for those born in the year 2000 - an increase of 4-6% each year since 1973 [224]. According to Cancer Research UK, 15,419 new cases of melanoma were reported in 2014 in the UK, with a +119% change in melanoma incidence rates since the early 1990s [225]. In Scotland, the incidence of malignant melanoma is projected to change by 64% between the periods 2003-2007 (4,634 cases) to 2018-2022 (7,617) [226]. According to World Cancer Research Fund International, Australia had the highest rate of melanoma in 2018 accounting for 33.6 cases per 100,000 subjects followed by New Zealand with 33.3 cases per 100,000 subjects. UK and U.S. cases were 15 and 12.7 per 100,000 subjects, respectively [227].

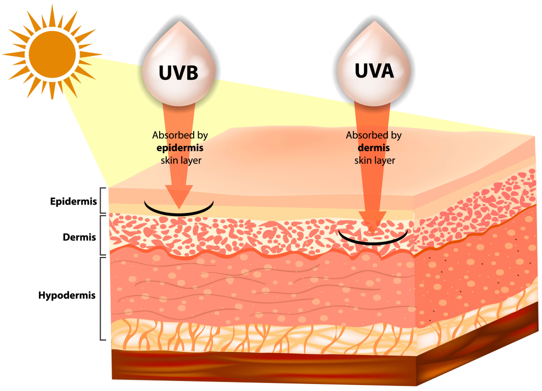


Figure 1.1: The effect of the sun on the skin [229]

The total cost of skin cancer in England in 2002 was estimated to be around £240 million [223]. This is expected to double in 2020, with projected total costs of £455.1 million [228]. In the U.S., the annual cost of treating skin cancer is $8.1 billion (non-melanoma: $4.8 billion, melanoma: $3.3. billion) [230].

A massive increase in demand is being faced by dermatologists in the UK [5]. 54% of the population are affected by skin disease each year, with 23-33% of the population having a skin disease that could benefit from dermatology care [6]. The increasing incidence of skin cancer (up 400% over the past 36 years) on the other hand has changed the practice of dermatology completely. It transformed the workload of dermatology departments, accounting for approximately 50% of their workload [7]. The RCP recommends 989 FTE (full-time equivalent) consultant dermatologists for a population of 61.8 million [5]. In 2012, the British Association of Dermatologists found there were 813 dermatology specialists in the UK (consultants, trainees, associates and associate trainees, not all full time) and a total of 729 consultant posts, 75 of which were vacant and 98 were occupied by locums [8]. It has been noted that in some parts of the UK it is almost impossible to see a dermatologist in the NHS. The North East and South East cost have only 0.69 and 0.64 full-time consultants per 100,000 population, respectively [231]. In England a patient would wait four or five months to see a dermatologist, and in Wales waiting for two years is not uncommon [232]. The U.S. also suffers from shortage in dermatologists. There are currently only 10,845 in the country, at the time that the American Academy of Dermatology (AAD) states that the U.S. needs at least 22,000 dermatologists in order to provide necessary healthcare treatments in an appropriate amount of time to those having skin issues, as Americans have to wait months for their scheduled appointments with a dermatologist [233].

1.1.1 Medical diagnosis

Although melanoma constitutes only 1% of the diagnosed skin cancer cases, 75% of deaths are caused by melanoma [10]. The increasing incidence of melanoma renders the attempts of the early detection of melanoma a continuing public health priority. Despite its aggressive infiltration of other body parts, melanoma is highly curable if diagnosed early and treated timely [11, 12]. Early detection is crucial since it contributes to a better survival; the 5-year survival rate for early stage melanoma is 94%, compared to a 5-year survival rate of only 17% for melanomas that have spread to other parts of the body [234].

There is an opportunity to increase the predictive value of melanoma diagnosis made by visual examination during pigmented lesion screening. This could help reduce the number of false-positives, the associated procedures, and corresponding scarring. Dermoscopy, or its synonyms dermatoscopy and epiluminescence light microscopy (ELM), is an in-vivo, non-invasive technique that has opened a new dimension of the clinical morphological features of pigmented skin legions (PSL) using different incident light magnification systems with an oil immersion technique [13, 14]. It provides higher accuracy in detecting suspicious cases that would not be possible via inspecting with the naked eye. Fig.1.2 shows a dermatoscope from different angles of view.

Figure 1.2: Dermatoscope [235]

Depending on the observer's experience, dermoscopy improves the diagnostic accuracy for melanoma up to 50% as compared to purely visual inspection. However, dermascopic diagnosis is subjective especially for inexperienced dermatologists. The accuracy of experts is in the range of 75% - 84% [15, 16]. Diagnostic accuracy can also be affected due to fatigue, especially with the large number of images the dermatologist has to interpret in a limited amount of time. A report from the National Cancer Intelligence Network in 2016 found that GPs failed to refer almost one in three patients with malignant melanoma for urgent tests [236].

Pattern Analysis set forth by Pehamberger et al. [17] in 1987, and updated by the Consensus Net Meeting of 2000 [18], is a method that defines the significant dermatoscopic patterns of pigmented skin lesions, allowing the distinction between malignant and benign growth features [19]. Revised Pattern Analysis distinguishes between local features consisting of individual or grouped characteristics that appear in the lesion, and global features that provide a quick preliminary categorization of a given pigmented skin lesion prior to more detailed assessment [16, 19]. In the Menzies method [20] developed in 1996, the authors identified 11 important criteria for diagnosing melanoma, such that for a lesion to be considered a melanoma it shouldn't have the two negative features, and have one or more of the nine positive features [21]. Stolz et al. developed a diagnostic scheme for dermoscopic image analysis by assessing the asymmetry (A), border (B), color (C), and diameter (D) of different structures [15]. The ABCD rule became the standard in dermoscopy for staging the PSL into benign, suspicious, or malignant moles [15, 16]. To handle the diagnostic accuracy of non-expert dermatologists, the 7-point checklist has been proposed as a diagnostic method that requires the identification of 7 dermoscopic criteria in order to aid clinicians in their use of dermoscopy, and has been shown to be reproducible with non-expert dermatologists who showed the ability to classify a high percentage of melanomas [22]. The CASH (color, architecture, symmetry, homogeneity) algorithm [23] is a simplified form of Pattern Analysis that even a novice dermatoscopist can use. Many of the features used in the ABCD rule are assessed by this algorithm, and it adds an element not found in any other algorithm, namely the architecture [24].

1.1.2 Computer-aided diagnosis

There is a niche to develop an objective, bedside tool that could be used as an adjunct in the clinical assessment of skin lesions. Tracking tumor changes manually is also labour-intensive, especially for patients with multiple moles on their skin. Computer-aided diagnosis (CAD) methods can assist dermatologists in different steps of analysis, such as detection of the lesion boundary, quantification of diagnostic features, classification into different lesion types, and visualization [25]. In fact, CAD systems are able to obtain a higher level of sensitivity for melanoma detection compared with inexperienced dermatologists [26]. Essentially, the standard pipeline in automatic dermoscopic image analysis consists of three main stages (Fig.1.3): (i) image segmentation (ii) feature extraction (iii) lesion classification.

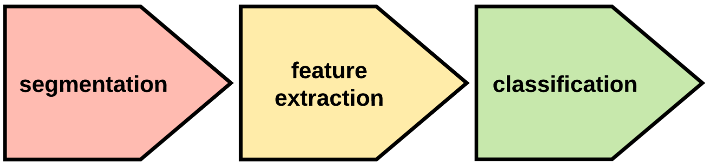


Figure 1.3: Automatic dermoscopic image analysis standard pipeline

1.2 Problem statement

The ABCD (Asymmetry, Border irregularity, Color variegation, and Diameter) rule is a popular method that provides a simple means for appraisal of pigmented skin lesions that might need further investigation by a specialist, and is commonly used by physicians, novice dermatologists, and non-physicians as a way to learn about the features of melanoma in its early curable stage, eventually leading to the early detection of melanoma. However, research gaps in evaluating the ABCD rule features have been encountered in literature.

Computerized methods proposed for measuring asymmetry demonstrated a dependency on the segmentation performance and how the axes locations are determined, asymmetry thresholds were heuristic and non-generalizable to other test images, and the extrinsic shape was mainly evaluated ignoring the asymmetry of the inhomogeneous pigmentation inside the lesion. Euclidean based measures, which don't characterize the complexity of the skin lesion shape, were the main measures used in characterizing border irregularity, with other measures in literature being empirical and prone to error, rely on the correlation of image pixels - which can be distorted when carrying out augmentation operations -, and sensitive to noise, rotation, and scale. No proper learning algorithms have been proposed to automate the aforementioned measures nor there were relevant datasets available to aid in building such algorithms. Although different approaches have been proposed to detect color variegation in skin lesions, they however don't accurately characterize the suspicious colors of melanoma and rather follow a trial-and-error approach in determining such colors. The use of an inappropriate color space which is not representative of the human perception was common amongst those approaches. Approximative measures that don't reflect the actual diameter have also been used in literature.

Due to its importance in determining the accuracy of subsequent stages, the first step in skin lesion analysis is usually image segmentation. However, classical image segmentation techniques can easily fail in skin lesion segmentation due to inherent dermoscopic image problems, such as weak edges and the presence of artifacts like light reflectance, apart from the great variety of lesion shapes, sizes, textures, and colors.

1.3 Research objectives

The aim of this thesis is to develop objective measures and automated methods for evaluating the ABCD rule features in dermoscopy images. To accomplish this objective, my PhD work was oriented into the following main directions:

- To develop image segmentation approaches that are robust to inherent dermoscopic image problems, and to investigate the efficacy of deep learning based segmentation methods in skin lesion segmentation in dermoscopic images.

- To develop objective measures that fill-in the research gaps encountered in literature for evaluating asymmetry, border irregularity, color variegation, and diameter.
- To build classification networks for automating feature detection.
- To build a database pertaining the asymmetry and border irregularity features and to raise the awareness of the scarcity and the need of such datasets.

1.4 Research contributions

In light of the above objectives, different contributions resulted from my PhD work. Two segmentation approaches have been proposed, namely gradual focusing and Otsu-II. The first approach was inspired by the human visual perception in that images are processed on a global scale to locate the main regions (coarse information) of the image and on a local scale to explicit the boundaries of the detected regions. In the second approach, an improvement of the traditional Otsu method has been proposed, coupled with pre-processing and post-processing stages to constitute a comprehensive robust image segmentation approach. The two approaches showed to outperform U-Net, a commonly used supervised deep learning based method for medical image segmentation.

Objective measures have been developed for the ABCD rules features. The asymmetry measure describes invariant features, provides a compactness representation of the image, and captures discriminative properties of skin lesions. A decision tree was used to learn the asymmetry measure to detect the asymmetry of new skin lesions, and has led to high accuracy. As a prerequisite to measuring border irregularity, the skin lesion border has to be extracted. A novel edge detection algorithm, FuzzEdge, which represents the image in terms of fuzzy sets was developed, depicting high similarity to the borders annotated by the dermatologist. The proposed border irregularity measure is rotation invariant, characterizes the complexity of the shape associated with the border, and robust to noise. To automate border irregularity detection, two classification approaches have been proposed. The fuzzy multilayer perceptron (F-MLP) which is a neural network that takes the ambiguity of data into account, and a CNN-Gaussian Naive Bayes ensemble which outperforms using the two classifiers individually. The proposed classification approaches demonstrated outstanding performance. A color palette composed of the image's dominant colors has been created from which the melanoma suspicious colors can be identified, eventually determining the presence of color variegation in the skin lesion. Finally, Feret's diameter was used as a shape descriptor for measuring the lesion's diameter, and spatial calibration was used to represent the diameter in millimeters.

Combining the contributed work could produce a CAD system that is able to automatically and objectively evaluate the ABCD rule, and which could also serve as an objective second opinion to the ABCD rule in general and its individual features in particular.

1.5 Thesis structure

The remainder of this thesis is structured as follows:

Chapter 2 presents a literature review of the different computerized methods developed for finding asymmetry, border irregularity, color variegation, and diameter in skin lesions, in addition to providing a critical analysis on the reviewed studies.

Chapter 3 introduces the background required to understanding the remainder of the thesis.

Chapter 4 explains the image segmentation approaches proposed in the thesis, in addition to providing a comparative analysis between supervised and unsupervised deep learning based methods for skin lesion segmentation in dermoscopy images.

Chapter 5 describes the proposed approaches for determining asymmetry, color variegation, and diameter in skin lesions.

Chapter 6 highlights the proposed approaches involved in determining skin lesion border irregularity in skin lesions.

Chapter 7 concludes the overall work presented in the thesis and gives recommendations for future work.

Chapter 2

Literature Review

2.1 Overview

Numerous efforts have been made to automate the ABCD rule by proposing objective approaches to finding the different features of the rule. This chapter starts by introducing the ABCD rule, followed by sections 2.3, 2.4, 2.5, and 2.6 that highlight the attempts geared towards finding the asymmetry, border irregularity, color variegation, and diameter of skin lesions in literature, respectively. Section 2.7 reviews the studies where a combination of the aforementioned features has been used as opposed to analyzing only an individual feature. A critical analysis is provided in section 2.8, and the chapter is summarized in section 2.9.

2.2 ABCD rule

The ABCD acronym refers to four parameters: (i) Asymmetry (ii) Border irregularity (iii) Color variegation (iv) Diameter greater than 6mm. The ABCD rule [27] emerged in 1985 by a group of researchers at the New York University as a simple framework that physicians, novice dermatologists, and non-physicians can use to learn about the features of melanoma in its early curable stage, enhancing the early detection of melanoma. It is more geared towards the public than the 7-point checklist which was designed for non-dermatological *medical* personnel. The approach has then been verified by the 1992 National Institutes of Health Consensus Conference

Report on the detection and treatment of melanoma in addition to other studies published at the time [28],[29],[30],[31], and is being promoted by the American Cancer Society as a method to help in seeking early medical evaluation of any suspicious pigmented lesions.

The ABCD rule provides a simple means for appraisal of pigmented cutaneous lesions that may need to be further examined by a specialist, which might result in further work of dermoscopy or biopsy, or both. The rule is not designed to provide a comprehensive list of all melanoma characteristics. It should be noted that the combination of the ABCD features (i.e., AB, AC, ABC) is what determines the suspicious lesions and has the greatest accuracy when used in combination, especially that it is not necessary for all melanomas to acquire all the four features. Combining the reliable sensitivity and specificity in the application of the ABCD rule in clinical practice supports the ongoing usage of the rule [32]; the ABCD rule is used in public education on a wide basis and is easy to memorize.

Asymmetry refers to the fact that when drawing a line across the middle of the mole, the two halves will *not* match, that is, the shape of one half doesn't match the other half, providing a warning sign of melanoma. Asymmetry evaluation is carried out by separating the lesion into four sectors using orthogonal axes that pass through the lesion centroid and are aligned so that minimum asymmetry (maximum symmetry) is obtained [33]. Fig.2.1 shows samples of symmetric and asymmetric skin lesions (skin lesion images shown in Fig.2.1-Fig.2.3 were obtained from the ISIC 2018: Skin Lesion Analysis Towards Melanoma Detection grand challenge datasets [81, 82]). Unlike benign mole borders that tend to be smooth and even, early melanoma borders possess uneven (*irregular*) borders. Fig.2.2 shows samples of skin lesions with regular and irregular borders. *Color variegation* refers to the presence of two or more shades of pigment (two or more colors) within the skin lesion border. Melanoma lesions often contain more than two colors as opposed to benign lesions which tend to be generally uniform in color. In particular, melanoma contains one or more of these six suspicious shades of color: white, red, light brown, dark brown, blue-gray, and black. Fig.2.3 depicts a skin lesion and its dominant colors. Melanomas usually have a larger *diameter* than benign moles, which is around 6mm (i.e. the size of a pencil eraser), although they might be smaller when detected earlier.

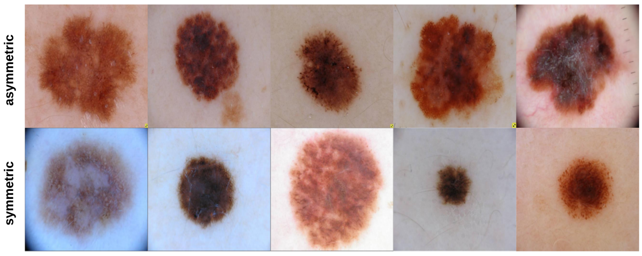


Figure 2.1: Samples of symmetric and asymmetric skin lesions

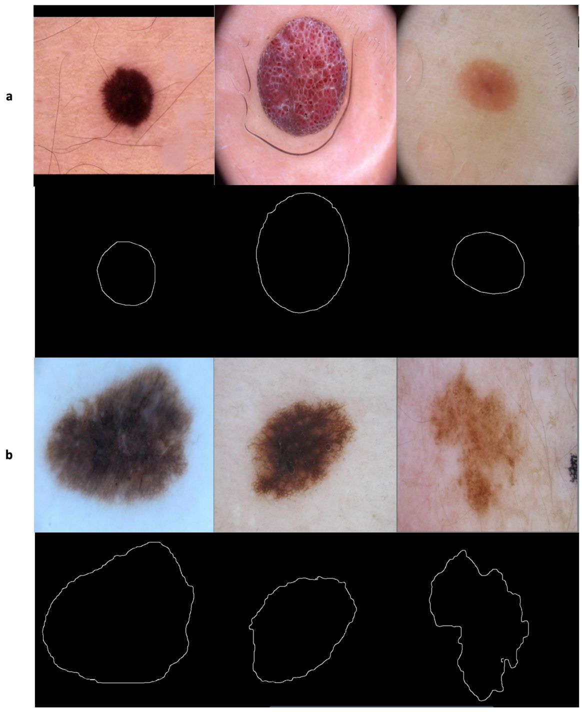


Figure 2.2: Skin lesions with (a) regular borders (b) irregular borders

Figure 2.3: A skin lesion image and its corresponding color palette showing the 7 dominant colors in the image (the first color is pure black and represents the background)

Since the interpretation of the ABCD rule features is considered subjective, different solutions have been proposed in literature to tackle such subjectivity and provide objective evaluations to the features. As explained in Chapter 1 (Fig.1.3), the standard pipeline of automatic dermoscopic image analysis consists of three main stages: image segmentation, feature extraction, and lesion classification. Fig.2.4 depicts the ABCD features and the stages involved in each feature based on the reviewed papers highlighted in the coming sections.

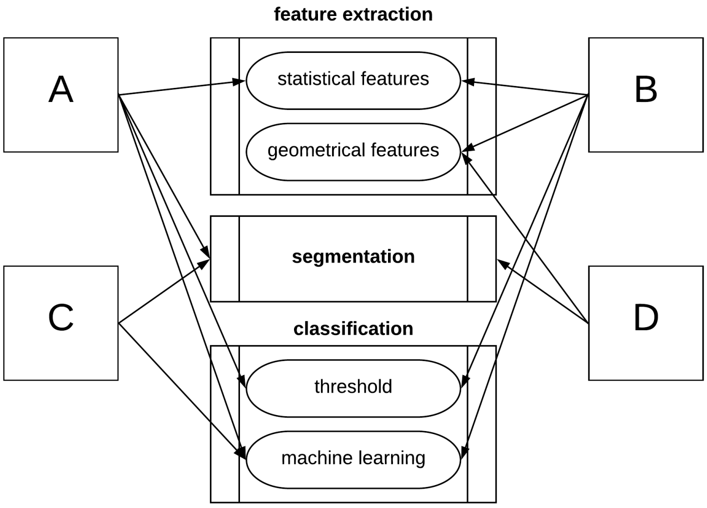


Figure 2.4: Dermoscopic image analysis stages involved in the ABCD rule features

2.3 Asymmetry

Different attempts have been made to automatically determine the asymmetry of skin lesions in literature. *Circularity* ($\frac{4\pi A}{P^2}$; *A: area, P: perimeter*) has been the main feature used in determining asymmetry, but the major drawback of this measure is that it relies mainly on the segmentation performance, especially that when skin lesion borders are highly irregular this will cause the segmented lesion to appear with a thick border and a small internal area, thus causing the circularity value to be small and misrepresenting the correct measurement [34]. To overcome such drawback, Ng and Cheung [34] used *Symmetry Distance (SD)*. However, this approach becomes computationally expensive when there are many points on the skin lesion border. Even if selecting fewer points, this selection will impact the SD value and is dependent on the real shape of the lesion to be measured. The authors propose an algorithm for combating the issue. To take the irregular borders into account and how they might affect the segmentation output, a *Fuzzy Symmetric Dstance (FSD)* is proposed. Vincent et. al. also used SD in their work [35].

Stoecker et. al. [36] proposed a method where the x and y coordinates of the image were made to coincide with the centroid of the image through shifting and translating the shape; the x and y coordinates are then aligned with the centroidal principle axes by rotating the image. The image is eventually reflected across the principal axis and its orthogonal counterpart. Two area differences are produced by subtracting the image on one side of the axis from the reflected image, where the least of the absolute values of the area differences is divided by the skin lesion area, resulting in the *asymmetry index*. A classification step is then carried out by finding the best asymmetry index threshold, such that skin lesions having an index value greater than the threshold are classified as asymmetric.

Seidenari et. al. [37] measured symmetry by dividing the skin lesion into 256 sectors using axes that were passed through the baricentre of the lesion. The difference in symmetric areas with respect to the centre was then found, evaluating to symmetry values that ranged from 0 (symmetric) to 10 (asymmetric). Andreassi et. al. [38] employed a similar approach were the area differences were taken between 360 lesion segments.

d′Amico et. al. [39] found asymmetry by splitting the skin lesion into two halves via a straight line that passes through the center of mass. Comparison of the two halves was then made by computing the distance between their *size functions* [**?**]. The splitting was repeated for 45 equally spaced radial lines, resulting in the distance as a function of angle. A set of characteristic numbers are extracted from the generated curve which are eventually fed to a Support Vector Machine (SVM) for classification. The advantage of this approach lies in detecting *qualitative* asymmetry as opposed to the traditional methods proposed which focus on geometrical asymmetry.

Ma et. al. [41] firstly define the major and minor axes that partition the skin lesion into four parts; the asymmetry measure of the contour is found by calculating the normalized asymmetry degrees Δk of a pair of contour segments with respect to the central, major, and minor symmetry. Features for each contour segment that would determine the degrees of asymmetry are then extracted, namely the area, averaged distance of local fractals, relative radial distance, and normalized zero-cross rate.

The integration of Fourier descriptors into a shape asymmetry quantifier was made in [42], where it has been shown that such descriptors could aid in identifying the skin lesion's principal axes of symmetry accurately, as when using the top two unique axes determined by the proposed approach a 92% match with the dermatologist has been achieved.

After segmenting the skin lesion, Tenenhaus et al. [43] considered both shape and texture asymmetry. In shape asymmetry, two symmetry axes calculated with the Hough transform algorithm are formed and asymmetry is estimated for each axis from the percentage of overlapping pixels after rotating the tumor surface around the axis. Texture asymmetry is estimated by finding the quadratic error averages between the intensity of overlapping pixels after rotating the tumor around the symmetry axes. Asymmetry is eventually described by a 4-dimensional vector, and this feature vector is fed to a KL-PLS (Kernel Logistic Partial Least Square regression) based classifier [44] as input which is then used to detect asymmetry. The approach is tested on 227 tumor images resulting in 73% accuracy.

She et al. [45] measured asymmetry by folding the lesion outline (i.e. segmented image) about the major axis of the best-fit ellipse, finding the non-overlapping region, and calculating the percentage of the non-overlapping region over the lesion area using the equation $A = \frac{\Delta T}{T} \times 100\%$, where ΔT is the number of pixels in the non-overlapping region, and T is the lesion area; A is called *bi-fold method* [46].

After the skin image is converted to grayscale and segmented, Zaqout [46] partitions the skin lesion (foreground of segmented image) across its centroid into two equal halves both vertically (left vs. right) and horizontally (top vs. bottom). To measure asymmetry, the entropy between each corresponding half is found and the entropy asymmetry is calculated based on a specified entropy threshold value. The bi-fold method is also applied and the overlapping asymmetry is calculated based on an overlapping threshold. The average value of those asymmetry results represent the overall asymmetry score of the segmented lesion.

Kasmi and Mokrani [47] first determine the axes of symmetry (horizontal and vertical) then measure color asymmetry, brightness asymmetry, and shape asymmetry. In *color asymmetry*, the image is divided into blocks of 20×20 pixels, such that blocks with at least 75% of their area inside the lesion are selected. The image is converted to the $L^*a^*b^*$ color space, the color average of each selected block is computed, and the color distance (i.e. Euclidean distance) between two symmetrical blocks along the principal axis is measured. The authors set a JND (just noticeable difference) as a threshold for differentiating between colors, such that along any axis for any pair of blocks if JND is smaller than the specified threshold the block pair is said to be color symmetric. If along any axis the number of blocks that are color symmetric is greater than the number of blocks that are color asymmetric, the image is considered to be color symmetric along that axis. *Brightness asymmetry* is measured by finding the difference between the average luminance of the two opposite halves of the skin lesion along a symmetry axis. If the difference is less than a specified threshold (3% of the total average luminance), the lesion is considered to be brightness symmetric along that axis. *Shape asymmetry* is measured by finding the difference between the lesion areas of the two opposite halves and comparing it with a specified threshold (2% of the lesion area). A lesion which is simultaneously color, brightness, and shape symmetric along an axis is considered symmetric along that axis. If the lesion is symmetric along both axes, along only one axis, not symmetric along any axis, the asymmetry score is set to 0, 1, or 2, respectively.

2.4 Border irregularity

Various studies attempting to detect the irregularity of borders in skin lesions and melanoma have been proposed in literature. In [48], a dermatologist was asked to score 60 skin tumor images as being regular or irregular (regular: 14, irregular: 46). A border was then found using a radial search algorithm [49], where different windows (i.e. sliding window) are automatically detected in the skin lesion, each of which represents the origin of a radii. Radii are searched for sufficiently high jumps in luminance that also possess sufficiently sustained luminance as those will form the candidate border points [50]. Irregularity is eventualy found using the following formula:

$$I = \frac{P^2}{4\pi A} \tag{2.1}$$

where P and A denote the *perimeter* and *area* of the closed boundary, respectively. The perimeter is measured by counting the points on the detected border, and the area is measured by counting the points on and within the border. The authors reached a conclusion that borders with an irregularity index greater than 1.8 were classified as being irregular. Using the proposed algorithm, 42 of the 46 irregular tumors were classified correctly. Of the 14 regular tumors, 8 were classified correctly. Thus, 83.3% of the tumors were classified the same as the dermatologist. This irregularity formula (i.e. I) has also been used by She et al. [45] and Messadi et al. [51] (combined it with compactness, fractal dimension, and radial variance).

Ng and Lee [52] used fractal dimensions (FDs) in measuring the irregularity of skin lesion borders. For each color image, four fractal dimension measures are found: direct FD, vertical smoothing FD, horizontal smoothing FD, and multi-fractal dimension of order two. Those FDs are also calculated on the blue band of the images. After being segmented by a multi-stage method [53], 468 melanocytic lesions (not hairy) are used to test the proposed approach. Results show that the multi-fractal method performs the best. FDs were also used in [54].

An automatic approach for analyzing the structural irregularity of cutaneous melanocytic lesion borders was proposed in [55]. The algorithm consists of two stages. In the first *preprocessing* stage, the lesion border was extracted from the skin images after removing the dark thick hair by DullRazor [56]. In the second stage, the structural shape of the lesion border was analyzed using a proposed measure, namely *sigma-ratio*, which is derived from the scale-space filtering technique with an extended scale-space image. Results showed that unlike shape descriptors such as *compactness index* and *fractal dimension* that are more sensitive to texture irregularities than structure irregularities (i.e. don't provide accurate estimation for the structure irregularity) [57], sigma-ratio is considered sensitive to the structural indentations and protrusions. The authors further improved their past work to propose a new border irregularity measure in [57, 60, 204]. The new method works first by locating all indentations and protrusions along the lesion border, and a new *irregularity index* is measured for each indentation and protrusion. Summing up all the individual indices provides an estimation on the overall border irregularity. A new feature was also introduced in the proposed method in that it is able to localize the significant indentations and protrusions.

Arbisala and Claridge [61] proposed a new measure of border irregularity based on *conditional entropy*, where it was observed that the entropy increases with the degree of irregularity.

98 skin lesions were used in the experiments, of which 16 were melanoma. The results of the proposed measure were compared with the Indentation Irregularity Index [60] and showed to have a better discriminatory power such that the area under the ROC curve was 0.76 compared to 0.73 for the Indentation Irregularity Index. In particular, the proposed measure gave 70% sensitivity and 84% specificity.

Ma et al. [62] used *wavelet decomposition* to extract the skin lesion border structure, based on which they would determine whether the lesion is a naevus or melanoma. Using the discrete wavelet transform (DWT), the 1D border is filtered into sub-bands down to level 9 where levels 6 to 9 (significant levels) have shown to contain information considered best for classifying between melanoma and benign samples. Some statistical and geometrical feature descriptors of border irregularity are extracted at each individual sub-band. A back-projection neural network is used as a classifier which receives a combination of features as input. 25 measurements are formed by applying 6 features in four significant sub-bands, and one feature in a single sub-band. Using a small training set of 9 melanomas and 9 naevi, the best classifier is obtained when the best 13 features are used.

A system was proposed by Jaworek-Korjakowska and Tadeusiewicz [63], which consists of the following steps: image enhancement, lesion segmentation, border irregularity detection, and classification. To find the border irregularity, the authors translated the border into a function with peaks that indicate the border irregularity. This is achieved by implementing a four step algorithm: (i) computing the bounding box of the segmented skin lesion; (ii) finding the boundary pixels lying on the lines that connect the center of the mass with the vertices; (iii) calculating the distance between the border and the edge of the image, which results in a function with an exact reflection of border irregularities. The signal is smoothed using a Gaussian filter in order to determine the ragged edges; (iv) finally calculating the derivative to find the local maximum points of the function, such that the local maximum is detected when the function crosses the zero point and the slope changes from + to -. The authors used a simple method to measure border irregularity, in which a simple semi-quantitative evaluation method is used to divide the lesion into eight similar parts such that the sharp abrupt cut-off in each part has a score of 1. Thus, a maximum score of 8 is obtained if the whole border is irregular, and a score 0 is obtained if the naevus is round with no ragged borders (authors in [64] divided the skin lesion into 8 segments and determined the pigment change in each segment for the measurement of border irregularity, with score ranging from 0-8. Zaqout [46] on the other hand found the *compactness index*: $I = \frac{P^2}{2\pi A}$ for each segment and the border irregularity index was calculated as the sum of those indices). As a rule of thumb, melanomas tend to have scores 4-8 [65]. This was tested on 120 skin lesion cases with border irregularity less than 3 and 180 skin lesion cases with border irregularity greater than 4, and the proposed approach achieved a 79% accuracy.

To determine border irregularity, Kasmi and Mokrani [47] divided the skin lesion into eight equal slices and approximated the sub-contour of each slice using a third-order spline function. A fitting error was then computed such that if the error is larger than (0.05× sub-contour length), the sub-contour is considered irregular. Each irregular sub-contour will be given a score of one, with the maximum score a lesion can have is eight.

2.5 Color variegation

In [66], Umbaugh et. al. proposed a color segmentation algorithm for the identification of *color variegation* in skin tumors. The algorithm is composed of six steps: (i) color averaging (ii) feature masking (iii) color space segmentation (iv) object filtering (v) object labeling, and (vi) high level processing, which is used to define color variegation based on two rules: (1) if the ratio of the tumor area while excluding ulcer, crust, and shiny areas (defined in step ii of the algorithm) to the entire tumor area is leas than 0.5, no color variegation is present (2) for any object of size greater than $2mm^2$, if it was composed of two or more colors then color variegation is present. The approach was tested on 160 images and achieved a 73% accuracy in determining the presence/absence of color variegation.

Mimicking how dermatologists determine color variegation, where they locate the areas with homogeneous color and decide the number of colors based on the number and size of those zones, Tenenhaus et. al. [43] propose two unsupervised classification methods for describing color homogeneity, namely Kohonen map and K-means clustering. In Kohonen map, a random selection of 5 pixels from each tumor in the database is obtained to form a 5×5 Kohonen map in the RGB color space (25 neurons in the map), which in turn represents the variegation of colors in the lesion. In a single color lesion most of the pixels will be projected in the same region, while pixels of a multiple color skin lesion will be projected across different regions of the map. Color variegation is determined based on a 25-dimensional vector which is obtained by finding the proportion of pixels projected onto each of the 25 neurons of the map. The authors also use K-means to classify the pixels of the skin lesion where the number of clusters are set to the number of colors detected by the dermatologists, which was set to $k = 4$ as the authors found that the dermatologists didn't spot more than 4 colors in the lesion in most cases. A 16-dimensional feature vector is formed in this step (4 clusters × RGB color of each cluster centroid). The detection of the blue coloration is finally made using the Hue component in the HSV (Hue, Saturation, Value/brightness) color space. The Hue values fall in the $[0, 360]$ range, and the blue color is coded by the proportion of pixels in the blue interval $[200, 250]$. Color features are eventually represented by a 42-dimensional vector, which is fed to a KL-PLS based classifier as input for the detection of color variegation. The approach is tested on 227 tumor images resulting in 66% accuracy.

To find color variegation, Kalwa et. al. [67] iterate through each pixel in the skin lesion HSV (Hue, Saturation, Value) image and extract the hue value of each pixel, eventually grouping pixels that have a hue value within a specified range together. However, a drawback of this approach is that the authors use trial and error to represent the HSV value equivalents of the melanoma suspicious colors.

She et al. [45] quantify color variegation using the normalized standard deviation of the red, green, and blue components of the lesion which are defined as: $C_r = \frac{\sigma_r}{M_r}$, $C_g = \frac{\sigma_g}{M_g}$, and $C_b = \frac{\sigma_b}{M_b}$, respectively. σ_r, σ_g, and σ_b are the standard deviations of the red, green, and blue components of the skin lesion area, respectively, and M_r, M_g, and M_b represent the maximum values of the red, green, and blue components of the lesion area, respectively.

Kasmi and Mokrani [47] measure the Euclidean distance between each pixel of the skin lesion and the six suspicious colors of melanoma (white, black, light brown, dark brown, red, and blue gray). A pixel is said to belong to some color if the distance is less than a threshold T which is calculated as the half of the distance between the two extreme colors (white and black). A lesion is considered to contain a suspicious color if the pixels that belong to this color represent more than 5% of the skin lesion pixels. This approach has been previously used in [68]. The same rationale as in [47], [68] is used in my work to determine the suspicious colors present in each skin lesion. However, the CIELab color space has been used in my work which is more representable of the human perception than the RGB color space. Moreover, the suspicious colors CIELab values are derived based on the color distribution of the used dataset, making it more accurate in determining the suspicious colors. It was not clear how the colors in [47],[68] have been derived (apart from the white, black, and red colors in [68] where the standard RGB values have been used); the RGB value for the white color in [47] did not represent the actual color (another color is produced rather then the color of correspondence). Authors of those two studies have used only one representative value for each suspicious color which might not be representative enough especially that we can have different levels (shades) of the same color (i.e. light brown). As opposed to those two studies, Minkowski distance is used instead of Euclidean distance.

2.6 Diameter

To find the diameter, authors in [68] measure the horizontal and vertical dimensions of the lesion as $h = max(i) - min(i)$ and $v = max(j) - min(j)$, respectively, and then calculate the diameter as $D = max(h, v)$ in pixels and eventually convert the value to millimeters based on the true size of the image. Kalwa et. al. [67] measured the diameter using the equation $D = 2\alpha\gamma$, where α represents the skin lesion's minimum enclosing rectangle in pixels and γ is the conversion factor from pixels to millimeters, which is found using the imaging system parameters (i.e. focal length, distance from the object to the lens). Garnavi et. al. [69] propose *greatest diameter* and *shortest diameter*. The greatest diameter is defined as the length of the line that connects the two *farthest* boundary points while passing across the lesion centroid. The shortest diameter on the other hand is the length of the line that connects the two *nearest* boundary points and passes across the skin lesion centroid. In [46], Zaqout used the major axis length of the segmented lesion as a measure for diameter in pixels and then converted the result to millimeters.

To find the diameter, Messadi et al. [51] first determined the coordinates (x, y) of each pixel in the skin lesion, and then calculated the distance between each pixel pair. The maximum of those distances is considered the lesion diameter. However, the authors did not mention the conversion to millimeters which made their measure lack the representation of the actual diameter in the real world.

Diameter in [45] was calculated using the formula $D = 2a$, where a is the semi-major axis of the best-fit ellipse. The result (in pixels) is eventually converted to millimeters using the prior knowledge of image pixel parameters and the spatial relationship at a particular magnification.

It should be emphasized that segmentation is a crucial step before diameter measurement can take place. Moreover, a major limitation in the methods attempting to measure the diameter is finding the correct conversion factor to millimeters, which mainly depends on the original image size taken in the real world, a feature which is not always available especially when working with online image datasets that lack such information.

2.7 Feature combination

Combining the ABCD features (e.g., AB, AC, ABC) has greater accuracy in determining suspicious lesions. Thomas et al. [70] showed the sensitivity and specificity of each individual criterion in the diagnosis of melanoma (Table.2.1), in addition to the sensitivity and specificity of the combination of these criteria in the diagnosis of melanoma (Table.2.2).

Table 2.1: Sensitivity and specificity of each ABCD criterion in diagnosing melanoma

No. of Lesions	No. of melanomas	Criterion	Sensitivity	Specificity
1140	460			
		A	57%	72%
		B	57%	71%
		C	65%	59%
		D	90%	63%

Table 2.2: Sensitivity and specificity of the combination of ABCD criteria in diagnosing melanoma

No. of Lesions	No. of melanomas	Criterion	Sensitivity	Specificity
1140	460			
		at least 1	97.30%	36%
		at least 2	89.30%	65.30%
		at least 3	65.50%	80%
		at least 4	54%	93.50%

She et al. [45] combined skin pattern features (skin line direction and intensity) with ABCD features to enhance classification performance. Before extracting the ABCD features, the authors used a snake-based algorithm [71] to detect the lesion area and form a binary image (i.e. segmentation); the lesion centre, orientation, and best-fit ellipse were determined from the binary image. The dataset used in the experiments was composed of 36 colored 24-bit images of size 230×350 (melanoma: 16, naevi: 20) which were converted to grayscale for skin pattern analysis. Classification was eventually carried out using individual features and a combination of features. In particular, the means of the *skin line direction* for the skin and lesion areas and their differences were calculated and a scatter plot of the line direction difference has been formed; the scatter plot would represent the separability used for classification. The area under the ROC curve evaluated to 0.84. Using the same approach but with the *skin line intensity* as the feature evaluated to 0.80. Using asymmetry, border irregularity, color variegation (red, green and blue components), and diameter as individual features evaluated to 0.66, 0.62, 0.54 (red), 0.76 (green), 0.78 (blue), and 0.62, respectively. The authors then combined the 8 features (skin pattern features and ABCD features) and used Principal Component Analysis (PCA) [72] to reduce the features used in classification to 2 features, resulting in an area under the ROC curve of 0.94.

Jaworek-Korjakowska et al. [73] proposed a software system for the detection of melanoma based on the ABCD rule. The system starts by converting the input colored image to a monochrome image using Otsu's method [74] and then utilizing DullRazor for hair removal. Image enhancement proceeds by blurring the edges and removing salt and pepper noise using the median filter, and tackling holes resulting from Otsu's method using different morphological operations (i.e. erosion, dilation); skin lesion borders are eventually found which represent the region of interest (i.e. lesion). The author then extracts the ABCD rule features via different standard methods. Asymmetry is evaluated using the asymmetry index, irregularity is determined using the Harris corner detector [75], color variegation is analyzed using color segmentation based on multidimensional thresholding, and if the diameter is larger than 6mm the factor $D = 5$. After the ABCD features were extracted, the author used the Total Dermoscopy Score (TDS) to indicate the malignancy (i.e. melanoma) of the lesion, which is defined as:

$$TDS = (A \times 1.3)\ (B \times 0.1)\ (C \times 0.5)\ (D \times 0.5) \tag{2.2}$$

where the score ranges of A, B, C, and D are $(0-2)$, $(0-8)$, $(1-6)$, and $(0-5)$, respectively. A skin lesion is classified as melanoma if $TDS > 5.45$, a benign lesion if $TDS < 4.75$, and as a suspicious lesion if the TDS lies between 4.75 and 5.45. The system was tested on 50 lesions (benign: 20, malignant: 30) and achieved 87% sensitivity and 80% specificity. TDS was also used in [64] and [46], with the latter applied on the PH^2 dataset[1] and resulting in an accuracy, sensitivity, and specificity of 90%, 85%, and 92.2%, respectively. Kasmi and Mokrani [47] used TDS, but the D feature referred to (Different Structures) as opposed to Diameter, where pigment network [76] and geometrical properties of the lesion (fractal dimension, asymmetry index, circularity, elasticity) were used to evaluate the D feature. The approach was applied on 200 dermoscopic images (bemign: 120, melanoma: 80) with size 712×454 obtained from the EDRA Interactive Atlas of Dermoscopy [65], and resulted in an accuracy, sensitivity, and specificity of 94%, 91.25%, and 95.83%, respectively.

After pre-processing the skin lesion image and determining the lesion area, Ramezani et al. [77] extracted a group of features that represent the ABCD rule traits. Asymmetry was represented by a group of 32 features (i.e. orientation angle, asymmetry index), border irregularity by a group of 34 features (i.e. irregularity index, compactness index), color variegation by 72 features (i.e. mean, standard deviation), and diameter by 7 features (i.e. best-fit ellipse diameter, major diameter). The authors further add 42 features of lesion texture (i.e. contrast, entropy) extracted using Grey Level Co-occurrence Matrices (GLCM) [78].

---

[1] https://www.fc.up.pt/addi/ph2%20database.html

To reduce the number of extracted features, PCA was utilized. This reduced the number of selected features from 187 to 13. An SVM was eventually used for classifying skin lesions into malignant or benign. The dataset used was composed of 282 macroscopic images collected from different online dermatology atlases such as Dermnet, Dermis and Dermquest atlases. The dataset included RGB images of 149 benign lesions and 133 malignant lesions ranging in dimension from 259×382 to 1186×1369 pixels. 70% of the data was used for training SVM and 30% for testing the classifier. That is, 197 and 85 images were used for training and testing the classifier, respectively. The approach resulted in accuracy, sensitivity, and specificity of 82.2±3.57%, 77.02±5.97%, 86.93±5.46%, respectively.

In [51], Messadi et al. pre-processed the skin lesion by removing the hair using the approach proposed in [79], then segmented the grayscale image using histogram thresholding and level sets [80]. The ABCD features were then extracted and fed to a multilayer perceptron (the approaches authors used to extract the ABD features are discussed in the previous relevant sections). For analyzing the C feature, four parameters were selected from the 14 proposed by Haralick et al. [78], namely: correlation, homogeneity, energy, and contrast). 320 color images containing both benign and melanoma samples from DermNetNZ were used in the experiments. The multilayer perceptron was run for 100 iterations and resulted in accuracy, sensitivity, and specificity of 87.32%, 90.34%, and 33.29%, respectively.

2.8 Discussion

We can notice from the above that several attempts have been made to automate the ABCD rule features measurement to come up with more objective evaluations of suspicious lesions (i.e. melanoma). In measuring the *A (Asymmetry)* feature, there has been reliance on statistical measures such as convexity and the symmetry distance, which in turn depend on the segmentation performance and the real shape of the lesion to be measured, respectively. New measures for asymmetry evaluation have also been proposed, such as the asymmetry index, where a threshold is derived using this measure and lesions are subsequently classified based on the threshold. Some studies evaluate asymmetry geometrically by dividing the images into sectors and building their assumptions based on those sectors. Others specify two symmetry axes around which the tumor would be rotated, and then an estimation of shape and texture asymmetry is made based on the overlapping pixels. The use of descriptors to evaluate the shape has been involved, such as Fourier descriptors. Classifying lesions in either symmetric or asymmetric has been mainly carried out using threshold or machine learning classifiers such as a support vector machine (SVM) where such classifiers are being trained on the extracted measures, and the learned model is used to classify new/test skin lesions. As a prerequisite to extracting asymmetry measures, some studies segment the skin lesion to form a region of interest (ROI) from which the measures are extracted and asymmetry is evaluated. The choice of the segmentation approach is thus crucial in determining the accuracy of the asymmetry evaluation. For instance, the multi-stage segmentation approach proposed in [53] provides poor results that degrade the performance of the evaluation as hair was present in the segmentation output.

Different research gaps have been encountered in the methods proposed for measuring asymmetry. Measuring asymmetry depends on the segmentation performance (where classical segmentation approaches are not robust to inherent skin lesion image problems such as noise and the presence of artifacts), and how the axes locations are specified. The asymmetry index is heuristic, and the obtained threshold is limited to the tested images and cannot be generalized to other images. The symmetry distance (SD) can be computationally expensive when there are hundreds of points on the skin lesion border. Evaluation was mainly performed on the extrinsic shape of the skin lesion, ignoring the asymmetry of the inhomogeneous pigmentation inside the lesion. Finally, the best-fit ellipse imposes an orthorhombic symmetry that the lesion itself might not possess.

In determining the *B (Border irregularity)* feature, border detection is considered a crucial prerequisite for characterizing this feature; different methods have been used for border detection such as the radial search algorithm [49] and Canny edge detector [83, 84]. Reconstructing the border using Fourier descriptors captures the general shape of the skin lesion but not the high-frequency information that delineates the border irregularities. In measuring border irregularity, most studies rely on statistical (e.g. area, perimeter) and geometrical (e.g. fractal geometry) features which have been extracted from the detected border. In classifying skin lesions as possessing irregular borders, some studies used a threshold-like measure (i.e. irregularity index

[57, 60, 204]), and other studies used a machine learning approach where the model is learned to predict border irregularity on new skin lesion images (i.e. Ma et al. [62] used a neural network as a classifier). The drawbacks of the approaches proposed for measuring border irregularity can be summarized as follows: Lack of robust lesion border detectors which are a required prerequisite before analyzing border irregularity. Euclidean based measures were mainly used in measuring border irregularity which don't characterize the complexity of the skin lesion shape, with other measures being empirical and prone to error, rely on the correlation of image pixels - which can be distorted when carrying out augmentation operations - such as in the case of entropy, and sensitive to noise, rotation, and scale. Small training samples were also used in classification (i.e. Ma et al. [62] used only 18 training samples for the neural network).

Different approaches have been proposed for determining the *C (Color variegation)* feature, such as segmenting the image into color objects and using heuristic rules in deciding the presence/absence of color variegation. Unsupervised learning algorithms (i.e. Kohonen map, K-means clustering) were used to form color feature vectors, and a KL-PLS based classifier was then used to detect color variegation based on those vectors. A trial-and-error approach was followed to determine color variegation. Other approaches determined the presence of a melanoma suspicious color by measuring the distance (e.g. Euclidean distance) between the pixel in the skin lesion and the six suspicious colors, such that the pixel was considered to belong to the color based on a threshold, and the lesion is said to contain the color if the pixels belonging to the color form a portion (i.e. more than 5%) of the skin lesion pixels. Segmenting the skin lesion before analyzing color variegation is beneficial as it narrows the search to a specific ROI, and eliminates the presence of any color outside the skin lesion that might otherwise be counted as a melanoma suspicious color. However, the approaches proposed in literature for detecting color variegation in skin lesions don't accurately characterize the melanoma suspicious colors, and a non-representative color space of the human visual system was commonly used when analyzing colors.

In measuring the *D (Diameter)* feature, the reviewed studies mainly relied on geometrical approaches in measuring the diameter in pixels, and eventually converted the result into millimeters based on a conversion factor. Such approaches are approximative measures and don't reflect the actual diameter, in addition to suffering from a major limitation when it comes to representing the result in a way that would reflect the real world (i.e. in millimeters) especially when an appropriate spatial calibration method is missing. Segmentation is also considered a crucial step before measuring diameter. No classifiers (i.e. threshold or machine learning based) have been used in the reviewed studies.

Combining the ABCD features improves melanoma detection accuracy, sensitivity, and specificity. This also applies to combining the ABCD features with other features extracted from the skin lesion.

As opposed to hand-crafted features (i.e. statistical and geometrical features) which are apparently the type of features utilized while tackling ABCD rule automation, data-driven features derived from deep learning methods have been recently proposed in literature. However, the main drawback of using such features is their inability to detect fine structures (i.e. ABCD rule features) [84]. The earlier attempts (to the best of our knowledge) in applying deep learning to melanoma detection were proposed in 2015 in [89], [90] (in Japanese) and [91]. The main papers published in 2016 can be referred to in: [92] - [106], and those published in 2017 in: [107] - [125]. Since 2018 the number of published papers on the topic has increased dramatically. Multiple skin lesion datasets have been publicly available that could aid in diagnosing skin cancer in general and in melanoma detection in particular. Table.2.3 provides a summary of such datasets along with their size and address. In this thesis, deep learning was particularly used in skin lesion segmentation (chapter 4), represented by U-Net and an unsupervised deep learning based approach, and in border irregularity classification (section 6.6).

The main advantage of extracting the ABCD features using the automated methods discussed in this chapter is the ability to provide an objective second opinion to the investigator (i.e. physician) which would otherwise be prone to subjectivity, especially that the ABCD features demonstrate fine structures of the skin lesion. The drawback however is that many approaches rely on the segmentation performance which could be degraded due to the presence of different artifacts (i.e. hair), affecting thereby the feature extraction process. Moreover, machine learning based approaches suffer from limitations in data availability especially when dealing with individual features (i.e. A, B, C, D). The approaches also lack the ability to reflect real-world measures such as in diameter measurement.

Different challenges need to be addressed when developing melanoma detection approaches

Dataset	No. of Images	URL
ISIC (International Skin Imaging Collaboration)	23,906	https://www.isic-archive.com
HAM10000	10,015	https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/DBW86T
MED-NODE	170	http://www.cs.rug.nl/ imaging/databases/melanoma$_naevi$
DermIS	6,800	https://www.dermis.net
Dermatology Atlas	10,682	http://www.atlasdermatologico.com.br
Danderm	1,869	http://www.danderm-pdv.is.kkh.dk
Dermnet	23,000	http://www.dermnet.com
DermNetNZ	20,000	https://www.dermnetnz.org
Dermatoweb	7,300	http://www.dermatoweb.net
Hellenic Dermatological Atlas	2,663	http://www.hellenicdermatlas.com
Dermofit	1,300	https://homepages.inf.ed.ac.uk/rbf/DERMOFIT
PH^2	200	https://www.fc.up.pt/addi/ph2%20database.html
Interactive Dermatology Atlas	1000	http://www.dermatlas.net
Interactive Atlas of Dermatology	2000	http://www.dermoscopy.org/atlast
Atlas of Clinical Dermatology	3,000	http://www.danderm-pdv.is.kkh.dk/atlas/index.html
Dermatologic Image Database (University of Iowa)	440	https://medicine.uiowa.edu/dermatology/education/clinical-skin-disease-images
Dermatology Atlas (Loyola University)	230	http://www.meddean.luc.edu/lumen/MedEd/medicine/dermatology/melton/atlas.htm
Skin Cancer and Benign Tumor Image Atlas (Loyola University)	80	http://www.meddean.luc.edu/lumen/MedEd/medicine/dermatology/melton/content1.htm

Table 2.3: Skin public datasets

in general and automating the ABCD rule in particular. Although public datasets are available (small image datasets have been used in different studies until the emergence of public datasets such as "ISIC 2018: Skin Lesion Analysis Towards Melanoma Detection grand challenge datasets" [81, 82] that enable the use of more images when evaluating different approaches geared towards the early detection of melanoma), there is still shortage in the availability of less quality skin lesion images that are mainly taken via mobile phone cameras as a result of the increase in melanoma apps usage; this type of images will introduce a variety of problems image processing and machine learning researchers need to handle. Datasets pertaining the different features in the ABCD rule need to be built, taking into consideration the labor intensive work accompanied with building such datasets (i.e. pre-processing operations, feature annotation). From a logistical perspective, there need to be more coordination between researchers and dermatologists (i.e. feedback) and the availability of more machine power (i.e. GPUs) especially for machine learning based solutions, as this would increase the pace of research outcomes significantly.

2.9 Summary

This chapter highlighted the different approaches proposed in literature for finding the ABCD rule features. It has been shown that the ABCD rule is a simple framework that physicians, novice dermatologists and non-physicians can use to learn about the features of melanoma in its early curable stage, enhancing thereby the early detection of melanoma. Since the interpretation of the ABCD rule features is subjective, different solutions have been proposed in literature to tackle such subjectivity and provide objective evaluations to the different features. The proposed solutions involved the automatic dermoscopic image analysis standard pipeline main stages, namely segmentation, feature extraction and classification. Some studies used machine learning based approaches in the classification stage, and an increase in using such methods is expected. As opposed to using an individual ABCD rule feature, a combination of features could lead to better accuracy, sensitivity, and specificity in melanoma detection. However, research gaps in evaluating the ABCD features have been encountered in literature, and this thesis attempts to propose solutions that would fill-in those gaps.

Chapter 3

Background

3.1 Overview

This chapter introduces background concepts relevant to understanding the remainder of this thesis. A brief overview on fuzzy sets is provided in section 3.2, section 3.3 explains the saliency detection process, and the trimap and KNN matting concepts are discussed in sections 3.4 and 3.5, respectively. This is followed by an explanation of statistical and geometrical measures (section 3.6), segmentation methods (section 3.7), and classification methods (section 3.8) used in the thesis. The chapter is finally summarized in section 3.9.

3.2 Fuzzy sets

Let $U = \{x_1, x_2, x_3, ..., x_n\}$ be the universe of discourse, a fuzzy set $A \in U$ is defined as the set of ordered pairs $\{(x_i, \mu_A(x_i))\}$, where $x_i \in U$, $\mu_A : U \rightarrow [0,1]$ is the membership function of A, and $\mu_A(x) \in [0,1]$ is the degree of membership of x in A. Such fuzzy sets are called type-I fuzzy sets. However, this kind of fuzzy sets is unable to model different types of uncertainties since their membership functions are crisp. Membership functions of type-II fuzzy sets are on the other hand fuzzy and can model different types of uncertainties. A type-II fuzzy set $A^{'}$ is characterized by a type-II membership function $\mu_{A'}(x,\mu)$, where $x \in U$ and $\mu \in [0,1]$, and is defined as:

$$A^{'} = \left\{(x,\mu), \mu_{A'}(x,\mu) \,|\hbar x \in U, \mu \in [0,1]\right\} \tag{3.1}$$

where $0 \leq \mu_{A'}(x,\mu) \leq 1$.

Type-II fuzzy sets can be simply formed by firstly defining a type-I fuzzy set and assigning lower and upper membership degrees to each element in order to construct the *footprint of uncertainty (FOU)*, that is, the interval between the lower and upper membership values (Fig.3.1 depicts the concept). A type-II fuzzy set can be defined as [141]:

$$A^{'} = \{(x, \mu_U(x), x, \mu_L(x)) \,|\mu_L(x) \leq \mu(x) \leq \mu_U(x), \mu \in [0,1]\} \tag{3.2}$$

where μ_L and μ_U represent the *lower* and *upper* membership degrees of the initial membership function $\mu(x)$, respectively, defined as follows [141]:

$$\mu_L(x) = [\mu(x)]^{\alpha} \tag{3.3}$$

$$\mu_U(x) = [\mu(x)]^{\frac{1}{\alpha}} \tag{3.4}$$

where $\alpha \in (1,\infty)$. The range of values $\alpha \in (1,2]$ are recommended to use for image data since $\alpha \gg 2$ is usually not meaningful for such data [141].

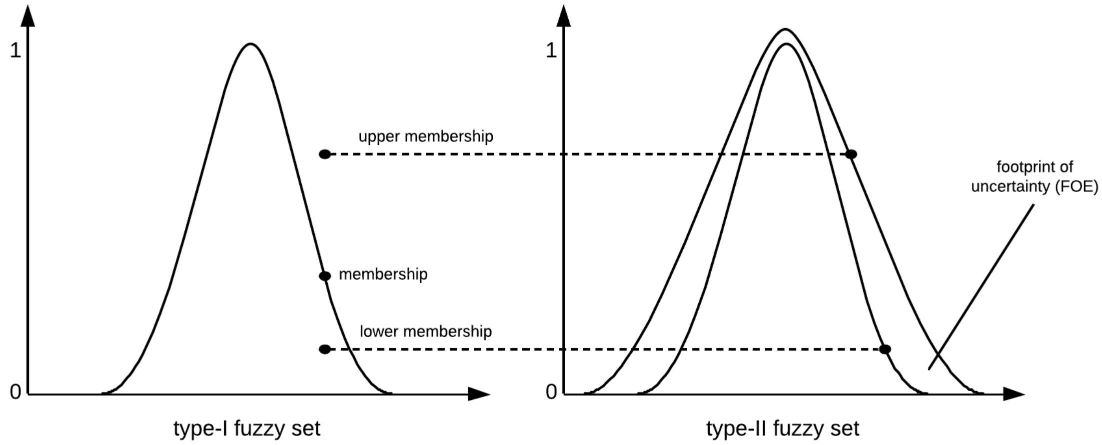


Figure 3.1: Creating a type-II fuzzy set

3.2.1 Fuzzy c-means clustering

Let $X = \{x_1, ..., x_i, ..., x_n\}$ be the set of n objects (i.e. pixels), $f(x_1, y_1), ..., f(x_i, y_j) : i \in [1, ..., m] ; j \in [1, ..., n]$, and $V = \{v_1, ..., v_i, ..., v_c\}$ be the set of c centroids in a p-dimensional feature space. In fuzzy c-means (FCM), X is partitioned into c clusters by minimizing the objective function J:

$$J = \sum_{j=1}^{n} \sum_{i=1}^{c} (u_{ij})^m \|x_j - v_i\|^2 \tag{3.5}$$

where $1 \leq m \leq \infty$ is the *fuzzifier* (set to 2 in this paper), v_i is the i^{th} centroid corresponding to cluster C_i, $u_{ij} \in [0, 1]$ is the fuzzy membership of x_j to cluster C_i, and $\|.\|$ is the distance norm, such that:

$$v_i = \frac{1}{n_i} \sum_{j=1}^{n} (u_{ij})^m x_j \quad where \quad n_i = \sum_{j=1}^{n} (u_{ij})^m \tag{3.6}$$

and,

$$u_{ij} = \frac{1}{\sum_{k=1}^{c} \left(\frac{d_{ij}}{d_{kj}}\right)^{\frac{2}{m-1}}} \quad where \quad {d_{ij}}^2 = \|x_j - v_i\|^2 \tag{3.7}$$

The process starts by randomly choosing c objects that represent the centroids (means) of the c clusters. Membership values u_{ij} are calculated based on the relative distance (i.e. Euclidean distance) of the object x_j to the centroids. The centroids v_i of the clusters are calculated after the memberships of all objects have been found. If the centroids at the previous iteration are identical to the centroids generated at the current iteration the process terminates [140].

3.3 Saliency detection

Saliency detection is the process of automatically locating the important parts of an image, where saliency refers to the unique features (i.e. pixels) of the image. In other words, saliency is a property of the image that stands from its neighboring pixels to the human eyes. The goal of saliency is to enable the segmentation approach to focus on a certain subset of visual information that matters most (i.e. lesion) to our task and omit other irrelevant information (i.e. skin). The output of the saliency detection step is a *map* where the intensity of each pixel represents the probability of the pixel belonging to the salient object. The Discriminative Regional Feature Integration approach (DRFI) is used for saliency detection [146],[147] as it is considered one of the most efficient algorithms for saliency detection [147]. DRFI consists of three main steps: (i) decomposing the image into multiple segmentations via supervised multi-level segmentation (a graph-based image partitioning algorithm is used) (ii) learning a random forest regressor to compute region saliency where the regional features (feature vector of each region) are mapped to a saliency score (iii) saliency maps across multiple layers of segmentations are fused together to generate a final saliency map.

In order to compute the *contrast* descriptor, each region belonging to a particular segmentation is represented by a feature vector which includes color (i.e. RGB, HSV, L*a*b*) and texture (i.e. LBP - Local Binary Patterns, LM - Leung Malik Filters) features that measure changes in a region appearance. A 29-feature vector is formed in this step, such that features extracted from each region are compared with features extracted from their neighboring regions. To represent more complicated textures present in the image, Jahanifar et. al. [147] added the response of Law's filter bank [148] to DRFI which emphasize more on edge, spot, ripple, and wave structures in the texture of the image. The regional *property* descriptor represents generic properties of the region which include appearance (distribution of colors and features in the region that characterize the common properties of the salient object and background) and geometric (size, shape, and position of a region that describe the spatial distribution of the salient object and background) properties, yielding a 35-feature vector descriptor. Jahanifar et. al. [147] added 5 colors (average of R, G, B, a*, and b* channels), 3 shapes (elongation, extent - dividing the region area by the area of its minimum area bounding box, and circle probability), and 14 texture-related features (Law's filter responses), extending the regional property descriptor by 22 features in order to improve skin lesion detection, in addition to applying color constancy to reduce the color variation in the image dataset using the Shades of Gray algorithm [149]; this makes the range of colors more consistent with other images in the dataset. Regional property features come in handy when the contrast features are insufficient such as the presence of a colored chart that could be misleading when determining the salient object, especially that it demonstrates a high contrast against the skin but it is not the salient object we are looking to specify (i.e. lesion). Since it is not enough to only use the property features to check if the region belongs to the background, that is identifying the background, a pesudo-background region is defined as the 15-pixel wide border region of the image from which the *backgroundness* descriptor is computed for each region. The backgroundness descriptor is composed of a 29-feature vector. As in the regional contrast descriptor, DRFI finds the difference of each region's features from the pseudo-background features to form the regional background descriptors. Jahanifar et. al. [147] introduced a new pseudo-background region in order to better distinguish between the skin lesion and background. A 93-dimensional vector is thus used to represent each region in total; the feature vector is illustrated in Fig.3.2.

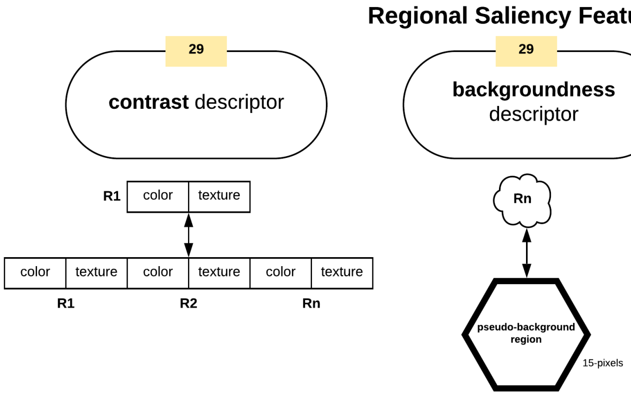


Figure 3.2: Regional saliency features (R_n: region number)

Labels of each region are determined by the groundtruth segmentation mask (0: region does not belong to the salient object, 1: region belongs to the salient object). Label vectors and regional features are then used to train a random forest regressor which is used to predict a saliency score (values between 0 and 1) for the saliency feature vector of a region. Predicting the saliency scores for all image regions at a specific level, a saliency map is created. The saliency score of each region is assigned to its contained pixels at each segmentation level, generating M saliency maps which are eventually fused together to get the final saliency map. A region is considered confident if more than 80% of its pixels belong to either the salient object or background. Training samples are created through supervised multi-level segmentation and the similarity score of each adjacent region is learned to determine the probability that the adjacent region belongs to the salient object (lesion) or background (skin), such that similar regions will be grouped together in a hierarchical way. Those similar regions will be used to train the saliency regressor (estimator). Multi-level segmentation is generated based on the learned similarity of two adjacent regions to gather a large amount of training samples. The learning process is depicted in Fig.3.3. As a final step, Jahanifar et. al. [147] threshold the saliency map of an image to create an initial mask which is then refined through level set evolution to represent the borders more properly. It should be emphasized that as opposed to salient object detection that attempts to extract the salient foreground object from the background, segmentation attempts to partition an image into regions that express homogeneous properties.

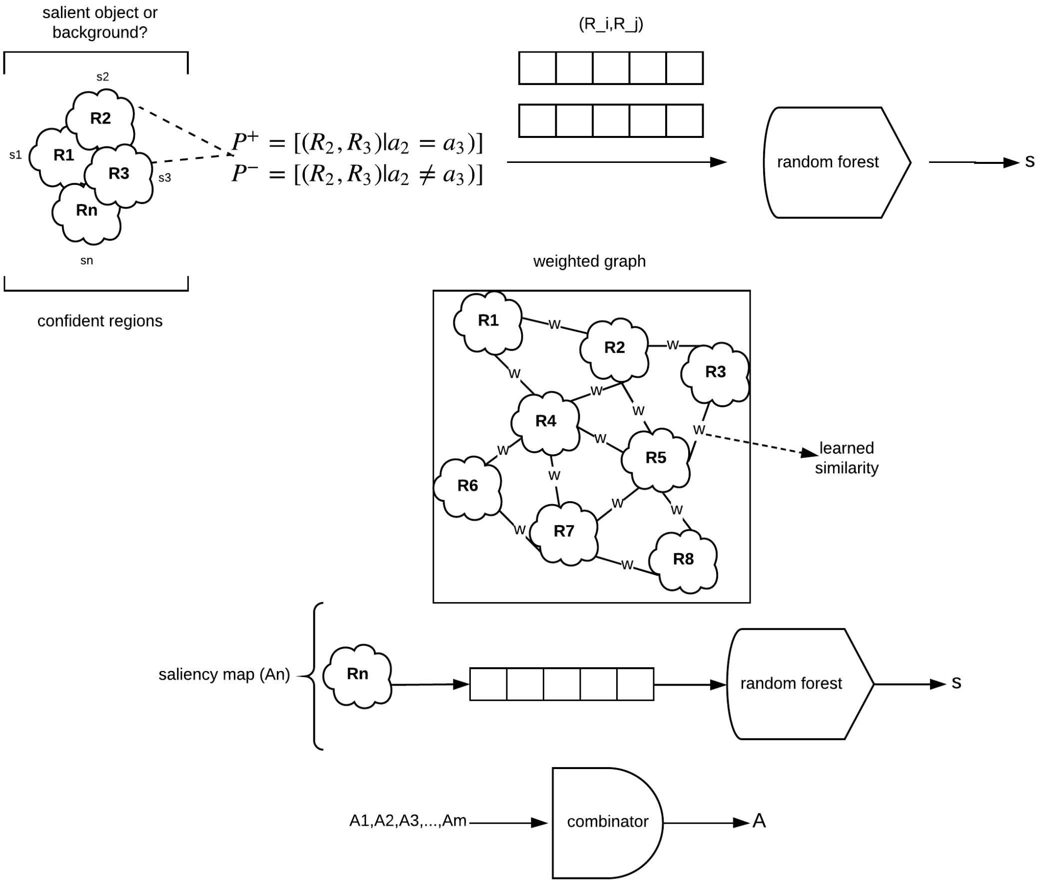


Figure 3.3: Learning the regional saliency regressor and combining the saliency maps - fusor - (R_n: region number, P+: positive part, P-: negative part, a_n: saliency label of R_n, s: saliency score, A_n: saliency map)

3.4 Trimap

A trimap means that each pixel in the image is assigned three possible values: definite foreground (lesion), definite background (skin), and uncertain (a mixture of foreground and background pixels). To create the trimap, erosion and dilation morphological operations are applied on the binary image mask of the skin lesions (i.e. salient objects). The trimap can be generated using the following formula:

$$M(R) = F(E) \bigcup B(D) \bigcup M(G) \tag{3.8}$$

where $M(.)$ refers to a set of pixels in the image, and $F(.)$ and $B(.)$ are the functions that extract the foreground and background pixels, respectively. E, D, G, and R denote the eroded image, dilated image, the gap between the foreground and background resulting from the morphological operations, and the trimap, respectively. $F(E)$ represents the foreground pixels (white), $B(D)$ represents the background pixels (black), and $M(G)$ represents the uncertain pixels (gray). Thus, the foreground in the eroded image is more likely to be the definitive foreground, and the background in the dilated image is more likely to be the definitive background. Moreover, labeling the G pixels as uncertain pixels can reduce the number of wrong labels in the trimap, and carrying out erosion and dilation operations allow us to remove holes and reduce noise in the binary image (i.e. image mask). As can be noticed, we can construct the trimap automatically by foreground/background segmentation and boundary dilation/erosion, where the trimap is represented as the union (combination) of the morphological operation results and the uncertain pixels. Such trimap will contain both a definitive foreground and a definitive background that will provide clues on the final segmentation result.

3.5 KNN matting

Matting is the process of finding (creating) the alpha matte/channel α_n which is used to accurately distinguish between the foreground and background, rendering the final segmentation. KNN (K-nearest Neighbors) matting [150] is represented as follows:

$$\alpha = KNN\{I, T\} \tag{3.9}$$

where the inputs I and T are the original input image and the corresponding trimap image, respectively. KNN matting basically specifies several neighbors for each uncertain pixel and forces them to have similar alpha values according to their distance in the feature space, as it assumes that pixels with the same appearance should be expected to have a similar alpha value. KNN finds the K nearest neighbors in a non-local neighborhood from the feature space and creates a matting from which the image foreground and background can be separated, provided that the alpha matte value is found for each pixel in the image. The main purpose of the matting process is to estimate the alpha values for the uncertain pixels under the constraints of the definitive foreground and background pixels. A more detailed description of KNN matting can be found in [150]. The trimap and matting processes are depicted in Fig.3.4.

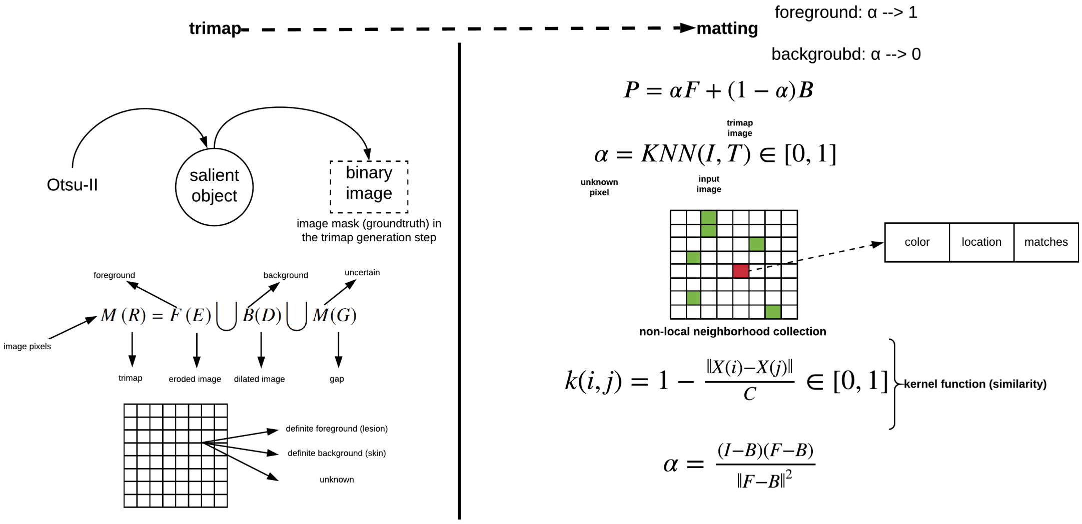


Figure 3.4: Trimap and matting processes (P: pixel color, F: unknown foreground layer, B: unknown background layer, $X(i)$: feature vector at pixel i, C: least upper bound of $\|X(i)-X(j)\|$)

3.6 Statistical and geometrical measures

3.6.1 SIFT based similarity

SIFT [166] is used as a sparse local descriptor where interest keypoints are detected in an image to describe invariant features (invariant to image scaling, rotation, and translation) in a local patch, such that an image will be represented as a collection of local feature vectors (shape descriptors) provided that a feature is a 128-dimensional vector representing a local region in the image. The goal of shape descriptors is to uniquely characterize the shape of the object and can be used to compare 2D object silhouettes. The group of feature vectors created by SIFT would thus represent the shape of the image. Using SIFT, we are detecting stable feature points in an image and then for each point a set of features that describe a small region around the point are provided, meaning that we are eventually extracting local information from digital images. Those extracted features are then used to match objects between different scenes.

Each keypoint in SIFT is a circular image region with an orientation, and described by three parameters: (i) keypoint center in the Cartesian plane (x,y) (ii) radius of the region (scale) (iii) angle expressed in radians (radius). SIFT features are computed over multiple scales, meaning that each pixel in the image will be compared with its 8 neighbors in addition to 9 pixels in both the previous scale and the next scale. The pixel is considered best represented in some scale if it was the local extrema amongst the pixels it was compared with (i.e. scale invariance). The SIFT feature is also expressed relative to its orientation (i.e. rotational invariance). Moreover, SIFT features possess robustness to some degree of affine transformations, noise, and illumination.

To extract features, SIFT applies a four stage filtering approach that consists of: (i) scale-space extrema detection (ii) keypoints localization (iii) orientation assignment (iv) building the descriptor.

In *scale-space extrema detection* potential locations for finding features are determined, such that locations and scales that are recognizable from different views of the same object are identified using a scale space function:

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \tag{3.10}$$

where $*$ is the convolution operator, x and y are the pixel location coordinates, $G(x, y, \sigma)$ is the Gaussian blur operator (Eq.3.11), σ is the scale parameter (amount of blur) such that the greater the value the greater the blur, and $I(x, y)$ represents the input image. The scale-space of an image is thus produced by convolving the input image with the Gaussian operator at different scales, producing the blurred image $L(x, y, \sigma)$.

$$G\left(x, y, \sigma\right) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2 y^2}{2\sigma^2}} \tag{3.11}$$

The blurred images resulting above are used to generate a new set of images, namely the Difference of Gaussians (DoG) which is used to detect the location of stable keypoints in the scale-space. The DoG is obtained by finding the difference between two subsequent Gaussian blurred images at different σ in the Gaussian pyramid in each octave. In order to detect the local maxima and minima of $D(x, y, \sigma)$, each point will be compared with its 8 neighbors on the same scale, and with its 9 neighbors below and above the current scale. The point is considered a local extrema if it was the maximum or minimum of all those points. Local extrema points are potential keypoints that are best represented in that scale. Images above and below the current scale are called *octaves*, where each octave is an image with half the size of the image in the previous scale.

Many keypoints will be generated at this stage. However, some keypoints might lie at the edges or might have low contrast, rendering them as undesired features. As the keypoints detected so far are scale invariant since we are aware of the scale at which the keypoint has been detected (scale of the blurred image), *orientation assignment* is used to assign an orientation to each detected keypoint to make it rotation invariant. This is achieved by collecting gradient directions and magnitudes for all the pixels around each keypoint, then the most prominent orientation(s) are figured out in that region and assigned to the keypoint. Gradient magnitudes and orientations are calculated using Eq.3.12 and Eq.3.13, respectively.

$$m\,(x, y) =\sqrt{(L(x\ 1, y) - L(x-1, y))^2\ (L(x, y\ 1) - L(x, y-1))^2} \tag{3.12}$$

$$\theta(x, y) = \tan^{-1}\left(\frac{(L(x, y\ 1) - L(x, y-1))}{(L(x\ 1, y) - L(x-1, y))}\right) \tag{3.13}$$

A histogram of 36 bins is then created for the 360 degrees of orientation, with each bin representing 10 degrees. The contribution of each keypoint neighboring pixel is determined by the gradient magnitude and a Gaussian window with a σ that is 1.5 times the scale of the keypoint of interest. The dominant orientations appear as peaks in the histogram, and a seperate keypoint is created for the direction that corresponds to the maximum histogram peak and any other direction within 80% of the maximum value.

As a final step, a *descriptor* is created for each keypoint by having a 16×16 (divided into 4×4 windows) window around the keypoint. Gradient magnitude and orientations are found within each 4×4 window, where the orientations are put into an 8-bin histogram. Each descriptor contains an array of 4 histograms around the keypoint, leading to a SIFT feature vector with 128 elements $(4 \times 4 \times 8)$.

3.6.2 Projection profiles

Projection profiles are data structures that are used to store the number of foreground pixels when the image is projected over the X-Y axes. They are one-dimensional representations of a two-dimensional image content, and are considered as compact representations of images as many useful information is retained in projections. In this approach, symmetry is measured by projecting the segmented skin lesion in the x and y directions and then comparing their histograms. Assume we have a binary image of size $M \times N$ (M: height, N: width), the projection of the image onto a line can be obtained by partitioning the line into bins and finding the number of 1 valued pixels that are perpendicular to the bin; horizontal and vertical projections can then be obtained by counting the number of 1 pixels for each bin in the horizontal and vertical directions, respectively.

The *horizontal projection* is the number of foreground (skin lesion) pixels in each row, and is defined as:

$$H[i] = \sum_{j=0}^{m-1} B[i,j]; \quad 0 < i < N \tag{3.14}$$

where $B[i,j]$ is the pixel value at (i,j), and $H[i]$ is the number of foreground pixels in the i^{th} horizontal row. Thus, for each horizontal line of pixels the number of foreground pixels are computed.

On the other hand, the *vertical projection* represents the number of foreground pixels in each column, and is defined as:

$$V[j] = \sum_{i=0}^{n-1} B[i,j]; \quad 0 < j < M \tag{3.15}$$

where $V[j]$ is the number of foreground pixels in the j^{th} vertical column. The vertical projection of each column is thus computed.

The horizontal and vertical projection profiles can be represented as a histogram as depicted in Fig.3.5. The values of each histogram represent the density distribution of the skin lesion.

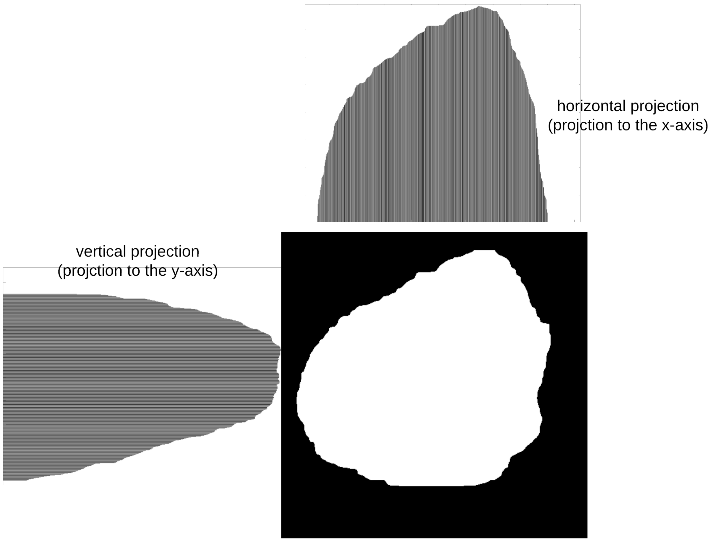


Figure 3.5: Horizontal and vertical projections histograms

After finding the two histograms they are compared using a *correlation* method. Let I denote the pixel intensity, $H_1(I)$ is the histogram which represents the number of pixels in the first image having pixel intensity I, and $H_2(I)$ is the histogram which represents the number of pixels in the second image having pixel intensity I. Using the correlation method, the two histograms are compared based on the following equations:

$$d(H_1,H_2) = \frac{\sum_I \left(H_1(I) - H'_1\right)\left(H_2(I) - H'_2\right)}{\sqrt{\sum_I (H_1(I) - H'_1)(H_2(I) - H'_2)}} \tag{3.16}$$

where,

$$H'_k = \frac{1}{N}\sum_I H_k(I) \tag{3.17}$$

is the mean value of each pixel in image k; N is the total number of histogram bins.

Symmetrical shapes (i.e. circle) evaluate to a correlation value of 1. The more asymmetrical the shape the less the correlation value.

3.6.3 Skewness

Image moments are useful in describing objects after the segmentation is carried out. They are scalar quantities that are used to capture the image's signficant features. A measure of asymmetry in an image can be given by its *skewness*, which is a statistical measure of a distribution's degree of deviation of the respective projection from symmetry. If the projection is symmetric with respect to the mean (origin), the corresponding skewness evaluates to *zero*. The *degree* of skewness can be determined using two third order moments: M_{30} and M_{03}.

To map from the image domain to the momenta domain, the *uniqueness theorem* of the momenta [167] can be used which states that the momenta sequence (general moment) M_{pq} is determined by the density distribution function (i.e. skin lesion image). A *general/standard moment* M_{pq} of an image $f(x, y)$ is defined as:

$$M_{pq} = \sum_{x} \sum_{y} x^{p} y^{q} f(x, y) \tag{3.18}$$

where $f(x, y)$ are the graylevels of individual pixels, p and q are positive integers, and $r = p\ q$ is the *order* of the moment.

The third order moment is the *skewness* of distances between the pixels in the image and its geometrical centre, measuring the bias of the distribution of pixels. The direction of the skewness can be obtained from the *sign* of the result of the moment, such that when the moment is *negative* the distribution will bias towards the left of the centre, and when *positive* will bias towards the right of the centre [168].

The skewness of the horizontal and vertical projections can be defined as shown in Eq.3.19 and Eq.3.20, respectively.

$$skewness_{H} = \frac{M_{30}}{\sqrt{{M_{20}}^{3}}} \tag{3.19}$$

$$skewness_{V} = \frac{M_{03}}{\sqrt{{M_{02}}^{3}}} \tag{3.20}$$

3.6.4 Fractal dimension

Fractal dimension has been used in characterizing skin lesion border irregularity as in [195], [52], and [196]. Fractal geometry [197] describes the space-filling capacity of irregular borders which is considered size independent and does not require any smoothing operations of irregular borders for measurement to be possible [198], meaning that structures don't need to possess a perfect geometric shape. The *fractal dimension* is a mathematical parameter that can quantify the irregularity (roughness or smoothness) of a skin lesion border via an objective observer-independent value. It is related to the complexity of the shape associated with the border such that a higher fractal dimension would stand for a higher degree of complexity of the analyzed pattern. In a 2-dimensional system a straight line will have a fractal dimension of *one*, and more complicated lines (having fractal properties) will have larger dimensions [199]. In general, fractal objects are those whose ratios are not whole numbers but fractions. This leads us to conclude that if the irregular borders of melanoma have fractal properties then they would be described more accurately by fractal dimension than Euclidean measures (i.e. perimeter) [200]. The *box-counting* method [201] is used for estimating the fractal dimension of the skin lesion border, which is defined as:

$$D = \lim_{\varepsilon \to 0} \frac{\log N(\varepsilon)}{\log\left(\frac{1}{\varepsilon}\right)} \tag{3.21}$$

where $D = [1, 2]$ is the box-counting fractal dimension of the skin lesion border, $\varepsilon > 0$ is the side (edge) length of the box, and N is the smallest number of boxes of side length ε needed to completely cover the skin lesion border (Fig.3.6). The fractal dimension is the *slope* in the $\log N(\varepsilon)/\log\left(\frac{1}{\varepsilon}\right)$ graph.

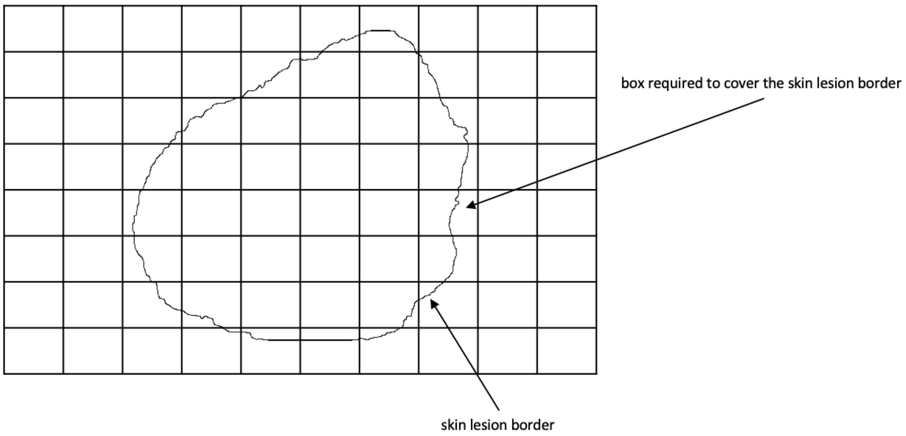


Figure 3.6: The box-counting method where 22 boxes are required to cover the skin lesion border

The lower the value D the straighter and smoother the skin lesion border, and vice versa. Melanoma borders, due to their irregularity, are more similar to fractals and are expected to have a higher fractal dimension than regular-boundary naevi. For instance, it was found in [198] that the fractal dimension of all lesions are greater than the topological dimension (i.e. one), which indicates that there exists a fractal element in their structure.

3.6.5 Zernike moments

Zernike moments are orthogonal moments, which means that no redundant or overlapping information exist between the moments, and are based on Zernike polynomials. They are invariant to rotation and are thus ideal for describing the shape characteristics of objects (i.e. skin lesions) [202][203]. Let (m,n) be a pair representing the Zernike polynomial order and the multiplicity (repetition) of its phase angle, respectively. The Zernike moment can then be defined as [202]:

$$V_{nm}(\rho,\theta) = R_{nm}(\rho)\, e^{im\theta}, \theta \leq 1 \tag{3.22}$$

where

$$\rho = \sqrt{x^2\ y^2}, \tag{3.23}$$

$$\theta = arctan\left(\frac{y}{x}\right) \tag{3.24}$$

$$R_{nm}(\rho) = \sum_{a=0}^{\frac{(n-|m|)}{2}} (-1)^a \frac{(n-a)!}{a!\left(\frac{n\lceil m\rceil}{2}-a\right)!\left(\frac{n-\lceil m\rceil}{2}-a\right)!}\rho^{n-2a} \tag{3.25}$$

ρ is the image pixel radial vector, θ is the angle between that vector and x-axis, and R_{nm} is the Zernike polynomial, which is an orthogonal polynomial equation over a circular space; polynomials are a function of the Cartesian coordinates (x, y) on the unit disc that are commonly expressed in terms of polar coordinates. In other words, the Zernike polynomial is defined in polar coordinates (ρ, θ) on a circle of unit radius $r < 1$. Zernike moment features describe the similarity of an input image to a set of Zernike polynomials. For an image $f(x_i, y_i) : 1 \le i \le M, 1 \le j \le N$, the Zernike moment Z_{nm} can be calculated as [203]:

$$Z_{nm} = \frac{n\ 1}{\pi}\sum_{i=1}^{M}\sum_{j=1}^{N} V_{nm}\left(x_i, y_j\right) f\left(x_i, y_j\right) \tag{3.26}$$

where $m = 0, 1, 2, 3, ..., \infty$ is the order of the Zernike polynomial, and n is the multiplicity of the phase angles in the Zernike moment, $x^2\ y^2 \le 1$. Zernike moments produce a 25-value vector (order n=8) as a description of the skin lesion contour.

3.6.6 Convexity

Convexity can be used to characterize the skin lesion border shape and irregularity [40], [204], [205]. It is the ratio between the perimeter (number of pixels divided by the length of the boundary) of the convex hull of the skin lesion (the smallest convex polygon that surrounds all of the skin lesion pixels) and the skin lesion perimeter. It shows the amount by which the object differs from the convex object. The convexity for the convex object evaluates to 1, and is less than 1 for non-convex objects (i.e. irregular skin lesion borders).

3.7 Segmentation

3.7.1 Otsu's method

The main goal of Otsu's method is to automatically, in an unsupervised manner, find an optimal global threshold (specific graylevel value) that would maximize the between-class variance or equivalently minimize the within-class variance, where variance can be perceived as a measure of homogeneity such that regions with high homogeneity will demonstrate a low variance. The threshold also transforms a grayscale image into a binary image; the image using this threshold is eventually divided into foreground (lesion) and background (skin).

Say we have two classes: skin (S) and lesion (M), the variance of the pixels in those two classes can be defined as:

$$\sigma_S^2(k) = \sum_{i=0}^{k-1} (i - \mu_S(k))^2 \frac{p_i}{P_S(k)} \tag{3.27}$$

$$\sigma_M^2(k) = \sum_{i=k}^{L-1} (i - \mu_M(k))^2 \frac{p_i}{P_M(k)} \tag{3.28}$$

where k is the graylevel value, $[0, L-1]$ is the range of graylevel (intensity) levels, p_i is the number of times pixel (graylevel) i occurred in the image which can be obtained from the image histogram, μ_S and μ_M represent the skin class mean and lesion class mean, respectively, which are defined as:

$$\mu_S(k) = \sum_{i=0}^{k-1} \frac{i p_i}{P_S(k)} \tag{3.29}$$

$$\mu_M(k) = \sum_{i=k}^{L-1} \frac{i p_i}{P_M(k)} \tag{3.30}$$

P_S and P_M represent the probabilities of class occurrence of the skin and lesion, respectively, defined as:

$$P_S(k) = \sum_{i=0}^{k-1} p_i \tag{3.31}$$

$$P_M(k) = \sum_{i=k}^{L-1} p_i \tag{3.32}$$

The *within class variance* which Otsu's method attempts to minimize by finding an optimal threshold is defined as:

$$\sigma_W^2(k) = P_S(k) \times \sigma_S^2(k)\; P_M(k) \times \sigma_M^2(k) \tag{3.33}$$

The *between class variance* on the other hand, which Otsu's method attempts to maximize, is defined as:

$$\sigma_B^2(k) = P_S(k) \times (\mu_S(k) - \mu_T(k))^2 \; P_M(k) \times (\mu_M(k) - \mu_T(k))^2 \tag{3.34}$$

where $\mu_T(k)$ is the total mean, defined as:

$$\mu_T(k) = P_S(k) \times \mu_S(k) \; P_M(k) \times \mu_M(k) \tag{3.35}$$

This is equivalent to Eq.3.36, that is subtracting the within-class variance from the total variance (σ^2) of the combined distribution.

$$\sigma_B^2(k) = \sigma^2 - \sigma_W^2(k) \tag{3.36}$$

The threshold k with the maximum between-class variance has also the minimum within-class variance.

The class separability η is:

$$\eta = \frac{\sigma_B^2(k)}{\sigma_W^2(k)} \tag{3.37}$$

In Otsu's paper [74] $\sigma_B^2(k)$ was measured for each graylevel value k (potential threshold) such that the optimal threshold k' is the one which maximizes $\sigma_B^2(k)$. In other words, $\sigma_B^2\left(k'\right)$ is the maximum. So, in Otsu's method we basically iterate through all the graylevel values such that at each graylevel value we separate the pixels into two clusters according to the graylevel (threshold). The variance is then measured for the pixels at each side of the threshold where pixel values less than the threshold belong to the skin (background) and those greater than the threshold belong to the lesion (foreground); the threshold k' will be that which maximizes $\sigma_B^2(k)$.

3.7.2 U-Net

U-Net [136] is an end-to-end encoder-decoder network for semantic segmentation which was firstly used for medical image segmentation and has also been utilized in skin lesion segmentation in dermoscopic images [86, 151]. The architecture is composed of left (down) and right (up) sides. The *down* part, which follows the typical convolutional network architecture, is the encoder part where convolution blocks are applied followed by maxpooling in order to encode the input image into feature representations at multiple levels. The number of feature channels are doubled at each downsampling step. The *up* part consists of upsampling the feature map followed by a convolution operation that brings the number of feature channels to half; a concatenation with the corresponding cropped feature map from the *down* part occurs, followed by two 3×3 convolutions which are also followed by two ReLU operations and one 2×2 max-pooling operation with stride 2 used for downsampling. The cropping process is essential as border pixels are lost at each convolution. While upsampling, the higher resolution features from the *down* part are thus concatenated with the upsampled features in order to localize and learn representations better. The resulting architecture is an expansive path which is symmetric to the concatenating path, yielding a u-shaped architecture. The final layer of U-Net uses a 1×1 convolution to map each 64 feature vector to the desired number of classes. The network is composed of 23 convolutional layers in total, provided that it does not have any fully connected layers and uses only the valid part of each convolution. For the border region of the image the pixels are predicted by an *overlap-tile strategy* such that the missing context is extrapolated by mirroring the input image, allowing the U-Net network to be applied on large images. The outcome of U-Net is a segmented image where the foreground (in white) represents the skin lesion and the background (in black) represents the skin.

3.7.3 Unsupervised segmentation

Unsupervised segmentation methods come in handy when labeled data is scarce (i.e. shortage of groundtruth). Since pixel-level annotations are considered difficult to get for image segmentation, weakly supervised learning approaches that utilize object bounding boxes [152, 153] or image-level class labels [154, 155] are widely used in the training process [156]. A CNN for image segmentation in a fully unsupervised manner (as proposed by Kanezaki [156]) has been used in segmenting skin lesion images. No training data and labels thus need to be available. This unsupervised setting could be crucial when it comes to skin lesion images (especially those taken using a mobile phone camera) that suffer from shortage in sufficient amount of data with defined ground truth that can be used for training a CNN.

Let $\{i_n \in \mathbb{R}^p\}_{n=1}^N$ be a set of p-dimensional feature vectors of image pixels, where N refers to the number of pixels in the input image. Cluster labels $\{c_n \in \mathbb{Z}\}_{n=1}^N$ are assigned to all pixels by $c_n = f(i_n)$, where $f : \mathbb{R}^p \rightarrow \mathbb{Z}$ is a mapping function. The parameters f and $\{i_n\}$ are trained in a fully unsupervised manner, predicting an unknown $\{c_n\}$.

Three constraints need to be met for the prediction of $\{c_n\}$: (i) feature similarity, (ii) spatial continuity, and (iii) number of unique cluster labels. In the *feature similarity* constraint, pixels with similar features are assigned the same label. A p-dimensional feature map $\{i_n\}$ (feature extraction) is computed from the input image through M convolutional components that are composed of a 2D convolution, p filters of size 3×3, ReLU activation function, and a batch normalization function (the batch is the N pixels in the input image). A linear classifier is applied to obtain a response map $\{y_n = W_c i_n \; b_c\}_{n=1}^N$, where $W_c \in \mathbb{R}^{q \times p}$ and $b_c \in \mathbb{R}^q$, which is eventually normalized to $\left\{y_n'\right\}_{n=1}^N$ with zero mean and unit variance. The cluster label c_n for each pixel is obtained by selecting the dimension that has the maximum value in $\left\{y_n'\right\}$ (argmax classification). Thus, the process clusters the feature vectors into q clusters and each pixel is assigned to the closest point amongst the q representative points, grouping similar pixels into clusters.

Based on the second constraint, the clusters of image pixels need1 to be *spatially continuous* (cluster labels of neighboring pixels are identical). Here, K superpixels (a group of connected pixels with similar colors) $\{S_k \in \mathbb{R}^n\}_{k=1}^K$ are extracted from the input image using Simple Linear Iterative Clustering (SLIC) [157], where S_k refers to the set of pixel indices that belong to the k^{th} superpixel, and each pixel belonging to some superpixel will have the same cluster label. The most frequent cluster label c_{max} for all $c_n \in \{1, ..., q\}$ is then selected. While the first two criteria aid in the grouping of pixels, the *third* criterion poses a constraint on the number of unique cluster labels, avoiding by that any undersegmentation (i.e. only one cluster) as this criterion gives preference to a large number of clusters.

The network starts by predicting the cluster labels with fixed network parameters (forward process of a network followed by superpixel refinement), and then training the network parameters with the fixed predicted cluster labels (backward process of a network based on gradient descent). Similar to supervised learning, the softmax loss is calculated between the responses $\left\{y_n'\right\}$ and the refined cluster labels $\left\{c_n'\right\}$, and error signals are backpropagated in order to update the parameters of both the classifiers and convolutional filters. This forward-backward process is iterated T times to obtain the final prediction of cluster labels $\{c_n\}$.

3.8 Classification

3.8.1 Decision trees

Decision trees offer a structured way of decision making in pattern recognition and are characterized by an order of set nodes, such that each of the internal nodes is associated with a decision variable of one or more features [169]. Decision trees are predictive models that work by finding the most informative covariate, partitioning the data based on the covariate, and then processing each partition recursively. The covariates selected after the training process to split the data can be represented as a tree-structure model. The wide usage of decision trees is due to their ease of implementation and interpretation and their good predictive performance. They can be used in regression analysis where the predicted result is most likely a real value number, and classification analysis where the predictive outcome is the class the data belongs to [170].

Taking classification decision trees as an example, such model consists of a group of $< covariate, partitions >$ pairs that are structured in the form of a tree, such that in each $< covariate, partitions >$ pair the covariate is represented as a tree node and each partition from partitions is represented as a branch from the covariate to the child node. The branch is a constraint of the node covariate such that the instances satisfying this branch's constraint are sent to the child node the branch links to. If a tree node has more than one branch, data instances are said to *split* on this tree node. If the tree node doesn't have branches, it is called a *leaf* node; the leaf node returns the final prediction. The *root* node of the tree has no incoming edges or branches [170].

Having a learned decision tree model, an instance is classified by starting at the root node of the tree, testing the covariate specified by this node, and eventually moving down the tree branch for which the constraint is satisfied by the value of the covariate. The process is repeated for the sub-tree rooted at the new node until it reaches a leaf node. A *decision tree learning algorithm* (i.e. ID3 [171], CART [172]), which is considered a searching algorithm in the space of all possible decision trees that can be built from data [170], is required in order to generate tree nodes and tree branches. Classification And Regression Tree (CART) has been used in the thesis.

3.8.2 Perceptrons

The perceptron is normally used in supervised linear classification tasks in which a hyperplane would be tuned to fit a training dataset. This tuned hyperplane can then be used to classify new unknown samples. This is achieved by minimizing the hyperplane's error as it is applied on the training dataset through minimizing the *error function*: $\epsilon\left(w\right) = -\sum_{i\in M} t_i \mathbf{w}^T \mathbf{x}_i$, where M is the set of misclassified samples, and $t_i \in \{-1, 1\}$ is the class of sample $\mathbf{x}_i$. If $\epsilon(\mathbf{w}) = 0$, this means that the hyperplane completely separates the classes. This minimization process is usually carried out in iterations such that after each iteration we move towards the minimum of $\epsilon(\mathbf{w})$. The $\mathbf{w}$ vector of iteration k 1 is obtained as the following weight updating step: $\mathbf{w}_{k1} = \mathbf{w}_k\ \Delta\mathbf{w}$ (weight update). Eq.3.38 shows the learning rule used in calculating the value for updating the weights at each increment:

$$\Delta\mathbf{w}_i = \eta\left(true_j - pred_j\right)x_i^j \tag{3.38}$$

where η is the learning rate, $true_j$ is the true class label and $pred_j$ is the predicted class label.

The perceptron's learning process starts by initializing the weights to small random numbers (or 0). For each training input sample the output value is calculated and the weights are updated until a minimum error is reached (i.e. backpropagation). The main drawback of perceptrons is that they are only able to converge when the two classes can be separated by a linear hyperplane.

A *multilayer perceptron* (also called Artificial Neural Network – ANN) is composed of neurons from the input layer, one or more hidden layers of neurons, and the output layer of neurons, where the input propagates through the network layer-by-layer in the forward direction where each layer of the network contains connections to the next layer. Such network is called a *feedforward* neural network and is typically used in supervised learning. The structure of the multilayer perceptron enables it to learn complex tasks by extracting more meaningful features from the input patterns. *Gradient descent* can be used to optimize model prediction by finding the local minimum of a function (i.e. minimize the network error), and is defined as follows:

$$\mathbf{w} = \mathbf{w} - \eta \times \frac{d}{d\mathbf{w}} F(\mathbf{w}) \tag{3.39}$$

where $\mathbf{w}$ are the weight values, η is the learning rate, and $\frac{d}{dw}F(\mathbf{w})$ is the derivative of the objective function $F(\mathbf{w})$ representing the slope (gradient).

3.8.3 Convolututional neural networks

Convolutional Neural Networks (CNNs) are analogous to Artificial Neural Networks (ANNs) in that they consist of neurons that self optimize through learning, such that each neuron would receive an input and perform an operation (i.e. scalar product) followed by a non-linear function. The neurons in the CNN are organized into height, width, and depth. Unlike ANNs, neurons within any given layer will connect to a small region of the preceding layer [207]. CNNs are thus considered a specialized type of neural networks that process data having a grid-like topology (i.e. images can be thought of as a 2D grid of pixels) [208]. CNNs are emerged from the study of the brain's visual cortex and have been used in image recognition since the 1980s. With the increase in computational power and amount of training data, CNNs are able to achieve superhuman performance on some complex visual tasks. The *convolutional layer* is considered the most important building block of the CNN. Neurons in the first convolutional layer are connected to pixels in their receptive fields as opposed to each pixel in the input image. Neurons in the second layer are thus connected to neurons that are located within a small rectangle in the first layer. Such architecture allows the network to focus on low-level features in the first hidden layer and then assemble them into higher level features in the next hidden layer, and so on, rendering CNNs to work well in image recognition tasks [209]. A neuron located in row i and column j of the feature map k in a given convolutional layer l is connected to the outputs of the neurons in the previous layer *l-1* located in rows i to $i\ f_h - 1$ and columns j to $j\ f_w - 1$, where f_h and f_w are the height and width of the receptive field, respectively. The neuron's weights (*filters* or *convolutional kernels*) can be represented as a small image of the size of a receptive field. A layer full of neurons using the same filter will give a *feature map* that highlights the areas in an image that are most similar to the filter. During the training process, a CNN attempts to find the most useful filters for the task and learns to combine them into more complex patterns. All the neurons share the same parameters (weights and bias) within one feature map. CNNs are composed of three types of layers: convolutional layers, pooling layers, and fully-connected layers. Stacking these layers together forms the CNN architecture as depicted in Fig.3.7.

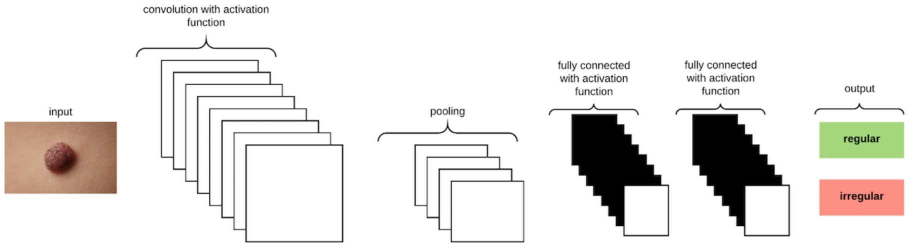


Figure 3.7: A CNN architecture composed of five layers

3.8.4 Gaussian Naive Bayes

The Naive Bayes classifier is a probabilistic classifier that applies Bayes' theory with strong (naive) independent assumption (i.e. independent feature model). The presence/absence of a particular feature of a class is not related to the presence/absence of any other feature. For instance, a skin lesion might be considered a melanoma if it has a larger fractal dimension and a smaller convexity. If those features depend on each other or upon the existence of other features (i.e. Zernike moments), the naive bayes classifier considers all those features to *independently* contribute to the probability that a skin lesion is considered a melanoma. An advantage of the naive bayes classifier is that it only requires a small amount of training data to estimate the *means* and *variances* required for classification. The probability model of the classifier can be represented (using Bayes' theorem) as:

$$p\left(C|F_1,...,F_n\right) = \frac{p\left(C\right)p\left(F_1,...,F_n|C\right)}{p\left(F_1,...,F_n\right)} \tag{3.40}$$

The probability model is a conditional model over a dependent class variable C with a small number of classes conditional on feature variables F_1 to F_n. The numerator is equivalent to the joint probability model: $p(C, F_1, ..., F_n)$. Since the denominator does not depend on C and the values of the features F_i are known, the denominator is considered constant. As each feature F_i is conditionally independent of every other feature F_j where $i \neq j$, the joint probability model can be written as:

$$p\left(C, F_1, \ldots, F_n\right) = p\left(C\right) p\left(F_1|C\right) p\left(F_2|C\right) p\left(F_3|C\right) \ldots = p\left(C\right) \prod_{i=1}^{n} p\left(F_i|C\right) \tag{3.41}$$

In *Gaussian naive bayes*, Gaussian distributions are used to represent the likelihoods of the features conditioned on the classes. Each feature can be defined by a Gaussian probability density function (having a Bell shape) defined as:

$$F_i \sim N\left(\mu, \sigma^2\right) \tag{3.42}$$

where μ is the mean, σ^2 is the variance, and:

$$N\left(\mu, \sigma^2\right)(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{3.43}$$

3.9 Summary

This chapter presented the theoretical background of the different concepts and methods used in the remainder of this thesis. The terms fuzzy sets, type-II fuzzy sets, and fuzzy c-means clustering were discussed. Saliency detection, as a method for locating the important parts of an image, was thoroughly explained. The method of finding the alpha channel, which is the result of the trimap and KNN matting processes, has been analyzed. Different statistical and geometrical measures were illustrated; those measures were used to form the asymmetry and border irregularity measures proposed in the thesis. An overview of Otsu's method, U-Net, and an unsupervised deep learning based approach have been introduced as segmentation methods. Decision trees, perceptrons, CNNs, and Gaussian Naive Bayes were introduced as classification methods.

Chapter 4

Skin Lesion Segmentation

4.1 Overview

The first stage in skin lesion detection CAD systems is usually image segmentation (Fig.1.3) due to its importance in determining the accuracy of subsequent stages in the detection pipeline. However, segmentation is considered a difficult task due to the great variety in lesion shapes, sizes, textures, and colors among different PSLs. Other difficulties are related to the presence of dark hair covering the lesions and the existence of specular reflections [127]. This chapter describes the image segmentation approaches proposed in my work. Section 4.2 provides a general overview of the different segmentation methods used in literature. The proposed gradual focusing and Otsu-II approaches are explained in sections 4.3 and 4.4, respectively, and compared with U-Net in sections 4.5 and 4.6, respectively. This is followed by a comparative analysis between an unsupervised deep learning based segmentation approach and U-Net in section 4.7, and the chapter is concluded by a summary in section 4.8.

4.2 Segmentation methods

In the last decades, hundreds of segmentation algorithms have been proposed in an effort to solve the segmentation problem through adopting different approaches. Elemental segmentation techniques use either boundary or region information. Boundary-based methods are often used to look for explicit or implicit boundaries between regions. This task can be driven from both following the image peaks or from a gradient plane. Elaborated techniques such as snakes proposed by Kass et al. [128] use the gradient information to deform a shape model initialized close to the object boundary in order to cope with edge discontinuities. Region-based methods are usually the first option in order to segment an image. The idea of most region-based approaches is to connect adjacent pixels with similar characteristics according to some user-specified criteria. Region growing, split-and-merge, morphological watershed transformation, fuzzy connectedness, or some active contour techniques (i.e. level set methodologies) come under region-based segmentation. In a poor image quality scenario however, the aforementioned classical segmentation techniques can easily fail due to inherent dermoscopic image problems, such as weak edges and the presence of artifacts like hair and light reflectance. Image segmentation methods can be categorized into: histogram thresholding, region growing, model-based (active contour, level set, Markov model), machine learning, and watershed methods [129].

The huge progress in machine learning, specifically in the area of deep learning, has recently revolutionized the image recognition and computer vision domains. Numerous studies have subsequently emerged employing machine learning in skin cancer and melanoma detection. The ultimate goal of applying machine learning to medical images is to recognize patterns in a better and quicker way than humans can, eventually increasing the productivity of doctors and patient healthcare outcomes. One such application is skin lesion segmentation. Jafari et al. [130] used a Convolutional Neural Network (CNN) that accepts input images (pre-processed for reducing the artefacts) and combines both local and global contextual information, eventually producing a label for each pixel which results in a segmentation mask that represents the skin lesion. The results have been further refined by some post-processing operations; comparisons to state-of-the-art algorithms have shown to outperform such methods in terms of segmentation accuracy and sensitivity, which were 98.5% and 95%, respectively.

A hybrid approach that uses convolutional and recurrent neural networks was proposed by Attia et al. [91]. The approach was tested on the ISIC (International Skin Imaging Collaboration) 2016 challenge [131] including 900 training images and 375 test images (no pre-processing has been made). The method achieved a segmentation average accuracy of 98% and Jaccard index of 93%, and demonstrated to be robust against different artificats with high sensitivity. The approach outperformed other state-of-the-art methods, including those using CNNs only.

Yuan et al. [132] proposed a fully automatic method that uses a 19-layer convolutional neural network for skin lesion segmentation. Moreover, the authors used a novel loss-function based on Jaccard distance which eliminates the need for sample re-weighting that is utilized when cross entropy is used as the loss-function. The approach was tested on the ISIC 2016 [131] challenge and the PH^2 [133] database, and showed to outperform other state-of-the-art methods.

Al-Masni et al. [134] proposed a segmentation method that uses Full resolution Convolutional Networks (FrCN) that learn the full resolution features of each individual pixel of the input image to achieve better pixel-wise segmentation of the skin lesions. The task was performed by eliminating the subsampling layers in the networks and enabling the convolutional layers in order to extract and learn full spatial features of the skin lesions, resulting in finely segmented contours of the lesions. No pre-processing or post-processing operations were used. The approach was tested on the ISIC 2017 challenge [81] and on the PH^2 database. The proposed method achieved an average Jaccard index and overall segmentation accuracy of 77.11% and 94.03% on the ISIC 2017 dataset, respectively. On the PH^2 datbase, the results were 84.79% and 95.08%, respectively. Compared with FCN [135], U-Net [136] and SegNet [137], the proposed method outperformed those networks by 4.94%, 15.47% and 7.48% for Jaccard index and 1.31%, 3.89% and 2.27% for the segmentation accuracy, respectively.

A deep learning architecture based on U-Net that uses multi-scale residual connections (Multiscale Residual U-Net) to maximize the learning capability and performance of the network was proposed in [138]. The residual connections preserve the information lost in the encoder stages due to max-pooling. A layer with binary cross-entropy loss function based on Jaccard index was included for the purpose of pixel classification. The method was applied on the ISIC 2017 challenge dataset and showed to have better performance compared to other methods in literature. The boundaries of lesion regions and background were well separated. Many other deep learning based skin lesion segmentation methods have been proposed; some of those methods can be referred to in [135]-[139].

4.3 Gradual focusing

This segmentation approach consists of the following main components [84]: (1) fuzzy c-means clustering (2) measuring the ambiguity threshold (inter-cluster threshold) that distinguishes between ambiguous and non-ambiguous pixels (3) revealing the ambiguous pixels (4) local treatment of the ambiguous pixels, and (5) final segmentation.

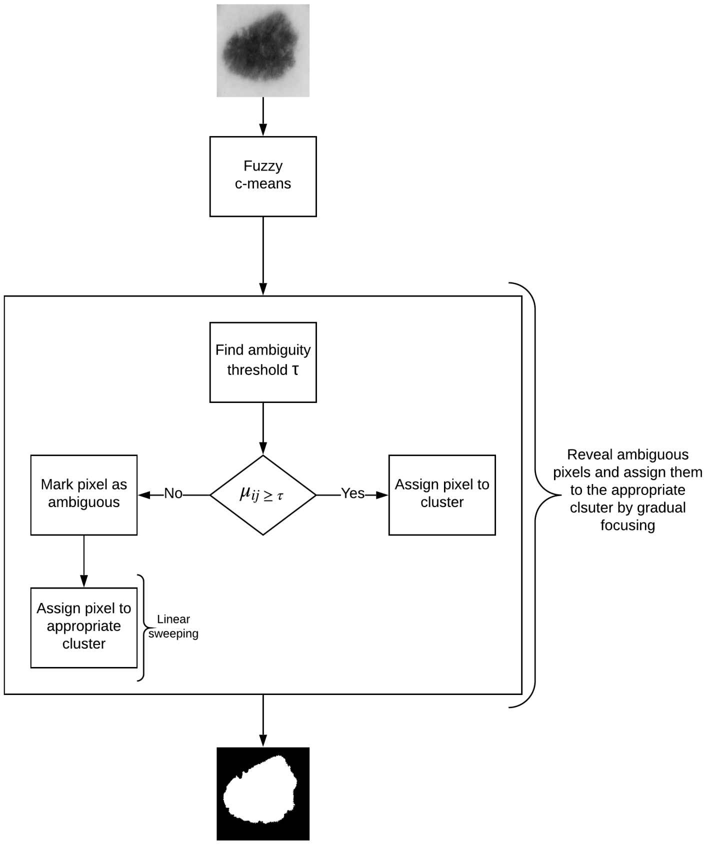


Figure 4.1: Defuzzification by gradual focusing

The measure of *ultrafuzziness* (linear index of fuzziness) $\tilde{\gamma}$ for an $M \times N$ image subset $\tilde{A} \subseteq X$ with L gray levels $g \in [0, L-1]$, histogram $h(g)$, and the membership function $\mu_{\tilde{A}}(g)$, can be defined as [141]:

$$\tilde{\gamma} = \frac{1}{MN} \sum_{g=0}^{L-1} h(g) \times [\mu_U(g) - \mu_L(g)] \tag{4.1}$$

The ambiguity global estimation is provided by the *ambiguity threshold* (τ). Algorithm.1 depicts the algorithm used for calculating the ambiguity threshold based on type-II fuzzy sets and the measure of ultrafuzziness [141]. The membership function used is the *S-function* (Eq.4.2) since it enhances the contrast of the fuzzy image (represented in terms of its membership values) and reduces the amount of ultrafuzziness [142].

```
Initialize the value $\alpha$ and determine the shape of the membership function
Find the image histogram
Initialize the position of the membership function, and shift it along the range of the
    gray-level values
At each position (gray-level value $g$), find the upper and lower membership values
    $\mu_U(g)$ and $\mu_L(g)$, respectively
Using Eq.4.1, find the amount of ultrafuzziness at each position
Find the maximum ultrafuzziness and use this value as the ambiguity threshold $\tau$
```

Algorithm 1: Measuring the ambiguity threshold

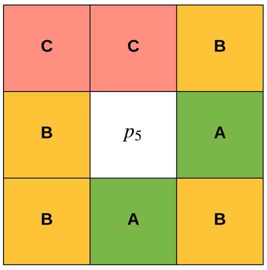
$$S(\mu; a, b, c) = \begin{cases} 0, & \mu \leq a \\ \frac{1}{2}\left(\frac{\mu - a}{b - a}\right)^2, & a < \mu \leq b \\ 1 - \frac{1}{2}\left(\frac{\mu - c}{c - b}\right)^2, & b < \mu \leq c \\ 1, & \mu > c \end{cases} \tag{4.2}$$

where $a = 0$, $c = 1$, $b = \frac{ac}{2} = 0.5$ (crossover point).

Different attempts have been made to measure the ambiguity threshold. However, such attempts suffer from limitations the proposed approach tries to overcome. In [143] to find the threshold a model of the membership function is found and the threshold is calculated with an $\alpha - cut$, such that the $\alpha - cut$ value is manually and heuristically chosen rather than in a systematic way as in the proposed approach. The choice of the most appropriate threshold using this method is thus very difficult. In Otsu's method [74] on the other hand, when calculating the threshold the histogram must be unimodal and does not take into account the level of fuzziness. By following the image thresholding algorithm proposed in Algorithm.1, the aforementioned limitations in calculating the ambiguity threshold are overcome.

In fuzzy clustering, the minimization of the function J (Eq.3.5) leads to partitions characterized by the membership degree matrix. A *defuzzification* step is required to obtain the final segmentation. While usually the data (pixels) get affected to the class with the highest membership degree, in skin lesion images such approach might not give appropriate results as lesion borders are sometimes not clearly defined (i.e. ambiguous).

The concept of *gradual focusing* (Fig.4.1), inspired by the human visual perception and introduced by Boujemaa et al. [144] proceeds in two steps: (i) membership values are compared with the ambiguity threshold τ to reveal the most ambiguous pixels (have a weak membership degree) from those that possess a high membership degree, in order to represent the coarse image information and locate the inner parts of the regions; ambiguous pixels are those that have a membership value smaller than τ (ii) weak pixels are affected to the appropriate cluster with regards to their spatial context. The notion of local ambiguity for a given pixel is introduced by considering a spatial criterion describing the neighborhood. The whole image has to be explored to deal with all the ambiguous pixels; *linear sweeping* is a method that can be used in such situation where the image pixel is affected to the major cluster of its neighbors. For instance, the weak pixel p_5 in Fig.4.2 evaluated against its neighbors in a 3×3 window will be affected to cluster B. If the cluster frequency is equal for each cluster (class) around the weak pixel, the pixel will be assigned to the original cluster it was classified to belong to. In other words, if we have more than one major cluster around the weak pixel, the assigned cluster to the pixel will be the one to which the pixel has the highest membership degree (i.e. the defuzzification step carried out by FCM). This process continues until all the weak pixels are treated.



C	C	B
B	p_5	A
B	A	B

Figure 4.2: Weak pixel p_5 will be assigned to cluster B

The use of fuzzy logic in gradual focusing stems from different motivations. Since fuzzy logic offers a suitable framework for modeling uncertainty, and dermoscopic images suffer from inherent ambiguity (i.e. how to accurately assign pixels to their relevant class), considering image features as fuzzy comes in naturally. Boundaries of skin lesions and the homogeneity of segments in dermoscopic images are usually vague and require fuzzy notions for their characterization. The contrast between the skin lesion (foreground) and the skin (background) can also be viewed as fuzzy in nature, since regions and edges in dermoscopic images are not always properly defined (i.e. no crisp boundaries exist between objects in the image). Classifying pixels into regions is thus considered a difficult task and is the rationale behind employing FCM in my work, which assigns each pixel to each cluster with some degree of membership reflecting thereby inherent issues in dermoscopic images where overlaps between clusters usually exist; this makes outlier cluster members more recognizable than by using hard (crisp) clustering since the degree of membership is continuous rather than binary, meaning that the fuzzy nature of FCM enables it to retain more original information than hard clustering.

4.4 Otsu-II

An improved version of Otionstsu's method [74] for skin lesion segmentation coupled with pre-processing and post-processing stages is proposed in [88]. In the proposed improvement which is referred to here as *Otsu-II*, the new measures $D_S(k)$ and $D_M(k)$ defined in Eq.4.4 and Eq.4.6, respectively are used instead of $P_S(k)$ and $P_M(k)$ (Eq.3.31 and Eq.3.32):

$$f_S(k) = ip_i; \{i \in \mathbb{Z} : 0 < i < k-1\} \tag{4.3}$$

$$D_S(k) = \sigma(f_S(k)) \tag{4.4}$$

$$f_M(k) = ip_i; \{i \in \mathbb{Z} : k < i < L-1\} \tag{4.5}$$

$$D_M(k) = \sigma(f_M(k)) \tag{4.6}$$

where σ is the standard deviation. σ_W^2 and σ_B^2 can be rewritten as:

$$\sigma_W^2(k) = D_S(k) \times \sigma_S^2(k)\ D_M(k) \times \sigma_M^2(k) \tag{4.7}$$

$$\sigma_B^2(k) = D_S(k) \times (\mu_S(k) - \mu_T(k))^2\ D_M(k) \times (\mu_M(k) - \mu_T(k))^2 \tag{4.8}$$

Thus, how spread out the pixel intensities are in each class is taken into account, such that when the pixel intensities are spread apart D will be large, and when they are tightly bunched together D will be small. Standard deviation is used since it expresses the statistical distribution of each class more accurately than variance, especially that dispersion of classes is measured as the distance between the mean of a class and any intensity value, which is proportional to the standard deviation rather than variance. This makes the optimal threshold less biased towards the larger variance among two class variances (i.e. larger dispersion of two classes) [145].

To improve the segmentation process, pre-processing and post-processing operations are proposed as depicted in Fig.4.3.

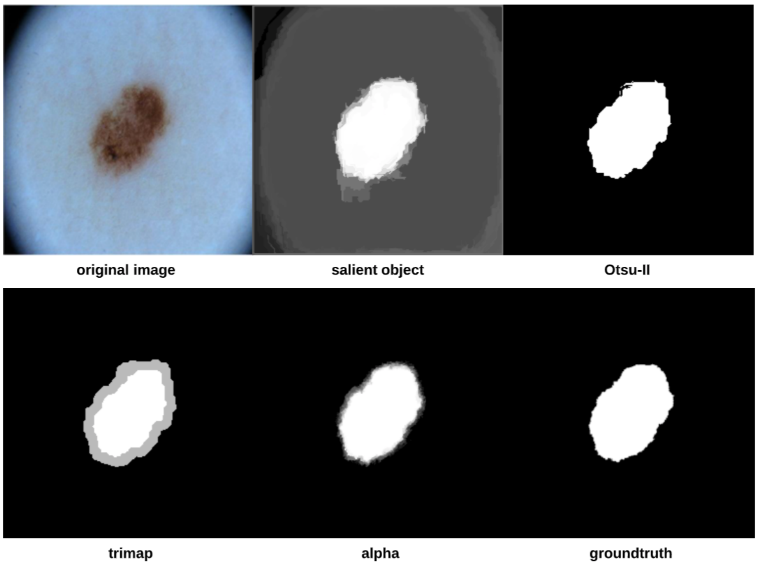


Figure 4.3: Otsu-II based segmentation

After detecting the salient objects, the improved Otsu method (i.e. Otsu-II) is applied on the salient objects to create binary images that will serve as image masks (i.e. groundtruth). The final step in the proposed approach is *matting*,

As we are going to measure asymmetry (and color variegation), especially when using SIFT, using the original *color* image we would like to focus only on the extracted skin lesion rather than the background and any accompanying artefacts for more accurate comparison of both halves. This can be achieved by merging the original image (i.e. color image) and its corresponding alpha matte described above. Fig.4.4 shows some examples on original images and their extracted lesions using this process.

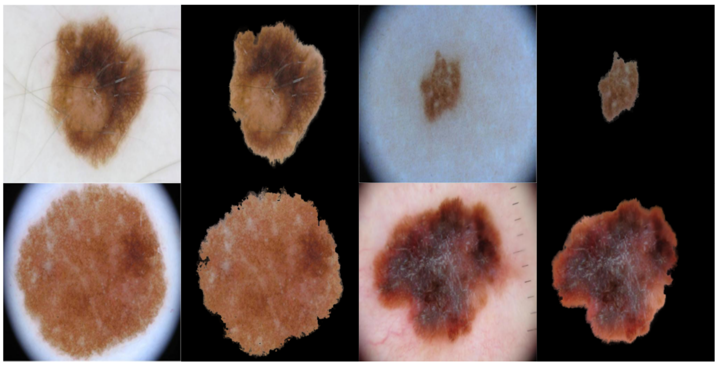


Figure 4.4: Original RGB image and the extracted lesion obtained by merging the original image and its corresponding alpha matte

4.5 Gradual focusing vs. U-Net

To compare gradual focusing with U-Net the two approaches were applied on 307 images from the "ISIC 2018: Skin Lesion Analysis Towards Melanoma Detection" grand challenge datasets [82, 158]. The U-Net architecture was trained on 2037 dermoscopy images along with their corresponding ground truth response masks (Fig.4.5 shows some samples of the training images). Images used to train U-Net were resized to 512×512 pixels, and the model was trained for 20 epochs on a Tesla P100 GPU. Training the model took 115.3 minutes and testing it on the 307 images took 46 seconds. Examples of test images along with their corresponding ground truth and segmentation results using the two approaches are depicted in Fig.4.6.

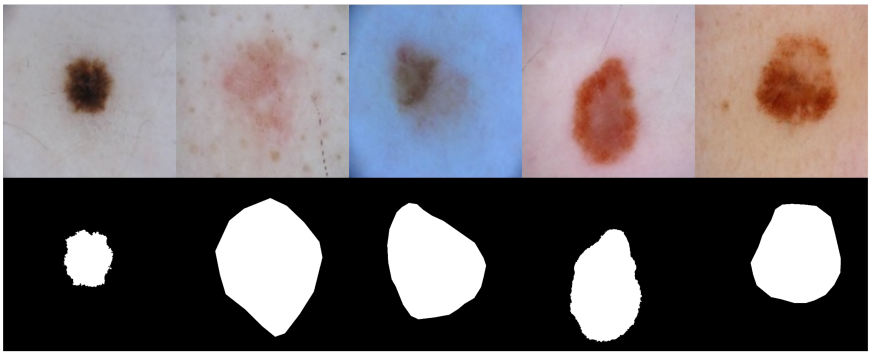


Figure 4.5: Samples of dermoscopy images (along with their corresponding ground truth) used to train U-Net

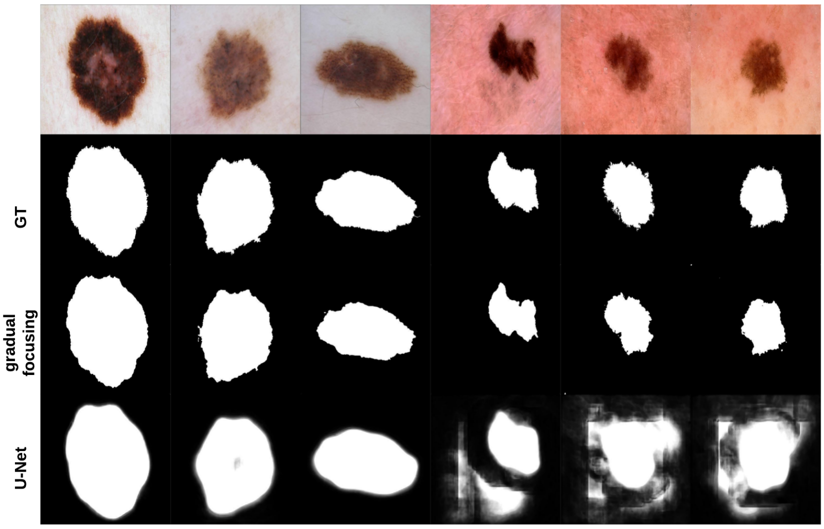


Figure 4.6: Segmentation results of gradual focusing and U-Net

The gradual focusing approach was able to detect the fine structures of the skin lesion borders as opposed to U-Net which lacks this ability. Detecting the fine structures is a crucial factor in determining the skin lesion border irregularity. Moreover, when the intensity between the background (skin) and skin lesion becomes closer (as in the last three images), U-Net produces noise in the segmentation results. The average Jaccard Index value of all the samples evaluated to 90.64% using the gradual focusing approach, while using U-Net evaluated to 58.31%. Jaccard index [159] is popular and frequently used as a similarity index. The area of overlap J is calculated between the segmented binary image A and its ground truth G as shown in the following equation [160]:

$$J = \frac{|A \cap G|}{|A \cup G|} \times 100\% \tag{4.9}$$

The value 100% means that the two values agree perfectly, while 0% means that there is no overlap.

4.6 Otsu-II vs. U-Net

The segmentation approach proposed in [88] was applied on 204 skin lesion images extracted from the "ISIC 2018: Skin Lesion Analysis Towards Melanoma Detection" grand challenge datasets" [82, 158]. Applying *Otsu-II* (the proposed Otsu thresholding improvement) and the original Otsu approach on the 204 images lead to a Dice similarity coefficient value of 87.7% and 82.5%, respectively. The Dice coefficient [161] is a measure that finds the spatial overlap between two binary images, resulting in a value that lies between 0% (no overlap) and 100% (agree perfectly). The Dice coefficient can be defined as:

$$D = \frac{2|A \cap G|}{|A \quad G|} \times 100\% \tag{4.10}$$

where A is the algorithm output and G is the ground truth.

Fig.4.7 shows some examples on segmentation results obtained by Otsu and Otsu-II methods, along with the ground truth of each corresponding image. As can be noticed, Otsu-II is able to improve the area coverage of the skin lesion.

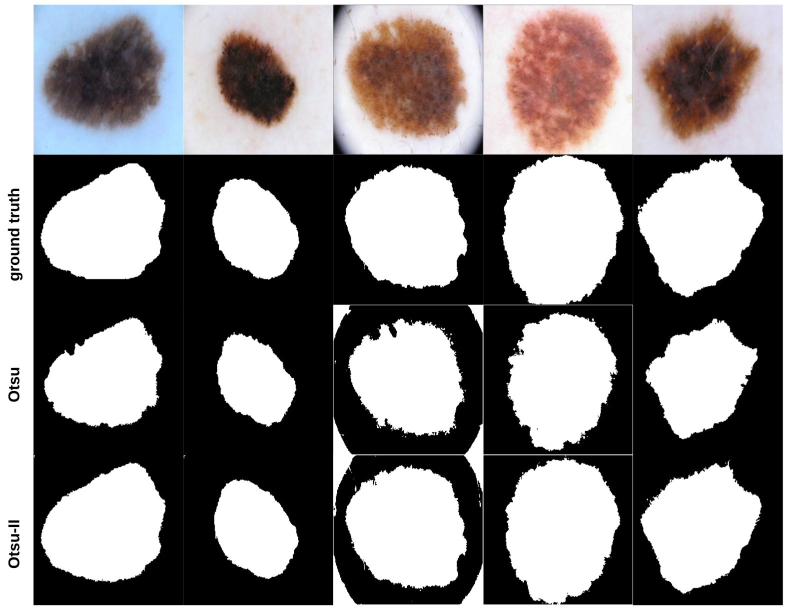


Figure 4.7: Skin lesion images segmented using Otsu and Otsu-II methods, along with their corresponding ground truth

Another adaptation to Otsu's method has been tested where $P_S(k)$ and $P_M(k)$ from Eq.3.31 and 3.32 were kept in Eq.3.33 and 3.34 (i.e. Otsu-II), resulting in the new within and between class variances shown in Eq.4.11 and Eq.4.12, respectively, which is referred to here as Otsu-II$'$:

$$\sigma_W^2(k) = D_S(k) \times P_S(k) \times \sigma_S^2(k) \; D_M(k) \times P_M(k) \times \sigma_M^2(k) \tag{4.11}$$

$$\sigma_B^2(k) = D_S(k) \times P_S(k) \times (\mu_S(k) - \mu_T(k))^2 \; D_M(k) \times P_M(k) \times (\mu_M(k) - \mu_T(k))^2 \tag{4.12}$$

Having three Otsu based methods, evaluation has been expanded to two more types of datasets. One dataset contained skin lesions that cover most of the image (i.e. large lesions) and another set was composed of skin lesions that comprise a small region of the image. The dataset used with Otsu and Otsu-II methods above contained a mix of such lesions (i.e. mixed). Fig.4.8 shows samples of large and small skin lesions. The datasets that contained images with large and small skin lesions were composed of 129 and 162 images, respectively.

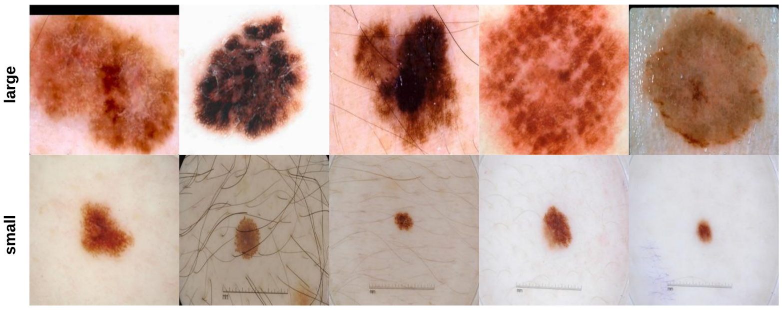


Figure 4.8: Samples of large and small skin lesions used to form two additional datasets, large and small, respectively

The Dice similarity of the three Otsu based methods on the different datasets is summarized in Table.4.1. The poor performance of Otsu-II$'$ on small images was due to the fact that it was not robust to artifacts (i.e. hair, ruler) that were more apparent in images with small skin lesions.

Table 4.1: Dice similarity of Otsu based methods on three datasets

Dataset	Otsu	Otsu-II	Otsu-II$'$
Small	71.2%	74%	49.6%
Large	84.6%	88.7%	88.5%
Mixed	82.5%	87.7%	85.6%

To evaluate the performance of the segmentation approach shown in Fig.4.3 against state-of-the-art methods, it has been compared with U-Net which was trained on 1935 dermoscopy images along with their corresponding ground truth response masks (Fig.4.9 shows some samples on the training images and their corresponding ground truth). Images used to train U-Net were resized to 256×256 pixels and the model was trained for 20 epochs on a Tesla P100 GPU. Training the model took 27.1 minutes and testing it on the 204 images took 18.1 seconds. Fig.4.10 shows some samples of test images, their corresponding ground truth, and the results using the proposed segmentation approach and U-Net. The Dice similarity on the 204 test images evaluated to 88% and 76.2% for the proposed segmentation approach and U-Net, respectively. Visual results (Fig.4.10) show that the proposed approach is able to detect the fine structures of skin lesion borders - a crucial factor when detecting skin lesion border irregularity (i.e. B feature in the ABCD rule) - better than U-Net. The results of U-Net on the other hand are blurry.

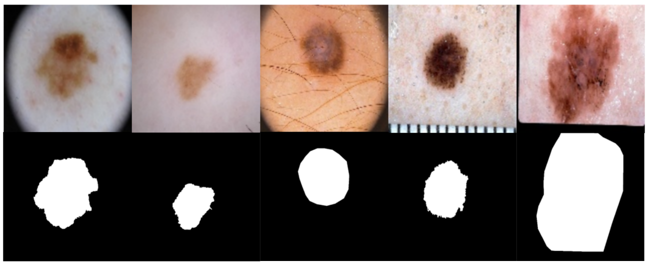


Figure 4.9: Samples of dermoscopy images along with their corresponding ground truth used in training U-Net

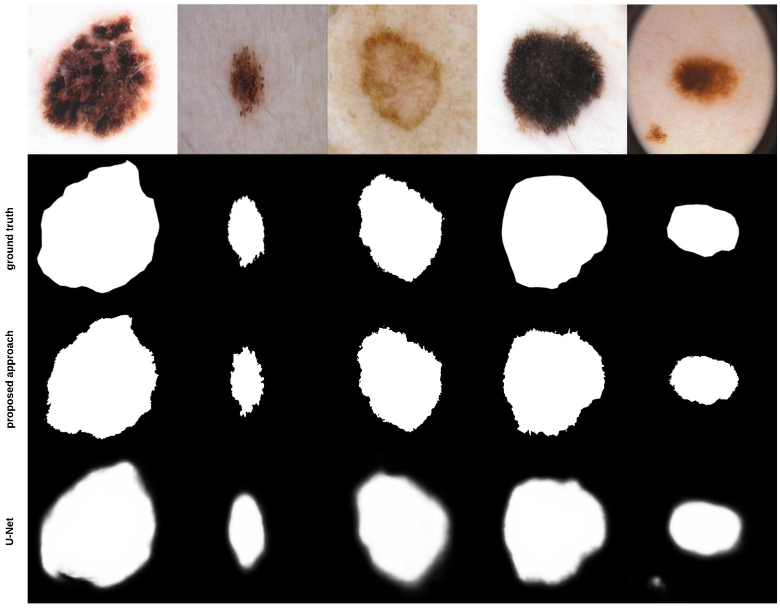


Figure 4.10: Skin lesion images segmented using the proposed segmentation approach and U-Net, along with their corresponding ground truth

4.7 Unsupervised segmentation vs. U-Net

In [151], the relative performance between U-Net (as a supervised deep learning based method) and an unsupervised deep learning based method [156] in segmenting skin lesions was evaluated. The U-Net architecture was trained on 2344 dermoscopy images along with their corresponding ground truth response masks from the "ISIC 2018: Skin Lesion Analysis Towards Melanoma Detection" grand challenge datasets [82, 158]. To make the most out of the training data, augmentation using some transformations has been applied (such as rotation, shifting horizontally and vertically, zooming, ...etc). This avoids the model from overfitting and leads to better generalization. Images used have been resized to 512 × 512. The U-Net model was trained for 20 epochs on a Tesla P100 GPU. Fig.4.11 shows how accuracy improved across the different epochs.

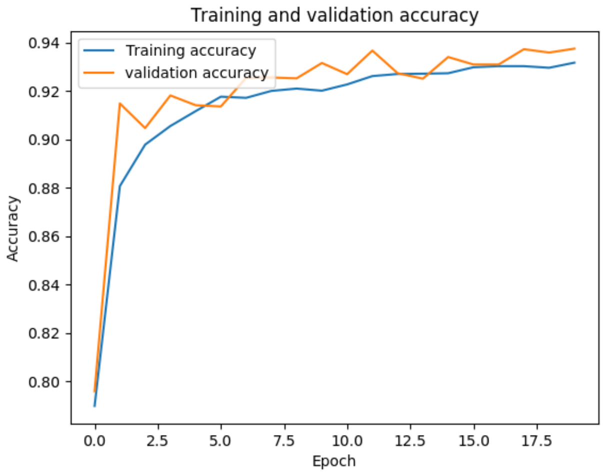


Figure 4.11: Accuracy improvement across different epochs

Different parameters have been set to the unsupervised deep learning approach (number of channels: 100, iterations: 1000, clusters: 2, learning rate: 0.1, convolutional layers: 2, super-pixels: 10000, and compactness of superpixels: 100); the approach has been run on a Tesla P100 GPU.

The approaches were tested on 250 images for segmentation. Fig.4.12 shows some examples on the tested images, corresponding ground truth, and the results of both approaches; the unsupervised approach shows less quality results than those of U-Net. This might be due to the close intensity values between the lesion and background which is apparent in the second image (counting from the left). This applies to the other images with different degrees of closeness. The presence of artifacts (hair) as shown in the fourth image could also be another reason of such poor performance. However, this is not the case with some other test images where the unsupervised approach could be better than U-Net in detecting the fine structures of the skin lesion. Fig.4.13 highlights some examples on such cases.

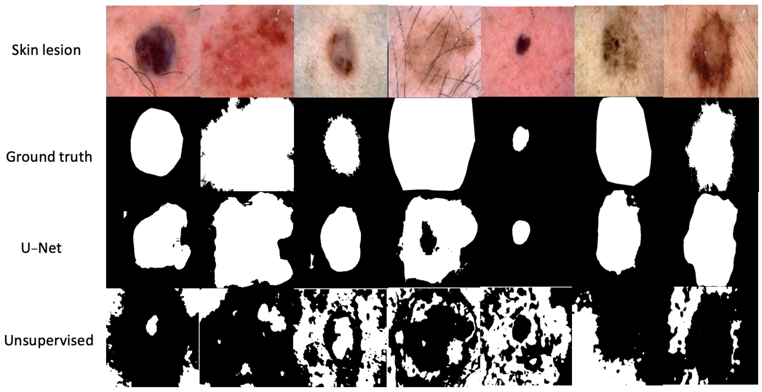


Figure 4.12: The unsupervised deep learning based approach shows less quality results than U-Net

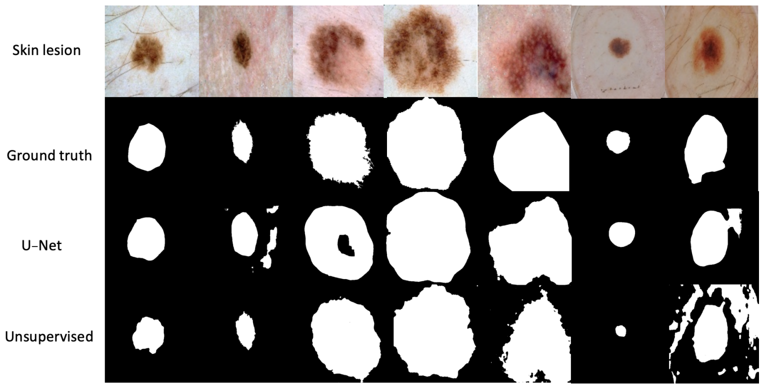


Figure 4.13: The unsupervised deep learning based approach can be better in detecting fine structures than U-Net in some cases

Table 4.2 shows the Dice coefficient values for U-Net and unsupervised based segmentation results shown in Fig.4.12, where image numbers reflect the order of the images from left to right. This applies to Table 4.3 that depicts the Jaccard index values of Fig.4.12. Dice coefficient and Jaccard index values of Fig.4.13 are shown in Table 4.4 and Table 4.5, respectively.

Table 4.2: Dice coefficient values for U-Net and the unsupervised based segmentation results shown in Fig.4.12. The image numbers represent the ordering of the images as read from left to right

Method	Image-1	Image-2	Image-3	Image-4	Image-5	Image-6	Image-7
U-Net	83.2%	86.2%	86.7%	77.5%	90.7%	78.5%	88.4%
Unsupervised	7.3%	13.1%	19.6%	33%	0%	18.3%	0.3%

Table 4.3: Jaccard index values for U-Net and the unsupervised based segmentation results shown in Fig.4.12

Method	Image-1	Image-2	Image-3	Image-4	Image-5	Image-6	Image-7
U-Net	71.2%	75.7%	76.5%	63.3%	83%	64.6%	79.2%
Unsupervised	3.8%	7%	10.8%	19.7%	0%	10%	0.2%

Table 4.4: Dice coefficient values for U-Net and the unsupervised based segmentation results shown in Fig.4.13

Method	Image-1	Image-2	Image-3	Image-4	Image-5	Image-6	Image-7
U-Net	94.7%	68.6%	89.8%	96.9%	91.4%	92.4%	85%
Unsupervised	75.5%	85.4%	94.4%	97%	89.8%	37.4%	52.4%

Table 4.5: Jaccard index values for U-Net and the unsupervised based segmentation results shown in Fig.4.13

Method	Image-1	Image-2	Image-3	Image-4	Image-5	Image-6	Image-7
U-Net	89.9%	52.2%	81.4%	94%	84.2%	85.8%	74%
Unsupervised	60.6%	74.5%	89.4%	94.3%	81.5%	23%	35.5%

The average Dice coefficient and Jaccard index values for the two approaches when applied on 250 test images are summarized in Table 4.6, which shows that U-Net performs better than the unsupervised approach in skin lesion segmentation.

Table 4.6: Average Dice and Jaccard values for U-Net and the unsupervised deep learning based approach

Method	Dice	Jaccard
U-Net	77.7%	67.2%
Unsupervised	40%	30.4%

Different variations of the two approaches have been used and tested on the 250 test images; evaluations are depicted in Table 4.7. For the unsupervised based approach, two variations have been made. The first variation consists of: (i) sigmoid is used instead of ReLU as an activation function in the convolutional components (ii) quick shift clustering [162] has been used instead of SLIC (iii) adam has been used as an optimization function instead of the stochastic gradient descent. The second variation consists of: (i) SELU (Scaled Exponential Linear Unit) [163] is used instead of ReLU (ii) using Felzenszwalb's efficient graph based image segmentation method [164] instead of SLIC.

For U-Net, two variations have been made. The first variation used tanh instead of ReLU as the activation function in the convolutional layers on both sides of the architecture (i.e. down and up). The second variation used SELU as an activation functions instead of ReLU on both sides of the architecture. Results show that the default setting of U-Net performs better on the test images (Tables 4.2-4.5), and that the unsupervised approach could be improved using some variations.

Table 4.7: Average Dice and Jaccard values for different variations of U-Net and the unsupervised deep learning based approach

Method	Dice	Jaccard
Unsupervised (sigmoid-quickshift-adam)	52.3%	41.8%
Unsupervised (selu-felzenszwalb-sgd)	54.3%	44%
U-Net(tanh)	68%	57.4%
U-Net(selu)	67.2%	55%

The unsupervised based approach is thus able to detect fine structures in skin lesions better than U-Net in some test samples. However, U-Net shows to provide better accuracy in terms of Dice coefficient and Jaccard index (77.7% vs. 40% and 67.2% vs. 30.4%, respectively). A proposed modification to the unsupervised approach can cause the Dice and Jaccard values to improve to 54.3% and 44%, respectively.

4.8 Summary

Classical image segmentation methods can easily fail in skin lesion segmentation due to inherent dermoscopic image problems, such as weak edges and the presence of artifacts like light reflectance, apart from the great variety of lesion shapes, sizes, textures, and colors. This chapter provided an explanation of the two proposed image segmentation approaches, namely gradual focusing and Otsu-II. Gradual focusing is inspired from the way the human visual perception works in that processing operates on two stages: global (locates the main regions of the skin lesion) and local (explicitizes region boundaries). Fuzzy c-means clustering was used in the global processing stage, and in the local processing stage a threshold was formed by which ambiguous pixels were revealed and eventually assigned to their appropriate clusters. Otsu-II is a proposed improvement on the traditional Otsu thresholding method, where the mean was replaced with standard deviation especially that the latter shows the spread of pixel intensities around the mean and is thus able to express the statistical distribution of skin lesion pixels. Otsu-II was coupled with preprocessing (salient object detection) and postprocessing (trimap and KNN matting) stages to form a robust image segmentation approach. Gradual focusing and Otsu-II outperformed U-Net, and were able to reveal the fine structures of the skin lesion borders in the segmentation results, which is a crucial factor in determining the skin lesion border irregularity. A comparative analysis was provided between U-Net as a supervised deep learning based approach and an unsupervised deep learning based approach, in which it was shown that U-Net provided better accuracy than the unsupervised approach.

Chapter 5

Asymmetry, Color Variegation, and Diameter

5.1 Overview

Skin lesion asymmetry is a strong indicator of malignant melanoma [165] such that the degree of asymmetry displayed by a skin lesion is indicative of its malignant potential. As opposed to benign pigmented skin lesions that are usually circular and symmetric, melanomas tend to develop in an anarchic fashion and grow at an irregular rate, rendering them to be asymmetric. In layman terms, asymmetry refers to the fact that when drawing a line through the middle of the mole the two halves will not match, meaning that the shape of one half doesn't resemble the other half (lopsided in shape), providing a warning sign to melanoma. There is no consensus on what asymmetry extent is required before one can tell that the skin lesion is considered asymmetric. This especially arises in moderately asymmetric lesions in contrast to the nearly symmetric lesions which dermatologists consider symmetric. It has been shown that the percentage of disagreement between dermatologists on the presence of asymmetry is around 5-10% [36]. A more objective measurement of asymmetry is thus deemed necessary.

Color variegation is considered the earliest sign of melanoma, and has a high predictability for the diagnosis of the disease. In fact, studies have demonstrated that color variegation might be the most important singular discriminator of melanoma [210]. It refers to the presence of two or more shades of pigment (two or more colors) within the skin lesion border. As opposed to benign lesions which tend to be generally uniform in color, melanoma lesions tend to often contain more than two colors, meaning that the color composition within the skin lesion is inhomogeneous. Melanoma in particular contains one or more of these six suspicious shades of color: white, red, light brown, dark brown, blue-gray, and black. The identification of colors in skin lesions is considered a subjective task even for experienced dermatologists, which deems it necessary to develop an automatic objective approach to identifying colors in skin lesions.

Most early melanomas tend to be larger than 6mm in diameter (i.e. size of a pencil eraser). Some studies argue that smaller diameters can exist in melanoma, which makes this criterion not absolute especially that 10% of melanomas tend to be missed in the early diagnosis (i.e. diameter <6mm) if the diagnosis was based only on diameter, making it necessary to use a computer vision system when evaluating diameter [217].

This chapter highlights the proposed approaches for determining asymmetry, color variegation, and diameter, which are explained in sections 5.2, 5.3, and 5.4, respectively. The chapter is summarized in section 5.5.

5.2 Asymmetry

A vector of three measurements has been built for measuring asymmetry. The vector will be used to train and test a decision tree which will eventually be used to predict the asymmetry of new skin lesion images. The extracted lesions (Fig.4.4) are then split vertically and horizontally across the centre into four equal halves, and SIFT is used to measure the image similarity (showing correspondences) between each opposite half (top vs. bottom and right vs. left) using the 128-dimensional local feature vectors. The total similarity score is the sum of the vertical (v_s) and horizontal (h_s) similarity, such that the similarity in each direction (i.e. vertical and horizontal) is normalized by dividing it by the number of keypoints in the other half (i.e. left vs. right, top vs. bottom). The greater the value the more similar the two halves of the skin lesion, and vice versa. Asymmetry means that the two halves are not similar. In measuring v_s and h_s, the best two matches for each keypoint are identified using OpenCV's brute-force matcher, where a feature descriptor on one image half will be matched with *all* feature descriptors on the other half and the best (closest) two matches will be returned based on the Euclidean distance. This is followed by the *ratio test* technique to remove any outliers (false positives) resulting from the brute-force matching step. Since the best two matches for each keypoint descriptor were selected, ratio test considers the match to be a good match if the distance ratio between the first and second match is smaller than a specific empirical value (chosen in this work to be 0.7 which is a typical value for Lowe's ratio). It should be emphasized that the asymmetry of *shape* is being investigated and not color, especially that dermatologists in general consider asymmetry in shape, color, and structural (border) distribution.

Fig.5.1 shows an example of an extracted skin lesion and its four halves (vertical and horizontal halves). The keypoints of each half (in red) are shown in Fig.5.2, and Fig.5.3 shows the lines between the matching keypoints between each half. The similarity between the left and right halves evaluates to 0.125 and between the top and bottom halves evaluates to 0.194. The total similarity in this skin lesion is: $(0.125\ 0.194) \times 100 = 31.9$, making it tend to be more asymmetric and the likelihood to be a melanoma lesion.

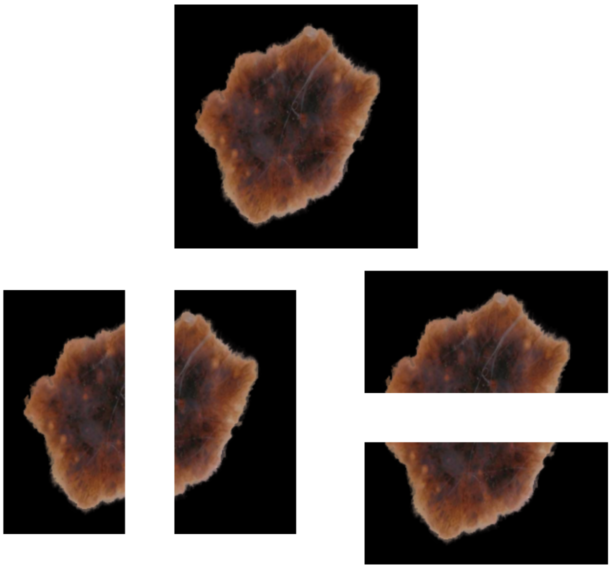


Figure 5.1: Extracted skin lesion and its four halves along the vertical and horizontal axes

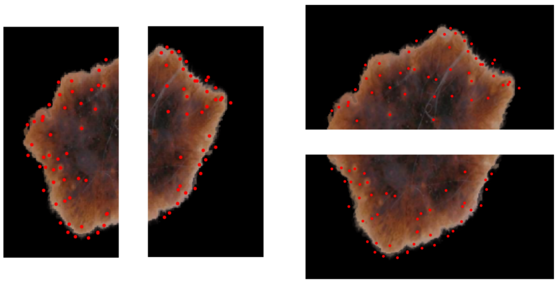


Figure 5.2: Keypoints (in red) of each half

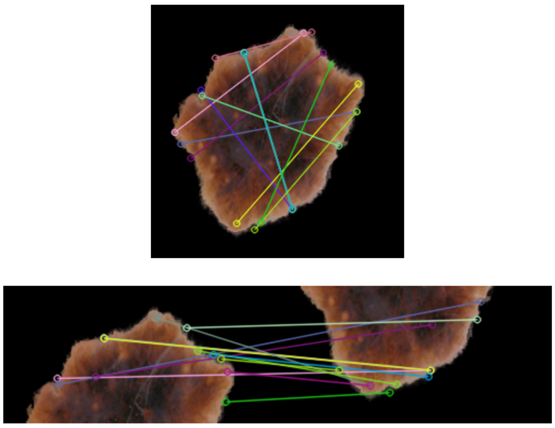


Figure 5.3: Matching keypoints between each half

204 skin lesion images were sent to a dermatologist to label as *symmetric* or *asymmetric* based on shape. Fig.5.4, Fig.5.5, and Fig.5.6 depict values of SIFT similarity, projection profiles, and skewness (horizontal and vertical projections) of the skin lesion images, respectively. The fluctuating values make it hard to come up with an accurate threshold that would aid in distinguishing between symmetric and asymmetric skin lesions, which is the reason a decision tree has been utilized to *learn* those asymmetric measures combined (i.e. vector).

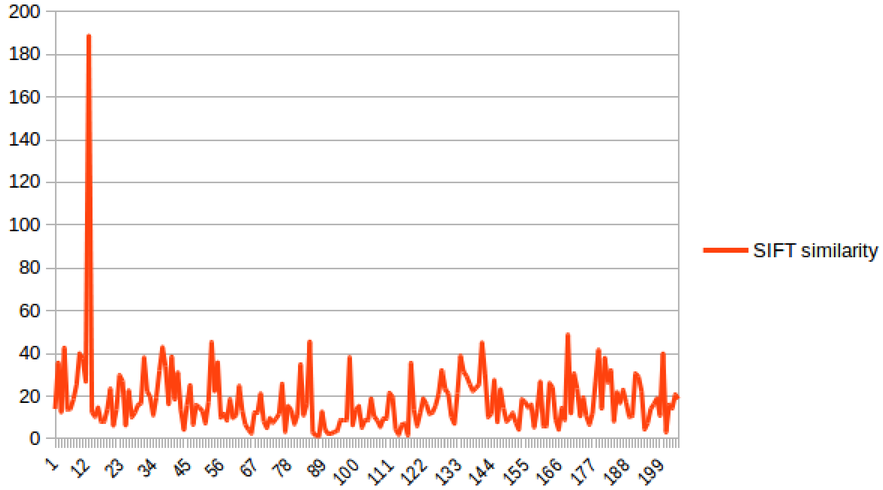


Figure 5.4: SIFT similarity values for the 204 images labeled by the dermatologist

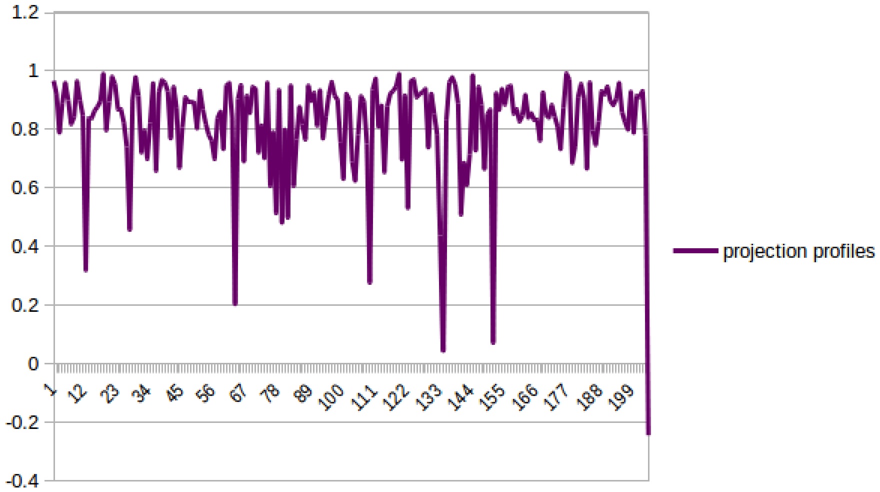


Figure 5.5: Projection profiles values for the 204 images labeled by the dermatologist

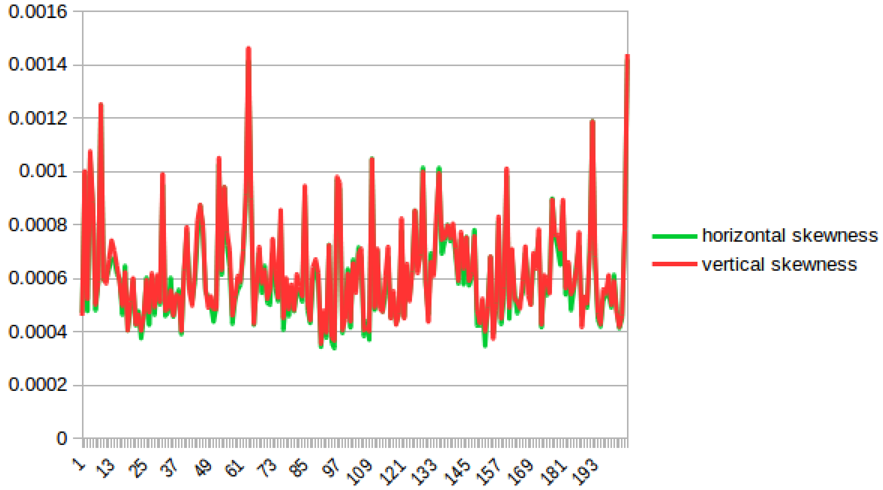


Figure 5.6: Skewsness of horizontal and vertical projections for the 204 images labeled by the dermatologist

Of the 204 images labeled by the dermatologist, 35 images were labeled as being *symmetric* and 169 images as *asymmetric*. Fig.2.1 shows some samples of symmetric and asymmetric images. The extracted asymmetry features/measurements (SIFT based similarity, projection profiles, and skewness) have been used to train and test a decision tree on an 80:20 ratio. That is, 80% of the data was used for training the decision tree and 20% of the data was used for testing the decision tree. 136 and 28 asymmetric and symmetric images were used for training the decision tree, respectively; 33 and 7 asymmetric and symmetric images were used for testing the decision tree, respectively. After training the decision tree, 30 and 2 asymmetric and symmetric images were predicted correctly, respectively, meaning that an 80% accuracy has been obtained. The decision tree performed better on asymmetric images; this can be due to having more asymmetric images in the training data. However, more balanced data samples (number of asymmetric and symmetric samples are the same) could improve the results significantly. This requires much labor work and is a topic of interest that could be investigated in a future work.

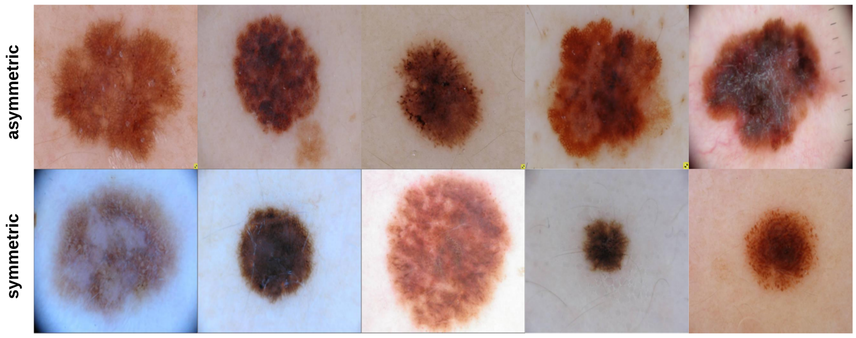


Figure 5.7: Samples of symmetric and asymmetric images labeled by the dermatologist

5.3 Color variegation

Before attempting to find color variegation (The C feature in the ABCD rule) in the skin lesion, the RGB image is converted to the CIELab (or CIEL*a*b*) color space [211] since the RGB color space does not closely match the human visual perception, whereas the CIELab color space is designed to approximate/model the human vision (i.e. the L component closely matches the human perception of lightness) and contains in theory every single color the human eye can perceive, allowing it to exploit the characteristics of the human visual pereception better. CIELab is device independent, meaning that the color model is based on the perception of the human eye and is designed to describe what colors look like regardless of what device they are displayed on. Colors in CIELab are described in three dimensions: L^* for lightness from black (0) to white (100), $a*$ from green to red, and $b*$ from blue to yellow. To thus be communicative with human perception, images used to find color variegation will be displayed in the CIELab color space.

To convert an image from RGB to CIELab, it has to firstly be converted to XYZ using Eq.5.1-Eq.5.3.

$$X = R \times 0.4124 \; G \times 0.3576 \; B \times 0.1805 \tag{5.1}$$

$$Y = R \times 0.2126 \; G \times 0.7152 \; B \times 0.0722 \tag{5.2}$$

$$Z = R \times 0.0193 \; G \times 0.1192 \; B \times 0.9505 \tag{5.3}$$

where R, G, B represent the colors of the red, green and blue channels, respectively. The (L^*, a^*, b^*) components are then computed using Eq.5.4-Eq.5.6.

$$L^* = 116 f\left(\frac{Y}{Y_n}\right) - 16 \tag{5.4}$$

$$a^* = 500\left(f\left(\frac{X}{X_n}\right) - f\left(\frac{Y}{Y_n}\right)\right) \tag{5.5}$$

$$b^* = 200\left(f\left(\frac{Y}{Y_n}\right) - f\left(\frac{Z}{Z_n}\right)\right) \tag{5.6}$$

where X, Y, Z are the tristimulus values of the color stimulus; X_n, Y_n, Z_n are the corresponding tristimulus values of the nominal white stimulus (those are constant values that can be found in tabulations of the illuminant spectra in texts such as [212]), and function f is defined as:

$$f(\rho) = \rho^{\frac{1}{3}} \; if \; \rho > \left(\frac{24}{116}\right)^3 \tag{5.7}$$

$$f(\rho) = \left(\frac{841}{108}\right)\rho \left(\frac{16}{116}\right) \; if \; \rho \leq \left(\frac{24}{116}\right)^3 \tag{5.8}$$

Fig.5.8 shows a skin lesion image and its 3D scatter plot representing the different pixel colors in CIELab color space.

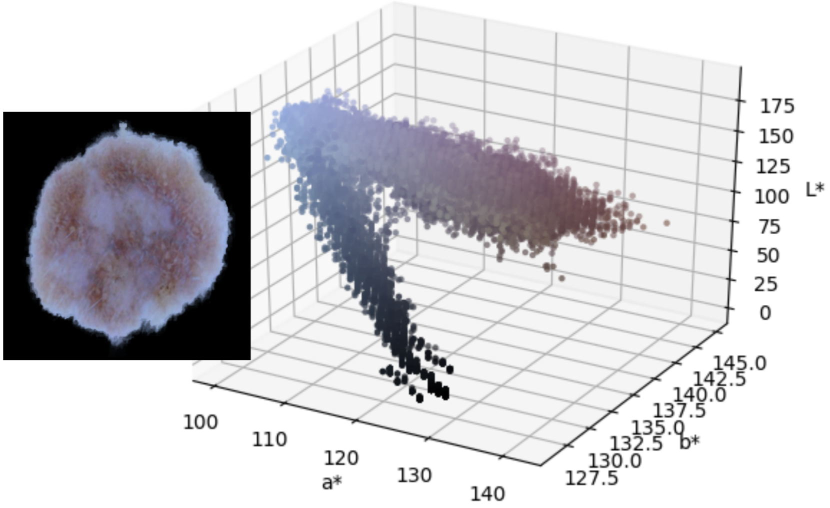


Figure 5.8: 3D scatter plot of skin lesion image pixels in CIELab color space

To determine the range of color shades that represent the six suspicious colors of melanoma based on the used image dataset, the color palettes of *all* images are extracted and analyzed based on which we are able to determine the range of shades that specify the suspicious colors (i.e. dark brown, light brown). In particular, each image in the dataset is represented in terms of its *dominant* colors which constitute the image's palette, such that those dominant colors would be the best possible colors by which we can display the image with the least amount of error. Clusters of dominant colors (group of pixels) are formed using k-means clustering [213, 214], where each pixel in the image has a CIELab value associated with it. The number of clusters is set to 7 (6 suspicious colors + pure black) since the pure black color (i.e. CIELab = [0,0,0]) is omitted in the measurement of the number of suspicious colors as it most likely belongs to the background (skin).

The process consists of applying k-means with a specific number of clusters (i.e. number of dominant colors) which is equal to the number of colors the color palette will be composed of. Each pixel color is then affected to the nearest cluster centroid according to the Euclidean distance. K-means minimizes the within-cluster sum of squared distances (i.e. Euclidean distance) between the centroid and the other pixels in the cluster. Using Euclidean distance in the CIELab space is uniform with difference perceived by the eye.

K-means finds the dominant colors in an image through an iterative corrective process, where colors can be thought of as points in the color space cloud that we aim to cluster around some mean (dominant color). K-means starts with a random palette of seven dominant colors $\{1, 2, ..., k\}$ as the starting point, where k is the number of clusters representing the dominant colors. Each pixel is then assigned a color label of the nearest dominant color. Image pixels are thus grouped by their dominant color value. New averages are then computed to update the cluster centers. If image pixels belong to the same clusters for two successive iterations, the process is considered complete and the final color palette is formed. Fig.5.9 depicts a skin lesion image and its corresponding color palette resulting from k-means clustering with $k = 7$. The CIELab values of the color palette from left to right are: $[0, 0, 0]$,[59.263,6.519,-1.826],[47.051,9.465,2.307],[67.222,3.355,-4.661],[22.237,0.779,-3.284],[40.018,10.696,4.107], [53.075,8.206,-0.027]. Fig.5.10 shows the dominant colors (based on 7 clusters) for the dataset used in the experiments, which is composed of 204 images; the process evaluated to 1221 *unique* colors (pure black color excluded).

Figure 5.9: A skin lesion image and its corresponding color palette showing the 7 dominant colors in the image (the first color which represents the pure black color will be omitted from the color variegation measurement)

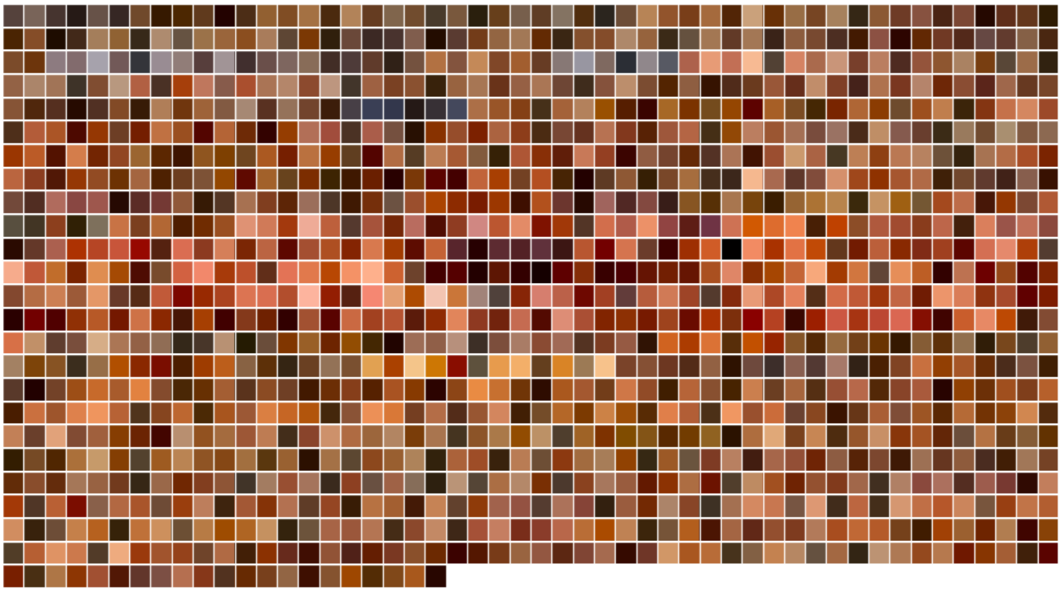


Figure 5.10: Unique dominant colors (1221 colors) of the 204 skin lesion image dataset

Representing the image in terms of its dominant colors is a more realistic approach when working on color variegation; for instance the image shown in Fig.5.9 is composed of 1480 unique colors, which makes it very difficult to determine the range of color shades that would represent the 6 suspicious colors of melanoma.

Analyzing the images used in the experiments, the suspicious CIELab color ranges (i.e. $[min] - [max]$) are determined as shown in Table.5.1. For the suspicious colors white, red and blue-gray, since they were not expressed in the utilized dataset, their standard CIELab color values are used. Skin lesions which have any of their pixel CIELab color value belonging to the range of black, dark brown, or light brown color values are considered to possess those colors. For the colors white, red and blue-gray, a skin lesion is considered to possess one of those colors if the (i) Minkowski distance between the pixel color and any of the aforementioned colors is less than a threshold T, measured as being the half of the Minkowski distance between the two extremes of colors (white and black), which evaluates to 50 (ii) pixels that belong to the suspicious color represent more than 5% of the skin lesion pixels.

Table 5.1: CIELab melanoma suspicious color values

Color	CIELab
Black	$[0.06, 0.27, 0.10] - [39.91, 30.23, 22.10]$
Dark brown	$[14.32, 6.85, 6.96] - [47.57, 27.14, 46.81]$
Light brown	$[47.94, 11.89, 19.86] - [71.65, 44.81, 64.78]$
White	$[100, 0, 0]$
Red	$[54.29, 80.81, 69.89]$
Blue-gray	$[50.28, -30.14, -11.96]$

Minkowski distance is considered a generalization of the Euclidean and Manhattan distances [215] and is defined as [216]:

$$d_m\left(x_i, x_j\right) = \sqrt[p]{\sum_{k=1}^{n} \left|x_{ik} - x_{jk}\right|^p} \tag{5.9}$$

where $p \geq 1$ is a real number. The distance represents the Manhattan distance and the Euclidean distance when $p = 1$ and $p = 2$, respectively.

The advantage of using Minkowski distance is that users can adapt the distance function to suit the needs of the application by modifying the Minkowski parameter p, which is set to $p = 3$ in this work. Based on the above, the image in Fig.5.9 is composed of 3 suspicious colors.

The distribution of the number of suspicious colors for the 204 images used in the experiments is depicted in Fig.5.11. It can be noticed that most of the skin lesions contained *one* suspicious color.

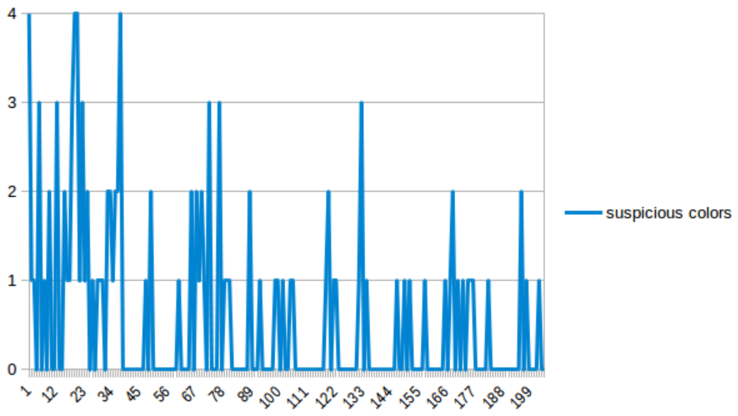


Figure 5.11: Number of suspicious colors of the 204 images labeled by the dermatologist

5.4 Diameter

In this work, Feret's diameter [218, 219] is used to measure the diameter of the skin lesion, which is the distance between two parallel tangents at the contour of the object (i.e. skin lesion) that are located on opposite sides of the object at an arbitrary selected angle. The maximum Feret diameter of an object is the distance between its two furthest points measured in a given direction. The average value over many orientations can also be used, meaning that Feret's diameter can be referred to as the average distance between two tangents in the opposite sides of the object parallel to some fixed direction; the maximum Feret diameter is used as the diameter measure.

Finding the Feret's diameter of some object shape is a commonly used measure in shape analysis. Feret's diameter is also called caliper diameter since the measurement involves placing the object for which we want to find the diameter inside the jaws of a caliper, with the caliper oriented at some specified angle (i.e. 0°, 45°, 90°, 135°). The jaws are then closed on the object tightly while maintaining the angle. The distance between the jaws is the Feret diameter at angle (direction) θ. In terms of digital images, this is made by isolating the corner pixels of the object's perimeter and taking the maximum distance between each corner pixel to all other corner pixels. It should be emphasized that Feret's diameter is based on the binary 2D image of the object. The main advantage of using Feret's diameter over other measures is its correspondence with the real physical diameter of the object. In other words, it corresponds to the length that would be measured if we handle the object between the teeth of a caliper.

As the skin lesion diameter in the real world is measured in millimeters (mm), and the diameter results using Feret's diameter are returned in pixels, representing diameter results in terms of the standard unit (mm) is required; this can be made using *spatial calibration* (geometric correction) which involves calibrating the image against a known value (i.e. mm) and then applying such calibration to the uncalibrated image (i.e. in pixels). The idea is to represent the diameter in units rather than pixels. An image produced in units is called a spatially calibrated image. However, to conduct such calibration one needs to know the original measure in real world and then map that to pixels. Since the original real world measures were not available, an image from the dataset (i.e. ISIC dataset) that had a ruler displayed (Fig.5.12) was used to get an estimate on the skin lesion measure in millimeters and deduce from that how many pixels would be in 1 mm. The image was also zoomed in to better reflect the sizes of the skin lesions used in the test dataset. Performing such calibration, it was found that *1 mm = 29.7 pixels* for the 256×256 images used in the test dataset. Fig.5.13 shows samples of skin lesion images, their segmentations, and corresponding Feret's diameter values. Fig.5.14 shows a plot of the first image in Fig.5.13 that demonstrates the Feret's diameter values at different angles.

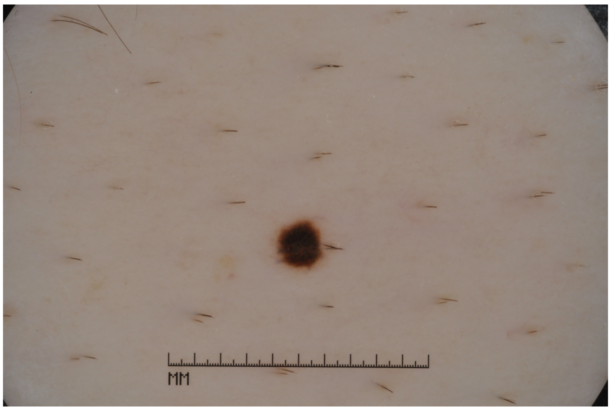
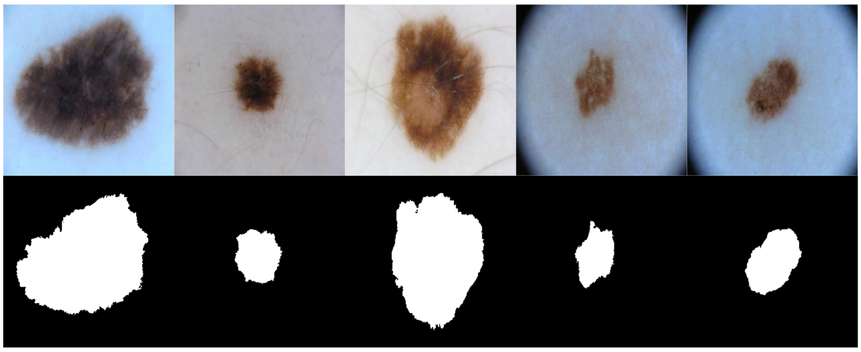


Figure 5.12: Skin lesion image with a ruler used to spatially calibrate the test images and deduce the pixels/mm value



211 pixels = 7.1 mm | 86 pixels = 2.9 mm | 205.5 pixels = 6.9 mm | 101.4 pixels = 3.4 mm | 109.6 pixels = 3.7 mm

Figure 5.13: Skin lesions' Feret's diameter (in pixels and millimeters)

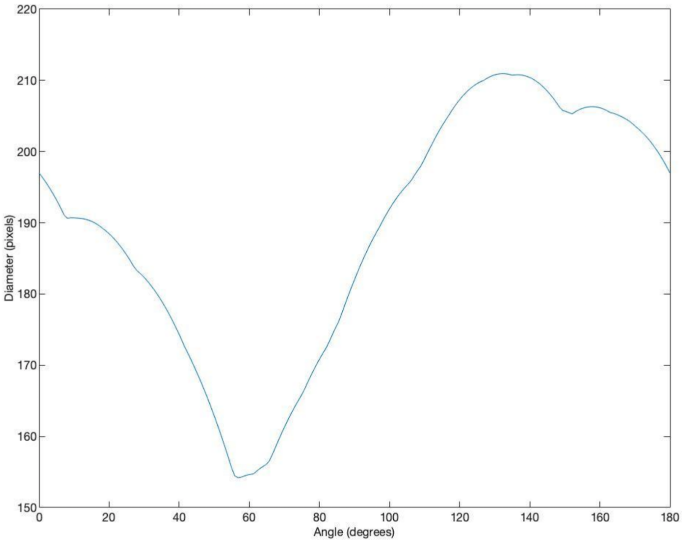


Figure 5.14: Feret's diameter values at different angles θ = [0°, 180°] for the first image in Fig.5.13

The distribution of Feret's diameter for the 204 images used in the experiments is depicted in Fig.5.15, where it can be noticed that most of the skin lesions had a diameter larger than 150 pixels (5.1 mm).

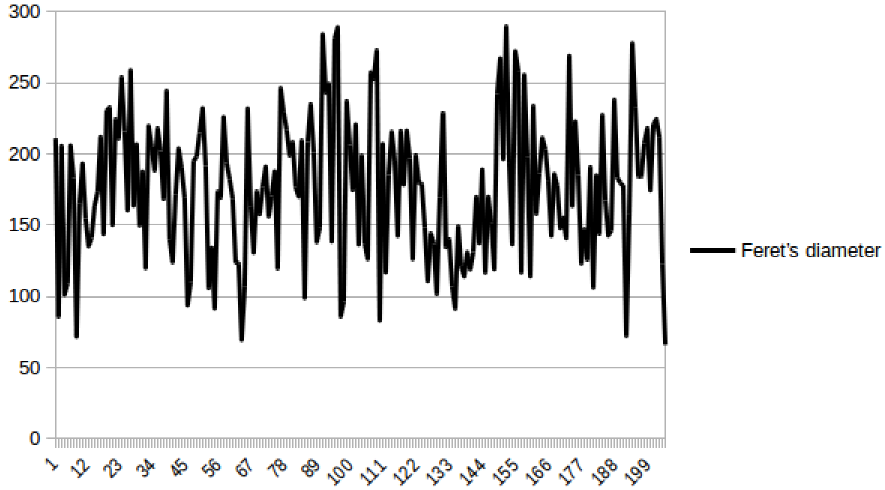


Figure 5.15: Feret's diameter values of the 204 images labeled by the dermatologist

Fig.5.16 shows a skin lesion and its extracted ACD features.

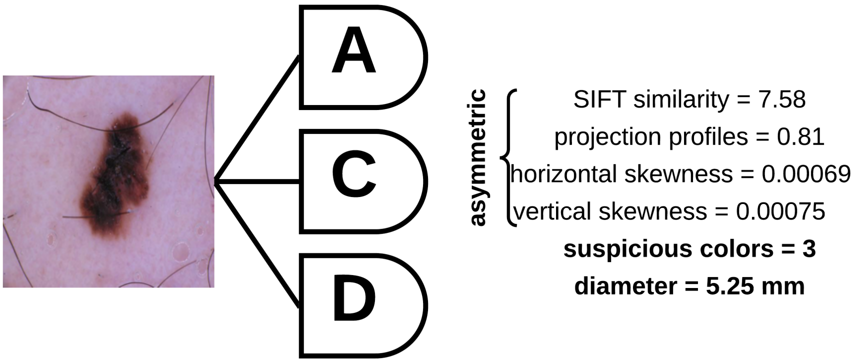


Figure 5.16: Skin lesion asymmetry, color variegation, and diameter features extracted using the work proposed in the thesis

5.5 Summary

Asymmetry, color variegation, and diameter are considered strong indicators of malignant melanoma. The subjectivity inherent in the first two features and the fact that 10% of melanomas tend to be missed in the early diagnosis due to having a diameter less than 6mm, deem it necessary to develop an objective computer vision system to evaluate these criteria. An asymmetry measure composed of a vector of SIFT based similarity, projection profiles, and skewness was proposed; a decision tree was used to learn this measure to automatically determine the presence of asymmetry in skin lesions. In finding color variegation, the skin lesion was represented in terms of its dominant colors using k-means clustering, from which the melanoma suspicious color value range that better reflect the utilized dataset was identified. Pixel colors that belong to that range are considered possessing the suspicious color. The Minkowski distance between the standard suspicious color value and pixel color value was used to determine the presence of suspicious colors that were not present in the dataset. Finally, Feret's diameter was used as a shape descriptor for measuring the lesion's diameter, and spatial calibration was used to represent the diameter in millimeters.

Chapter 6

Border Irregularity

6.1 Overview

Border irregularity has been reported to be the most significant factor in melanoma diagnosis [173]. Unlike benign pigmented lesions which tend to be present with regular borders, melanomas possess irregular borders due to the uneven growth rate [174], the spread of melanocytes in various directions, and the regression of invasion and/or genetic instability of the lesion [60]. The computerized analysis of skin cancer images normally starts by lesion border detection, which is considered a crucial step for subsequent diagnostic steps, given that clinical features such as border irregularity are measured from the border, an important feature (i.e. B feature) in the ABCD rule that physicians, dermatologists and non-physicians use to detect features of melanoma in its curable stage [27, 175]. Detecting skin lesion borders is however challenging due to the low contrast between the surrounding skin and the lesion, ambiguous lesion borders, artifacts such as light reflectance, and the variation of colors inside the lesion [176].

This chapter describes the proposed approaches involved in determining skin lesion border irregularity. Section 6.2 provides an overview of lesion border detection approaches used in literature. FuzzEdge, a proposed edge detection method, is explained in section 6.3. Section 6.4 highlights the proposed skin lesion border irregularity measure. Border irregularity detection using a proposed fuzzy multilayer perceptron (F-MLP) and a CNN-Gaussian Naive Bayes ensemble are depicted in sections 6.5 and 6.6, respectively. The chapter is concluded with a summary in section 6.7.

6.2 Skin lesion border detection

Different methods have been proposed in literature for lesion border detection. Erkol et. al. [177] used Gradient Vector Flow (GVF) snakes to obtain the border of skin lesions in dermoscopy images, where a luminance image blurring approach was presented for automatic snake initialization. The approach was tested on 100 images (30 melanoma and 70 benign). The results showed that 76 out of the 100 images had percentage border errors less than 20%, and 96 out of the 100 images had percentage border errors less than 30%. In [178] Celebi et. al. proposed a fast unsupervised approach to border detection of pigmented skin lesions based on the Statistical Region Merging (SRM) algorithm [179]. The authors attempted to facilitate the border detection process by the pre-processing operations of black frame removal and image smoothing. The process was also followed by a post-processing step, namely morphological dilation. The approach was tested on 90 dermoscopy images and did not perform well on images with significant amount of hair. Celebi et.al. [180] proposed an unsupervised approach to border detection in dermoscopy skin lesion images based on a modified version of the JSEG algorithm [181] for segmentation, preceding that with pre-processing steps that aim at facilitating the border detection procedure (image smoothing, color quantization and approximate border localization), and carrying out a post-processing step as the final stage of the proposed approach. The approach was tested on 100 dermoscopy images (30 melanoma and 70 benign) and showed to suffer from several limitations in that the bounding box determined by the approximate lesion localization method did not contain the whole lesion, and the method might not perform well on images with a significant amount of hair. An automatic approach for skin lesion border detection was proposed by Tzekis et. al. [182] where the process was carried out through a combination of different phases. The first phase attempts to determine if some points belong to melanoma or not, which is carried out by calculating the base color between the skin color and the melanoma color. The second phase of the approach locates a random point in the mole either from the entire area of the image or from random points around the center of the image. A point on the border of the mole is then found in the third phase, which adds more border points and terminates when a point is very close to where the first one is located, eventually drawing the actual boundaries of the lesion. The results of the approach were fast, simple and accurate. The authors also utilized an algorithm for hair removal. An Artifact Removal and Border Detection (ARBD) approach was proposed by Abbas [183]. The approach is composed of multiple steps: pre-processing to enhance the contrast of lesions, artifact removal to reduce the effects of specular reflection, dermoscopic-gel, and lines (e.g. hair, blood vessels, skin lines, ruler markings and camera flash), plane-fitting to reduce the effect of texture patterns from tumor regions, detecting candidate regions, and segmenting the tumor region and detecting the border using Adaptive Dynamic Programming (ADP). The approach was applied on 250 dermoscopic images (30 benign melanocytic, 60 Malignant Melanomas (MM), 45 Basal Cell Carcinoma (BCC), 25 Merkel Cell Carcinoma (MCC), 70 Seborrhoeic Keratosis (SK), and 20 Acral volar Melanocytic (AM)), and has shown to be robust to boundary detection of the tumor in cases of fuzzy or smooth lesion types (BCC, MCC and AM). Other studies attempting to detect skin lesion borders can be found in [184, 185, 186, 187], and a comprehensive survey on lesion border detection in dermoscopy images can be found in [188].

With the great advances in deep learning along with the outstanding results it provided on different computer vision tasks, some attempts have been made in utilizing this technology in skin lesion border detection. Sabouri and GholamHosseini [189] used a CNN to detect lesion borders. The CNN was trained on two classes: the lesion class (containing 480 lesion images) and the background class (containing 1200 images of the normal skin with artifacts such as hair and ruler marker). The output is a binary mask which is resized and multiplied with the original image to obtain the segmented lesion. Lesion borders using this approach were not accurately detected. The authors utilized morphological closing as a post-processing step to determine the border more accurately. The advantage of this approach is that no pre-processing operations are carried out for hair removal and illumination correction. In [190] the authors employed a CNN as a feature extractor avoiding any pre-processing operations for the input images, and a Support Vector Machine (SVM) as a skin lesion classifier. The CNN was trained on 23,000 patches of normal and skin lesions divided in such a way that each class contains the same number of images. The border of the lesion was obtained by multiplying the generated binary mask (resulting from the previous steps) with the original image. Results showed that some lesion borders were not detected perfectly, and thus a post-processing approach for smoothing the images was applied. Ali et al. [86] proposed a skin lesion border detection approach composed of two stages: (i) segmenting the skin lesion dermoscopy image using U-Net (ii) detecting the edge (border) using a novel method called *FuzzEdge*. The approach requires neither pre-processing nor post-processing operations.

6.3 FuzzEdge

A novel fuzzy filter for skin lesion border detection, namely *FuzzEdge*, was proposed in [86]. The method is based on [191]. However, in FuzzEdge the filter calculates the standard deviation for the purpose of edge detection, while in [191] it calculates the weighted fuzzy mean for the purpose of noise removal. The algorithm code has been open sourced and can be accessed at: https://github.com/abderhasan/fuzzedge.

Assume we have an image G of size $M \times N$ pixels, and has L gray levels. Such image can be denoted as $G = [g(i,j)]_{M \times N}$, where $g(i,j) \in \{0, 1, \ldots, L-1\}$ refers to a pixel in the image. Fuzzy sets that represent a particular concept for the gray level (i.e. intensity feature) of the image pixels can be constructed. Examples of such fuzzy sets are *dark pixels* and *bright pixels*. The membership function of each fuzzy set determines the membership grade in the range $[0, 1]$ (1: full membership, 0: no membership) by which a pixel with a certain gray level belongs to a fuzzy set (concept). Those fuzzy sets can be eventually used to describe image pixels.

The fuzzy concepts can be derived from the image histogram, defined as follows:

$$h(g_k) = \frac{n_k}{n} \tag{6.1}$$

where g_k denotes the k^{th} gray level of image G, n_k is the number of pixels with the k^{th} gray level in G, n is the total number of pixels in G, and $k = 0, 1, 2, \ldots, L-1$.

In order to generate the fuzzy sets (concepts), the membership functions for those fuzzy concepts have to first be defined. A heuristic algorithm is utilized to define the membership functions of the different fuzzy concepts from the histogram of the image. Three fuzzy concepts that represent an image have been created: *Bright*, *Dark*, and *Median*.

The fuzzy concepts used in this work are of the $L-R$ type fuzzy number [192] defined as follows:

$$f_{LR\_FC}(x) = \begin{cases} L\left(\frac{m-x}{\alpha}\right), & x \leq m \\ R\left(\frac{x-m}{\beta}\right), & x \geq m \end{cases}$$

(6.2)

where $L(y) = R(y) = max(0, 1-y)$, and $f(x)$ can be represented as a triplet $[m, \alpha, \beta]$, such that m corresponds to the modal value of the membership function, and α and β are the spreads that correspond to the left-hand and right-hand curves of the membership function, respectively. Algorithm 2 highlights the steps required to create fuzzy concepts.

```
input  : grayscale image I
output : fuzzy set (concept)
For the fuzzy concepts $Bright$, $Dark$, and $Median$, specify the intervals of $[Bright_{begin}, Bright_{end}]$, $[Dark_{begin}, Dark_{end}]$, and $[Median_{begin}, Median_{end}]$, respectively.
Let $Bright_{begin} = \left(N_f - 1\right)\left[\frac{L-1}{N_f}\right]$, $Dark_{end} = \left[\frac{L-1}{N_f}\right]$, $Median_{begin} = Dark_{end} - left\_overlap$, and $Median_{end} = Bright_{begin}\ right\_overlap$ /* $N_f$ is the number of fuzzy concepts, and $left\_overlap$ and $right\_overlap$ determine the overlapping range of the fuzzy concepts (the overlap range was set to 0) */.
Set $Dark_{begin}$ to be the first $g_k$ from 0 to $Dark_{end}$ with $n_k \neq 0$ /* $k = 0, 1, 2, ..., L - 1$) */.
Set $Bright_{end}$ to be the last $g_k$ from $Bright_{begin}$ to $L - 1$.
In the interval $[Dark_{begin}, Dark_{end}]$, find $g_k$ with the maximum value of $p(g_k)$ in the image /* the most frequent $g_k$ is selected */.
For the fuzzy concept $Dark$, create its membership function $f_{Dark}$ as follows: $m_{Dark} \leftarrow g_k$, $\alpha_{Dark} \leftarrow m_{Dark} - Dark_{begin}$, $\beta_{Dark} \leftarrow Dark_{end} - m_{Dark}$.
In the interval $[Median_{begin}, Median_{end}]$, find $g_k$ with the maximum value of $p(g_k)$.
For the fuzzy concept $Median$, create its membership function $f_{Median}$ as follows: $m_{Median} \leftarrow g_k$, $\alpha_{Median} \leftarrow m_{Median} - Median_{begin}$, $\beta_{Median} \leftarrow Median_{end} - m_{Median}$.
In the interval $[Bright_{begin}, Bright_{end}]$, find $g_k$ with the maximum value of $p(g_k)$.
For the fuzzy concept $Bright$, create its membership function $f_{Bright}$ as follows: $m_{Bright} \leftarrow g_k$, $\alpha_{Bright} \leftarrow m_{Bright} - Bright_{begin}$, $\beta_{Bright} \leftarrow Bright_{end} - m_{Bright}$.
```

Algorithm 2: Fuzzy set (concept) creation process

For initiating the filtering process, a 3×3 window (kernel) is affected on the input image, where the window determines the gray level values (intensities) of the filtered area, and the pixel to be filtered would stand in the central cell of the 3×3 kernel. Let $X = [x\,(i,j)]_{M \times N}$ be the original input image, and $Y = [y\,(i,j)]_{M \times N}$ be the filtered output image. The $(i,j)^{th}$ pixel of the filtered image Y is represented as: $\bar{y}(i,j) = FuzzEdge(\underline{X}(i,j))$, where $\underline{X}(i,j)$ is a 3×3 kernel centered at the input pixel $x(i,j)$ that will be affected by the filter, and $FuzzEdge(.)$ denotes the function of the $FuzzEdge$ (fuzzy filter). The kernel can be represented as follows:

$$\underline{X}(i,j) = \begin{bmatrix} x(i-1,j-1) & x(i-1,j) & x(i-1,j\ 1) \\ x(i,j-1) & x(i.j) & x(i,j\ 1) \\ x(i\ 1,j-1) & x(i\ 1,j) & x(i\ 1,j\ 1) \end{bmatrix}$$

(6.3)

The $FuzzEdge$ operation is composed of three standard deviation processes, one for each fuzzy concept (*Bright*, *Dark*, or *Median*). Each standard deviation process determines the value of the filtered pixel on one of the three fuzzy concepts. The weight associated with each pixel is determined by referring to the membership function (i.e. f_{Bright}) of the associated fuzzy concept. The standard deviation is then found for the pixels located in the kernel, eventually affecting the result to the pixel located at the center of the kernel (i.e. $\bar{y}(i,j)$). This process allows us to find $y_{\bar{Bright}}(i,j)$, $y_{\bar{Dark}}(i,j)$, and $y_{\bar{Median}}(i,j)$.

After $y_{\bar{Bright}}(i,j)$, $y_{\bar{Dark}}(i,j)$ and $y_{\bar{Median}}(i,j)$ are produced, the *decision process* of $FuzzEdge$ is utilized in order to determine the final filtered output of each pixel in the input image by referring to a fuzzy estimator derived from a fuzzy interval. A fuzzy interval is of $LR-type$ if two shape functions L and R exist, in addition to the parameters (i.e. $(m_l, m_R) \in R^2$, α and β) that are used to form the membership function of the fuzzy interval. The fuzzy interval can be denoted as $FI = [m_l, m_r, \alpha, \beta]_{LR}$, and the membership function of FI can be defined as shown in Equation 5.4. A standard deviation process similar to the one described above is applied on a sliding kernel (window) centered at the input pixel $x(i,j)$. In the fuzzy estimator step, the parameters are assigned the following values: $m_r = Bright_{end}$, $m_l = Dark_{begin}$, $L(y) = 0$ and $R(y) = 0$ for the membership function of the fuzzy interval.

$$f_{LR\_FI}(x) = \begin{cases} L\left(\dfrac{m_l - x}{\alpha}\right), & x \leq m_l \\ 1, & m_l \leq x \leq m_r \\ R\left(\dfrac{x - m_r}{\beta}\right), & x \geq m_r \end{cases}$$

(6.4)

The final output of each filtered pixel is determined by selecting the pixel that is nearest to the fuzzy estimator from $y_{\bar{Bright}}(i,j)$, $y_{\bar{Dark}}(i,j)$ and $y_{\bar{Median}}(i,j)$ (Algorithm 3).

```
if $\left|\bar{y}_{Dark}(i,j) - f_{LR-FI}\left(\underline{X}(i,j)\right)\right| < \left|\bar{y}_{Median}(i,j) - f_{LR-FI}\left(\underline{X}(i,j)\right)\right|$
    $y(i,j) \leftarrow \bar{y}_{Dark}(i,j)$
else
    $y(i,j) \leftarrow \bar{y}_{Median}(i,j)$
if $\left|\bar{y}_{Bright}(i,j) - f_{LR-FI}\left(\underline{X}(i,j)\right)\right| < \left|y(i,j) - f_{LR-FI}\left(\underline{X}(i,j)\right)\right|$
    $y(i,j) \leftarrow \bar{y}_{Bright}(i,j)$
```

Algorithm 3: Decision process of $FuzzEdge$

The general framework of the proposed approach is depicted in Fig.6.1.

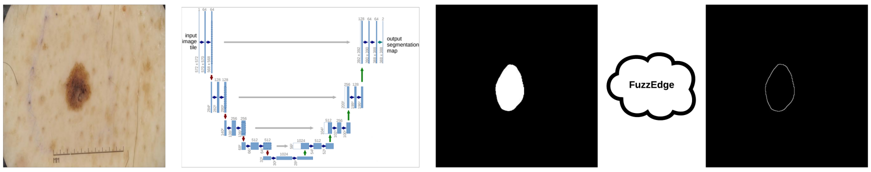


Figure 6.1: Framework of the approach proposed in [86]

The U-Net architecture was trained on 2594 dermoscopy images along with their corresponding ground truth response masks from the "ISIC 2018: Skin Lesion Analysis Towards Melanoma Detection" grand challenge datasets [18-19] (the ISIC archive contains more than 23,000 dermoscopic images). Fig.6.2 shows some examples on the training dataset and the corresponding ground truth.

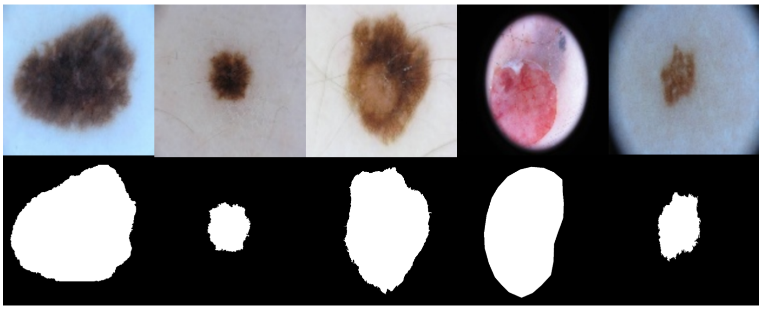


Figure 6.2: Samples of images and their corresponding ground truth used to train U-Net

To make the most of the training data, augmentation using some transformations has been applied (rotation, shifting horizontally and vertically, shearing, zooming, horizontal flip, and filling in newly created pixels which can appear after a rotation or a horizontal/vertical shift).

This prevents over-fitting and leads to better generalization. Images used have been resized to 512×512 pixels. The U-Net model was trained for 20 epochs on a Tesla P100 GPU. Fig.6.3 shows how accuracy has improved across the different epochs.

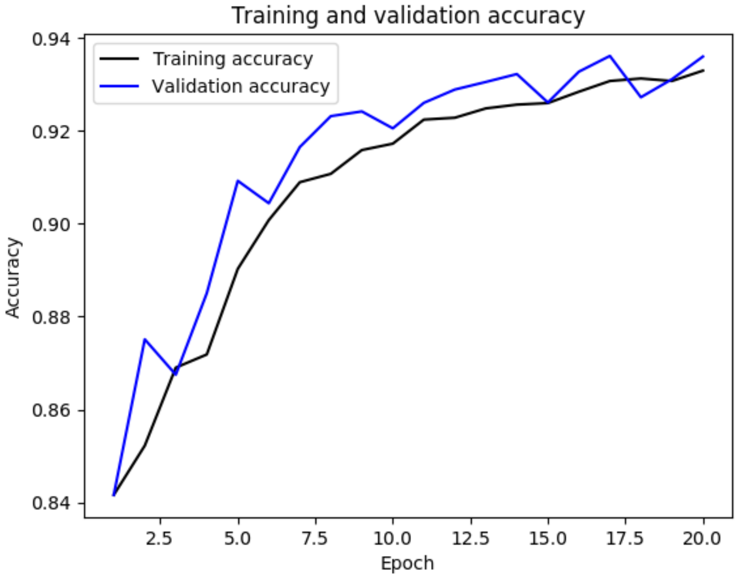


Figure 6.3: Accuracy improvement across different epochs

6.3.1 Ground truth

A ground truth is required to evaluate the automatically detected borders; the manual borders which represent the ground truth were created by a dermatologist. Fig.6.4 shows an example on such annotated images which represent a sample of skin lesion dermoscopy test images that the model was not trained on before (i.e. did not see).

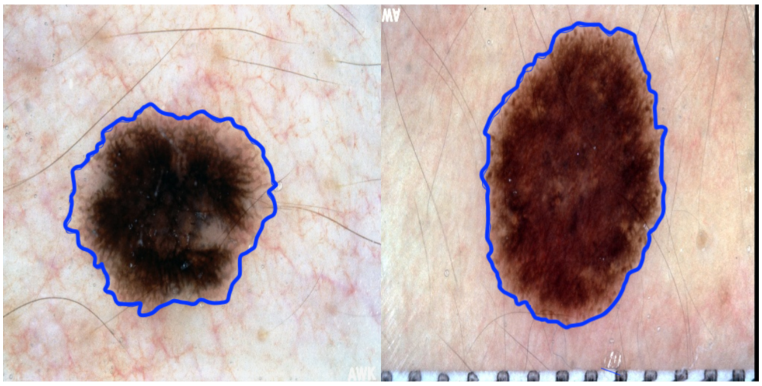


Figure 6.4: Borders (in blue) manually drawn by the dermatologist

6.3.2 Comparison with another automated method

Fig.6.5 shows the test images, segmentation results (i.e. U-Net), and detected borders (i.e. FuzzEdge).

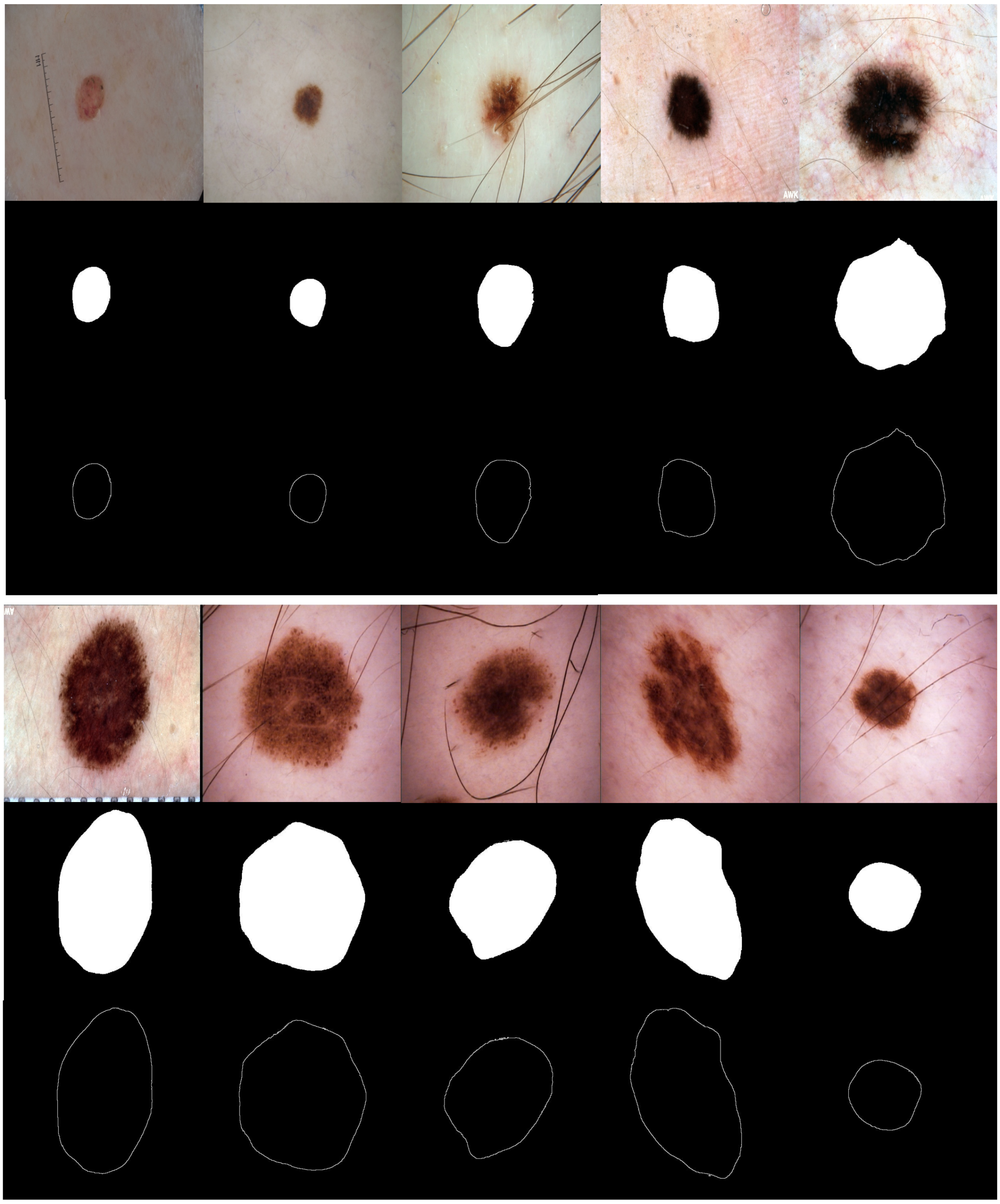


Figure 6.5: Results of the proposed approach. The first row in the two sets of images represents the original dermoscopy images, the second row shows the segmentation results of U-Net, and the third row displays the detected borders after applying FuzzEdge

The proposed approach was compared with another automatic border detection approach developed by Hua [193]. Fig.6.6 illustrates Hua's two stage results: (i) pre-processing and (ii) edge detection using the Sobel edge detector [194] (the author used different edge detectors, but one edge detector was chosen for brevity, namely the Sobel edge detector). Results of the pre-processing stage had to be binarized for better edge detection results.

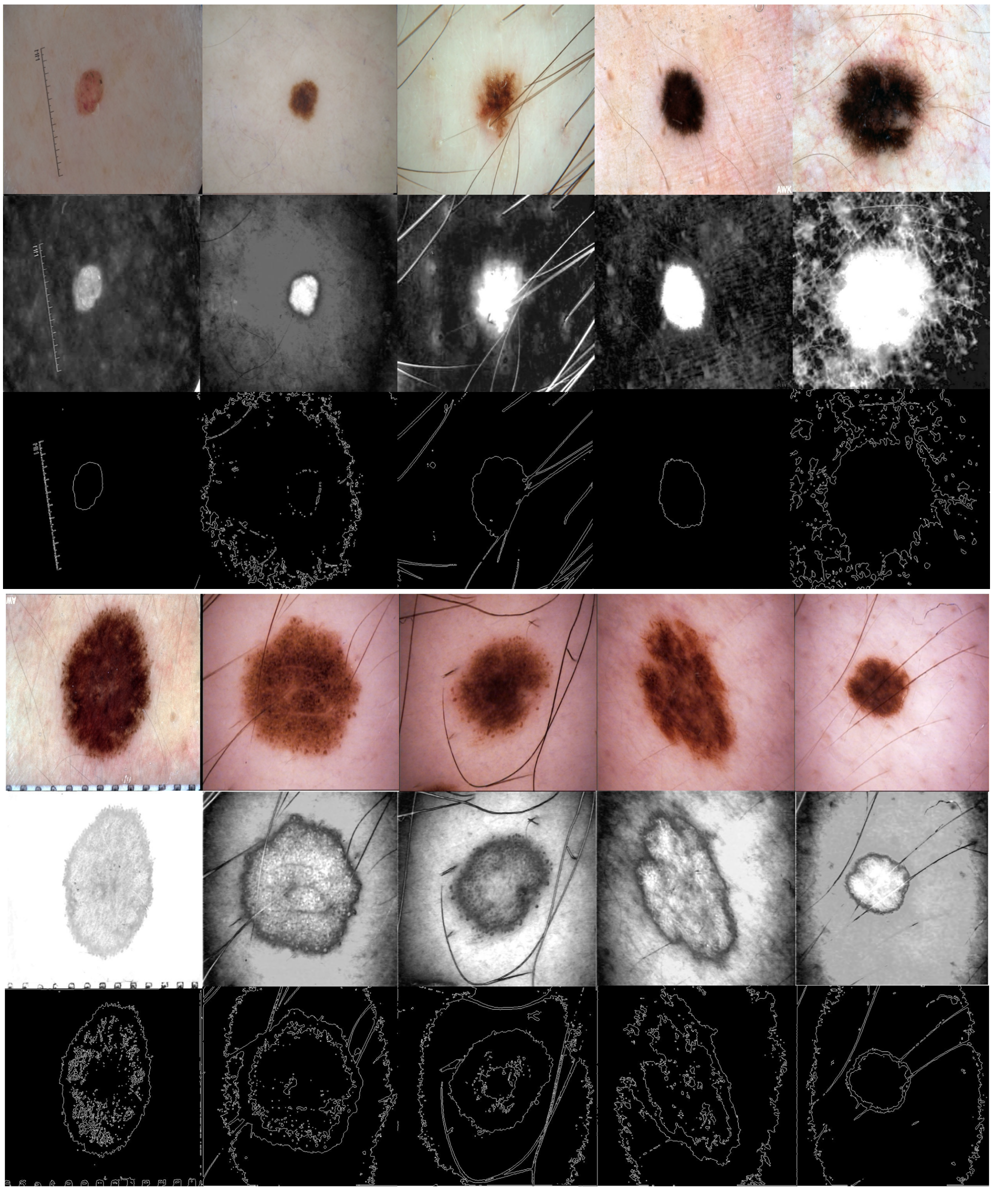


Figure 6.6: Results of the approach proposed by Hua [193]. The first row in the two sets of images represents the original dermoscopy images, the second row shows the pre-processing results, and the third row displays the detected borders after applying the Sobel edge detector on the binarized results

From Fig.6.5 and Fig.6.6 we can notice that the proposed approach is able to detect the main border around the skin lesion, and is robust to artifacts presented in the images as opposed to the approach proposed by Hua [193]. Moreover, the approach by Hua is heuristic and based on trial and error as different threshold and sigma values need to be experimented with. However, the proposed approach lacks the ability to detect the exact fine structure of the skin lesion border. This might be due to the many structures skin lesions possess, unlike other medical imaging shapes (i.e. colon) that tend to be very similar,

6.3.3 Comparison with the ground truth

Fig.6.7 shows a sample of the proposed approach results (white border) overlaid against the dermatologist's annotations (blue border). A visual comparison between the borders drawn by the dermatologist and the borders of the proposed approach shows that the automatically extracted borders are very close to those manually outlined by the dermatologist.

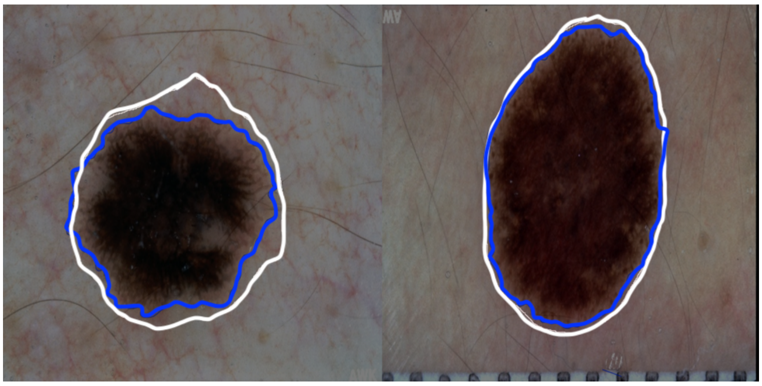


Figure 6.7: Borders of the proposed approach (in white) overlaid on the borders annotated by the dermatologist (in blue)

To quantitatively evaluate the proposed method against the ground truth, the areas surrounded by the dermatologist's border and the proposed method's border are filled for each image. The Dice coefficient [161] is used to measure the similarity between the two images. The Dice coefficient is a measure that finds the spatial overlap between two binary images, resulting in a value that lies between 0 (no overlap) and 1 (agrees perfectly). The Dice coefficient can be defined as follows:

$$D = \frac{2\,(A \cap G)}{A\ G} \times 100\% \tag{6.5}$$

where A is the algorithm output and G is the ground truth.

Table.6.1 lists the Dice coefficient results of 10 test images (the names reflect the original names as used in the dataset). The average Dice similarity achieved by the proposed approach is 87.7%. The reason the Dice coefficient results of Hua's approach are not included is since it was hard to determine the region to fill out for ground truth comparison purposes due to the existence of much noise as depicted in Fig.6.6, as opposed to the proposed approach which shows only one region of interest (border).

Table 6.1: Dice values for dermoscopy test images

Image	Dice
ISIC_0012456	82.6%
ISIC_0012560	89%
ISIC_0012647	81.1%
ISIC_0015094	84.7%
ISIC_0015117	88.2%
ISIC_0015165	95.4%
ISIC_0017789	74.8%
ISIC_0018917	94.4%
ISIC_0019133	94.5%
ISIC_0020005	92.1%

6.4 Skin lesion border irregularity

After detecting the skin lesion border we need to measure the border's irregularity which represents the B feature of the ABCD rule. For this task, fractal dimension is combined with both Zernike moments and convexity that would together serve as an *objective* quantitative measure of border irregularity, especially when many of the signs that the clinician relies on in diagnosis involve *subjective* judgment. This applies to visual signs such as border irregularity [195]. It has been shown that both clinicians and patients find it hard in agreeing upon whether a naevus border is considered irregular or not [195]. Such measure could thus aid in improving the diagnostic accuracy.

Although the fractal dimension D provides values consistent with the rules normally used in clinical practice in the aspect that D values significantly increase in melanoma lesions as compared to benign lesions, using D as a single parameter in distinguishing skin lesion border irregularity could be limited, especially that it relies on the quality of the edge detection algorithm used. Thus, combining it with other parameters should be considered [196].

Each skin lesion is now represented by a 27-value vector as depicted in Fig.6.8, reflecting the proposed lesion border irregularity measure.

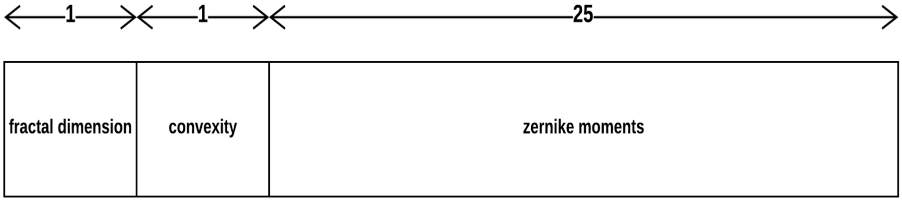


Figure 6.8: Skin lesion border irregularity measure

The feature vector combination shown in the figure produces a robust border irregularity measure. Since convexity is an Euclidean based measure that characterizes circle and sphere-like shapes, it is more suitable for benign skin lesions that tend to have regular borders and possess Euclidean geometrical shapes. However, melanomas have irregular borders, and this thus required the introduction of fractal dimension in the feature vector since it gives an indication on the complexity of the shape associated with the border, in addition to being better at describing irregularly shaped objects that have fractal properties. Zernike moments have been used due to the fact that the accuracy of fractal dimension relies on the quality of the edge detection algorithm and the edge thickness [237], while Zernike moments do not require knowledge of the precise boundary of the skin lesion as they are not sensitive to image segmentation results, making them suitable for representing complex objects with obscure boundaries [238]. Moreover, Zernike moments are able the characterize the global shape of the object, robust to noise, and rotation invariant (an important feature when augmenting skin lesion images) [239].

6.5 Border irregularity detection using F-MLP

A fuzzy multilayer perceptron was proposed in [87], which incorporates the membership degree of each neuron in the classes of interest (e.g. regular vs. irregular) in the learning process. Moreover, the gradient descent benefits from the membership values by reducing the effects of *ambiguous* features/neurons (i.e. features that have a membership degree of 0.5) when updating the weights (learning). Membership degrees are obtained by clustering each layer in the neural network (except the output layer) using fuzzy c-means. The proposed architecture is depicted in Fig.6.9.

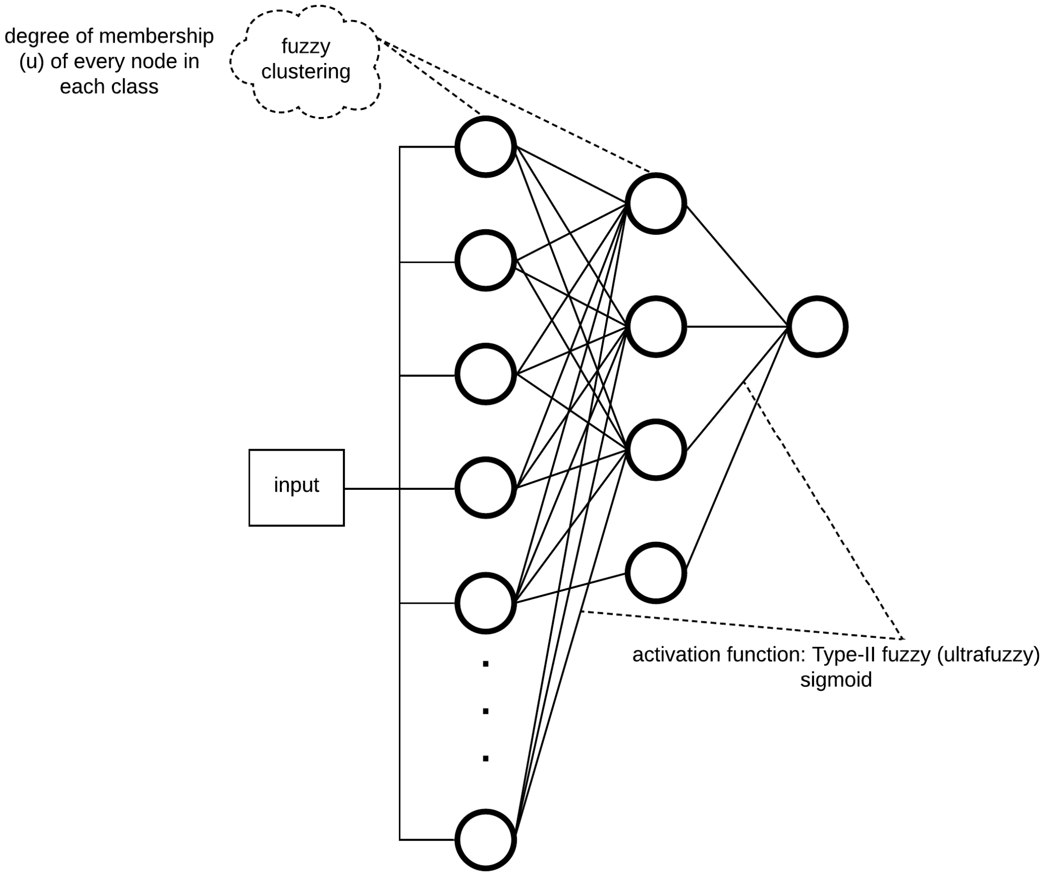


Figure 6.9: Fuzzy multilayer perceptron (F-MLP) architecture

A commonly used activation function in multilayer perceptrons is the *sigmoid* activation function (Eq.6.6). The sigmoid function is suitable for binary classification and provides continuous values in the range $[0,1]$ that represent the probability of a class in the binary classification problem. As the sigmoid function introduces non-linearity in the hidden layers, it allows the neural network to learn more complex features [206].

$$sig\left(x\right) = \frac{1}{1\ e^{-x}} \tag{6.6}$$

Assuming that φ is the fuzzy sigmoid activation function, the *type-II fuzzy sigmoid* activation function can be represented as:

$$\varphi_L\left(x\right) = \left[\frac{1}{1\ e^{-x}}\right]^{\alpha} \tag{6.7}$$

$$\varphi_U\left(x\right) = \left[\frac{1}{1\ e^{-x}}\right]^{\frac{1}{\alpha}} \tag{6.8}$$

where φ_L and φ_U are the lower and upper sigmoid activation functions, respectively, and $\alpha = 2$ since $\alpha >> 2$ is not meaningful for image data [141].

The proposed *fuzzy gradient descent* is defined as follows:

$$\mathbf{w} = \mathbf{w} - mean\left(|\mathbf{u}_1 - \mathbf{u}_2|^2\right) \times \eta \times \frac{d}{d\mathbf{w}}\varphi \tag{6.9}$$

where $\mathbf{w}$ are the weight values, $\mathbf{u}_1$ and $\mathbf{u}_2$ are the degrees of membership of each neuron to *class1* and *class2*, respectively; φ is the type-II fuzzy sigmoid function, and *mean* is used to represent the square differences between the degrees of membership for a particular layer with a single value, which can be perceived as an *ambiguity* parameter. Notice that for ambiguous nodes (i.e. $u_1 = u_2 = 0.5$) $|u_1 - u_2|^2$ will evaluate to 0, thus having no effect on how weights are being updated. Incorporating degrees of membership in optimization will determine how features/neurons contribute to the learning process based on their ambiguity, such that more ambiguous features/neurons will have less effect on learning, and will rather be based on more non-ambiguous features/neurons. The cost function used in this work is simply represented as the difference between the actual values and the predicted values. The F-MLP algorithm code has been open sourced and can be accessed via https://github.com/abderhasan/F-MLP.

The skin lesion images were first segmented using U-Net, which was trained for 20 epochs on a Tesla P100 GPU on 1777 dermoscopy images - resized to 256 × 256 pixels - along with their corresponding ground truth response masks from the “ISIC 2018: Skin Lesion Analysis Towards Melanoma Detection" grand challenge datasets [82, 158], and tested on 158 images from the same dataset. Training U-Net and testing it took 27.6 minutes and 25.9 seconds, respectively. Fig.6.10 shows samples of the training dataset along with their ground truth, and Fig.6.11 shows samples of the segmentation results using U-Net (i.e. test dataset). The average Dice similarity achieved on the 158 images was 83.8%.

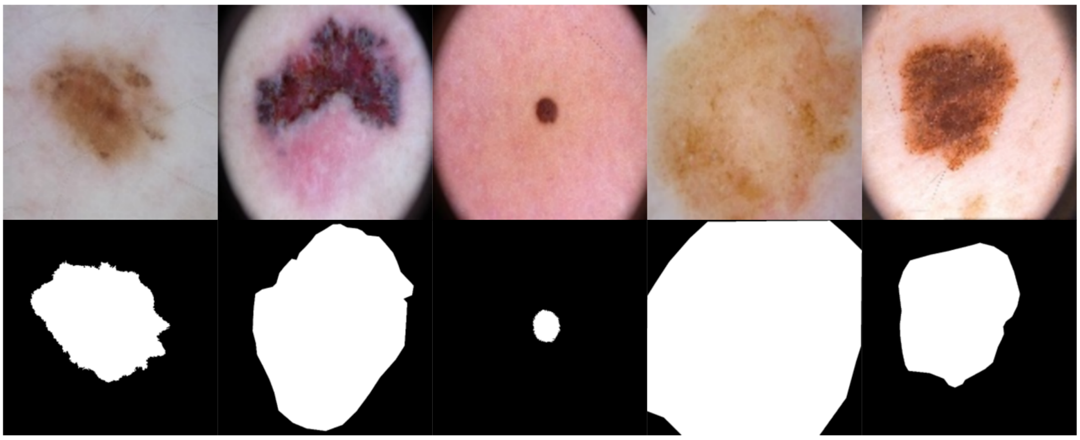


Figure 6.10: Samples of images used to train U-Net along with their ground truth

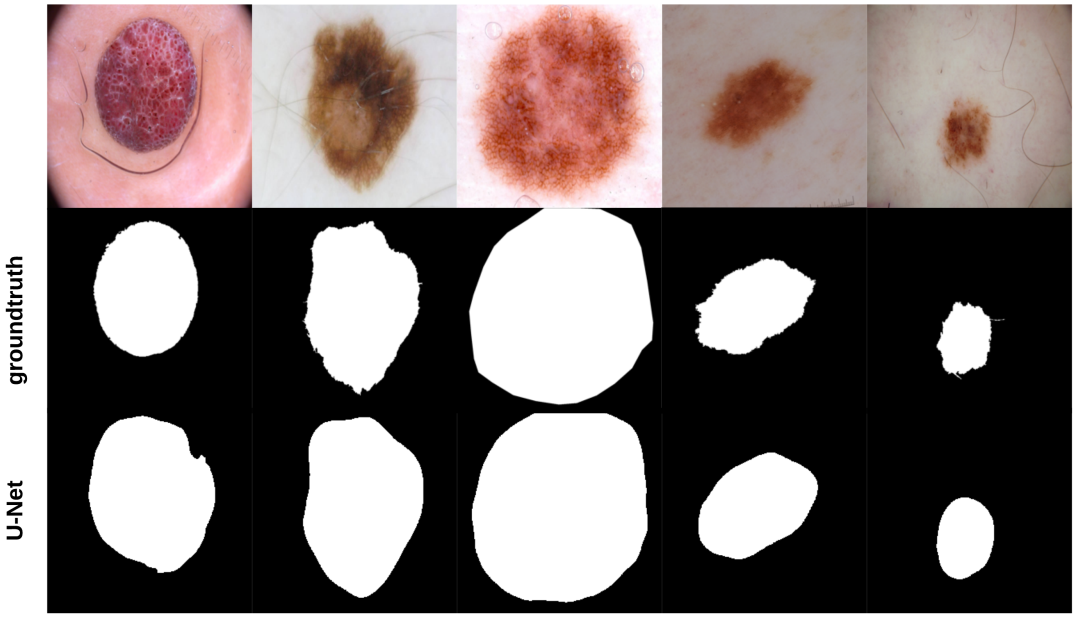


Figure 6.11: Samples of images used to test U-Net, their ground truth, and segmentation results

To detect the skin lesion border, FuzzEdge (section 6.3) was used. Fig.6.12 shows some samples of borders detected using FuzzEdge. Skin lesion border irregularity was measured based on the method proposed in section 6.4, excluding Zernike moments.

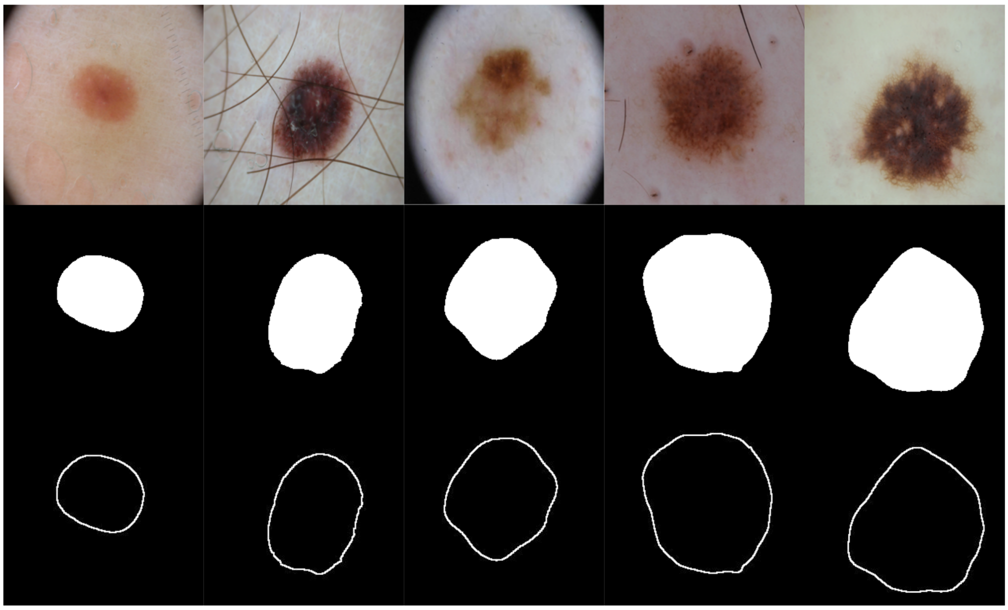


Figure 6.12: Samples of test images, their segmentation using U-Net, and borders detected using FuzzEdge

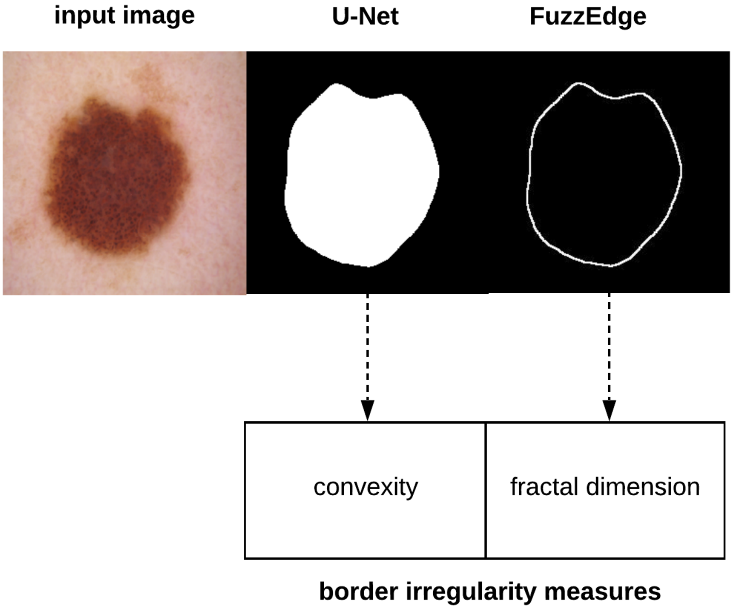


Figure 6.13: Skin lesion border irregularity measures extraction

To prepare the training and testing data for F-MLP, 158 images were segmented using U-Net, their skin lesion borders extracted using FuzzEdge, and the extracted borders sent to a dermatologist to label as *regular* or *irregular* borders (regular: 5, irregular: 153), which will eventually serve as the ground truth (labels) for the training data. Fig.6.14 shows some samples of regular and irregular borders along with their original and segmented images. However, due to the imbalance in data, an *augmentation* step (rotating, and flipping horizontally and vertically) has been carried out to increase the regular samples. Augmentation was carried on the 5 regular bordered images, producing multiple versions of those imgaes. The total number of images after augmentation was 310 images (regular: 157, irregular: 153).

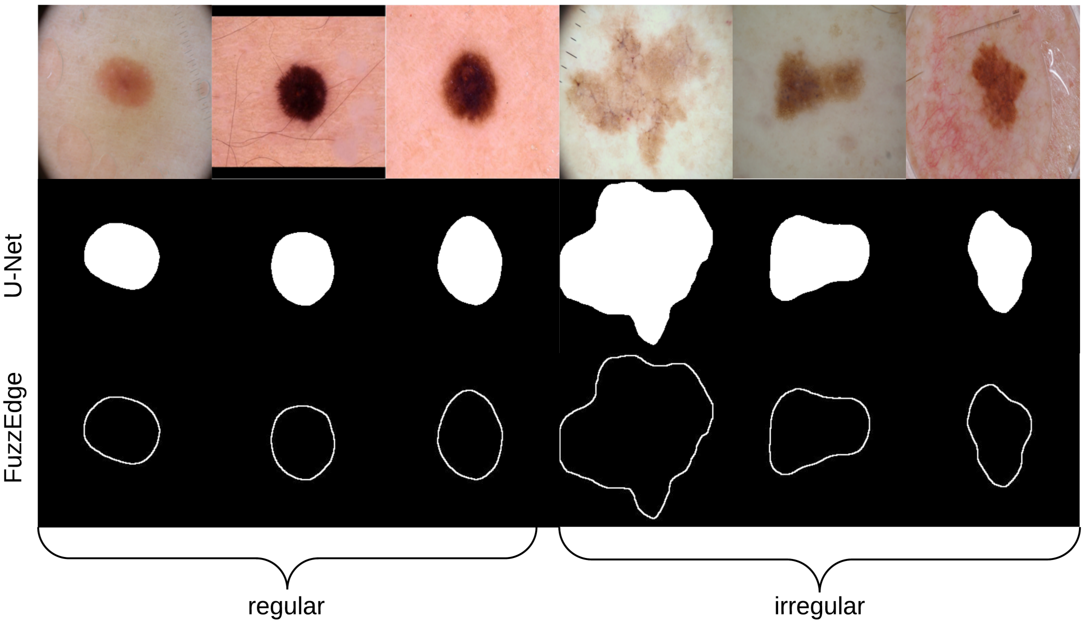


Figure 6.14: Samples of regular and irregular borders labeled by the dermatoloigist

Table.6.2 shows the fractal dimension and convexity values for the images in Fig.6.14. It should be emphasized that fractal dimension is found for the edge images (i.e. FuzzEdge) and convexity is found for the segmentation (i.e. U-Net) results of the image, as demonstrated in Fig.6.13. Fig.6.15 shows box-and-whisker plots depicting the distribution of fractal dimension and convexity values for the regular and irregular skin lesions used in training and testing the neural networks.

Table 6.2: Border irregularity measures for the images presented in Fig.6.14. Images 1.r, 2.r, and 3.r from left to right refer to the first three images (regular), and images 1.i, 2.i, and 3.i refer to the last three images (irregular)

Image	Fractal Dimension	Convexity	Label
1.r	1.2527	0.9898	1
2.r	1.2599	0.9890	1
3.r	1.2875	0.9893	1
1.i	1.4499	0.9031	0
2.i	1.3056	0.9531	0
3.i	1.3125	0.9586	0

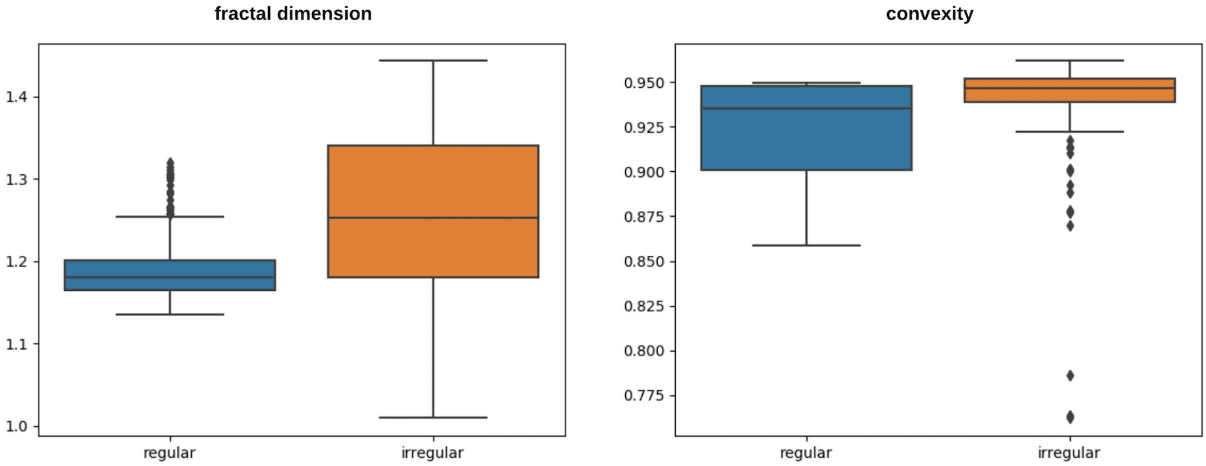


Figure 6.15: Box-and-whisker plots representing the fractal dimension and convexity distributions of the skin lesions (regular and irregular) used in training and testing the neural networks

The extracted skin lesion border irregularity measures were used to train and test a standard neural network and a type-II F-MLP. For both networks the number of neurons in the input layer is equal to the number of input features (2 features), the first hidden layer is composed of 4 neurons, the second hidden layer is composed of 2 neurons, the output layer is composed of 1 neuron which represents the final classification result, and the learning rate is 0.001. Experiments were run on a machine with an Intel Core i7 processor of speed 2.2 GHz and 16 GB memory.

After obtaining the *prediction probability* $\in [0, 1]$ of each test sample, a *threshold* is generated from the prediction probabilities to decide the final prediction (regular or irregular) according to Eq.6.10.

$$threshold = \frac{\sum_{i=1}^{n} p_i}{n} \tag{6.10}$$

where n is the number of test samples, and p_i is the prediction probability of test sample i. The final decision is obtained using Eq.6.11.

$$Decision = \begin{cases} regular, & P_i > threshold) \\ irregular, & P_i \leq threshold) \end{cases} \tag{6.11}$$

Table.6.3 and Table.6.4 show the different training and testing split ratios used to evaluate the networks, number of iterations used in each network, time consumed (in seconds), and accuracy for both the standard neural network and F-MLP, respectively. The networks were run for only 1 iteration since more iterations didn't improve the accuracy.

Table 6.3: Standard neural network evaluation on classifying regular and irregular borders using different training and testing split ratios

Ratio	Training	Testing	Training time	Testing time	Accuracy
80:20	248	62	0.02	0.007	91.9%
70:30	217	93	0.02	0.008	91.4%
60:40	186	124	0.01	0.007	87.9%
50:50	155	155	0.02	0.009	79.4%

Table 6.4: F-MLP evaluation on classifying regular and irregular borders using different training and testing split ratios

Ratio	Training	Testing	Training time	Testing time	Lower sigmoid acc.	Upper sigmoid acc.
80:20	248	62	0.58	0.01	95.2%	90.3%
70:30	217	93	0.8	0.07	91.4%	89.2%
60:40	186	124	0.6	0.08	90.3%	87.9%
50:50	155	155	0.7	0.08	83.9%	75.5%

Using two sigmoid activation functions reflects the type-II fuzzy set in that the error rates represent the range of performance that could be achieved using the fuzzy neural network (F-MLP), modeling thereby the potential uncertainty occurring within the input data. Two versions (lower and upper sigmoid) of F-MLP can be obtained, and the one with the best performance (maximum accuracy) can be used as shown in the following equation:

$$\lambda FMLP = \max\left(\lambda FMLP_{lower}, \lambda FMLP_{upper}\right) \tag{6.12}$$

where $\lambda FMLP$ is the accuracy of the fuzzy multilayer perceptron, $\lambda FMLP_{lower}$ is the accuracy of the fuzzy multilayer perceptron utilizing the lower sigmoid activation function, and $\lambda FMLP_{upper}$ is the accuracy of the fuzzy multilayer perceptron utilizing the upper sigmoid activation function.

In comparing the standard neural network and F-MLP, the 80:20 ratio is considered as it results in better accuracy amongst the other ratios, evaluating to 91.9% and 95.2% for the standard neural network and F-MLP, respectively. Table.6.5 and Table.6.6 depict the confusion matrices of the classification results of both networks, from which we derive the sensitivity and specificity values that evaluate to 100% and 82.8% for the standard neural network, respectively, and 100% and 89.7% for the F-MLP, respectively. Fig.6.16 depicts the receiver operating characteristic (ROC) curves of the standard neural network and F-MLP. To evaluate the proposed approach further, it is compared with other state-of-the-art classification methods as shown in Table.6.7 which shows that F-MLP outperforms most of those methods.

Table 6.5: Standard neural network confusion matrix

	Predicted		
	Regular	Irregular	Total
Regular	33	0	33
Irregular	5	24	29
Total	38	24	62

Table 6.6: F-MLP (lower sigmoid) confusion matrix

	Predicted		
	Regular	Irregular	Total
Regular	33	0	33
Irregular	3	26	29
Total	36	26	62

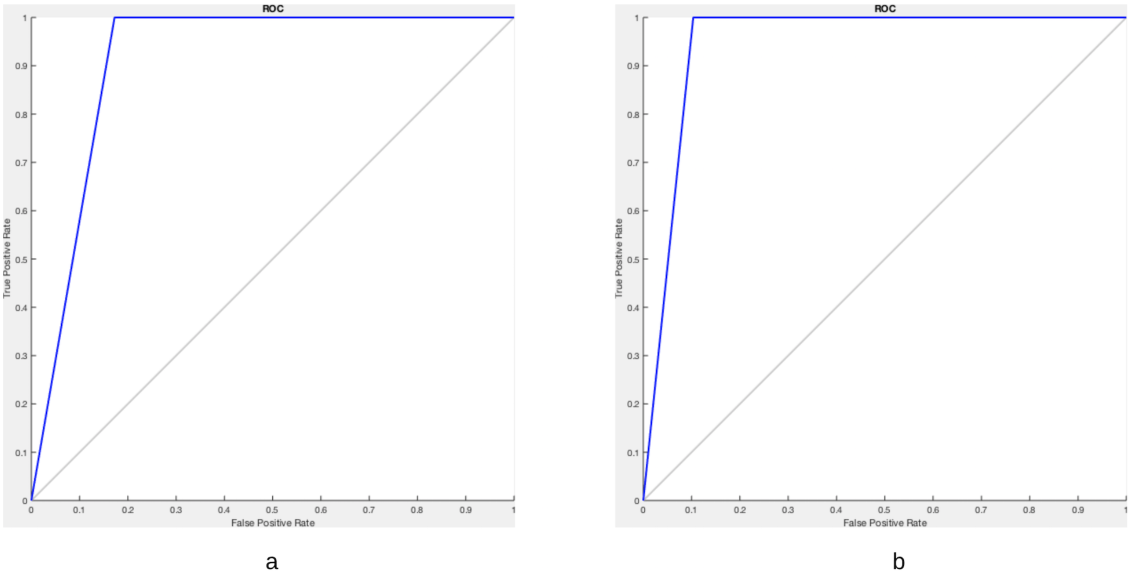


Figure 6.16: ROC curve (a) standard neural network (b) F-MLP (lower sigmoid)

Table 6.7: Comparing F-MLP (lower sigmoid) with other state-of-the-art classification methods

Method	TP	TN	FP	FN	Accuracy
F-MLP (lower sigmoid)	33	26	3	0	95.2%
Random forests	32	28	1	1	96.8%
Stochastic gradient descent	33	26	3	0	95.2%
Random forests	32	28	1	1	96.8%
Logistic regression	32	22	7	1	87.1%
K-nearest neighbors	32	26	3	1	93.5%
Gaussian naive bayes	32	22	7	1	87.1%
Support vector machine	28	23	11	0	87.1%
Decision tree	32	26	3	1	93.5%

Incorporating the membership degree in the gradient descent (Eq.6.9) helps in reducing the effects of ambiguous features/neurons when updating the weights, and thus increases the performance of learning (i.e. higher accuracy predictions). The proposed type-II F-MLP is able to perform better than its traditional neural network counterpart with fewer iterations. However, training F-MLP is more time-consuming than its traditional neural network counterpart.

To test the generalizability of F-MLP, it was compared with the standard neural network on some standard datasets, namely, the *Titanic* dataset (from the Department of Biostatistics at the Vanderbilt University School of Medicine [1]) and the *Pima Indians diabetes* dataset (from the National Institute of Diabetes and Digestive and Kidney Diseases [2]). Training and testing data was split on an 80:20 ratio basis. The classification problem in the Titanic dataset involved predicting whether a person survived the tragedy or not. The dataset contained the information of 1309 passengers; the following features of each passenger were used: survival (whether the passenger survived or not) which acts as the label of the input, gender, age, number of siblings and spouses aboard, number of parents and children aboard, class (passenger class: first, second, or third), the port of embarkation (C = Cherbourg; Q = Queenstown; S = Southampton), and the passenger fare. Two layers have been used in the networks (traditional and fuzzy neural

---

[1] http://biostat.mc.vanderbilt.edu/wiki/pub/Main/DataSets/titanic3.csv

[2] https://www.kaggle.com/uciml/pima-indians-diabetes-database

networks). The input layer was composed of 7 neurons (i.e., the number of input features) and the hidden layer was composed of 1 neuron. The learning rate was 0.1 and the networks were run for 10 iterations on a machine with an Intel Core i7 processor of speed 2.2 GHz and 16 GB memory. The accuracy obtained using the traditional neural network was 39.3%, while the type-II fuzzy neural network evaluated to 79% and 29% for the lower and upper sigmoid activation functions, respectively. The time elapsed for training and testing the traditional neural network was 0.07 seconds and 0.08 seconds, respectively; and took 58.5 seconds and 0.19 seconds to train and test the type-II fuzzy neural network, respectively. Increasing the number of neurons in the hidden layer of the traditional neural network to 6 improved its accuracy to 63.7%. However, it is not necessary that increasing the number of neurons in the hidden layer will always lead to better accuracy. For instance, increasing the number of neurons to 10 lead to a 59.5% accuracy. The Pima Indians diabetes dataset is composed of predictor variables for diabetes, namely: number of pregnancies the patient had before, plasma glucose concentration, diastolic blood pressure (mm Hg), triceps skin fold thickness (mm), 2-hour serum insulin (mu U/ml), body mass index (BMI), diabetes pedigree function, and age. The outcome variable refers to the diagnosis of the patient and takes the values 0 or 1. The classification problem stems around the prediction of the onset of diabetes based on the diagnostic measures provided (i.e. predictor variables); 768 samples were present in the dataset. The structure, hyperparameters (i.e. learning rate and number of iterations) of the neural network, and the machine settings resemble those used with the Titanic dataset, but the number of input neurons is 8 which reflects the number of features used in the Pima Indians diabetes dataset. The accuracy obtained using the traditional neural network was 36.1%, while for the type-II fuzzy neural network was 64.5% and 35.5% for the lower and upper sigmoid activation functions, respectively. The time elapsed for training and testing the traditional neural network was 0.03 seconds and 0.02 seconds, respectively; and took 7.3 seconds and 0.96 seconds to train and test the type-II fuzzy neural network, respectively. Increasing the number of neurons in the hidden layer of the traditional neural network to 2 improved its accuracy to 63.9%.

The process of annotating regular and irregular skin lesion borders used in training the classifiers is considered laborious and would involve a larger team to be able to label thousands of lesion borders, a task that could eventually improve the prediction accuracy.

6.6 Border irregularity detection using a CNN-Gaussian Naive Bayes ensemble

An ensemble based approach for detecting skin lesion border irregularity was proposed in [84]. The approach is depicted in Fig.6.17. After segmenting skin lesions using gradual focusing (Section 4.3) from the "ISIC 2018: Skin Lesion Analysis Towards Melanoma Detection" grand challenge datasets, the edges of the segmentation results are made sharper for better edge detection. This is carried out using the approach proposed in [85] which basically removes texture at any level without distorting edges through the use of a local regularization named Relativity-of-Gaussian (RoG) on which a global optimization is applied to identify potential edges at different scales. In other words, different scale edges are defined using different Gaussian kernels to preserve important structures with high resolution; edges that possess similar patterns in their neighbors would show more similar direction gradients. A global optimization function is subsequently defined to smooth the edges at different scales. 250 skin lesion borders extracted from those segmented images using the Canny edge detector were sent to a dermatologist to label as *regular* or *irregular*. Most of the images had an irregular border, summing up to 244 versus 6 regular bordered images; images were resized to 512×512 pixels. To make the most of the training data and to deal with the data imbalance, augmentation using some transformations has been applied (i.e. rotating, and flipping horizontally and vertically). 2000 images were generated after the augmentation process, with each class (regular or irregular) having 1000 images. This step was required for the training phase of the proposed approach. Fig.6.18 shows some samples of the skin lesion border images used in the training phase.

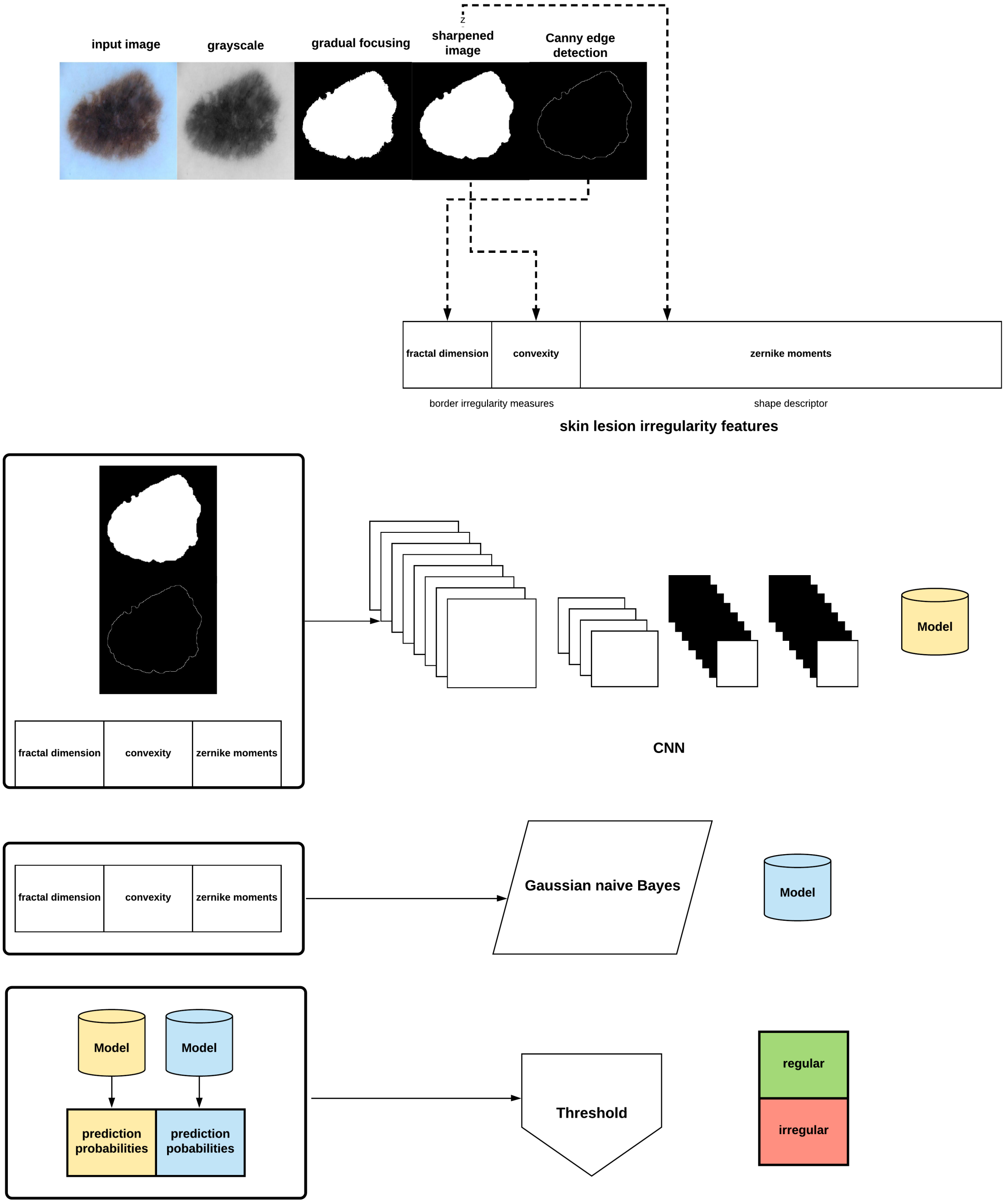


Figure 6.17: Border irregularity detection using a CNN-Gaussian Naive Bayes ensemble: The skin lesion image is firstly converted into grayscale, after which the skin lesion is segmented and smoothed, lesion border (edge) detected, and the lesion irregularity features extracted. A CNN is trained on the smoothed segmented image, skin lesion border, and the skin lesion irregularity features; while a Gaussian naive Bayes is trained on the irregularity features. The generated models are used to predict class probabilities, and a threshold is eventually used to determine the final decision (regular or irregular skin lesion border)

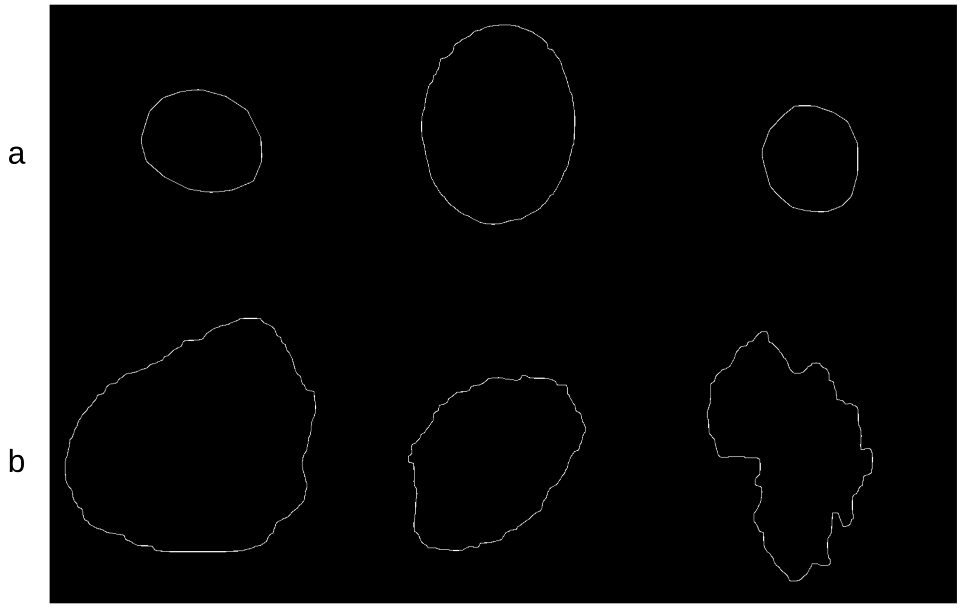


Figure 6.18: Samples of skin lesion border images used in the training phase and labeled by the dermatologist: (a) regular borders (b) represents irregular borders

The skin lesion border irregularity measurement step is then applied to the extracted borders, producing a 27-value feature vector (Fig.6.8) that describes the irregularity inherent in each extracted border. Table.6.8 shows the extracted border irregularity measures for the images shown in Fig.6.18, and the log-log graphs of the corresponding images are depicted in Fig.6.19 where the fractal dimension values are determined from the slope (the amount of change along the y-axis divided by the amount of change along the x-axis) of each plot. Convexity and Zernike moment values are extracted from the smoothed segmented images (the first 10 values of the 25-value Zernike moment vector are shown in the table). Fig.6.20 and Fig.6.21 show the original and smoothed segmented results corresponding to the skin lesion borders shown in Fig.6.18, respectively. The label column L in Table.6.8 is manually added and reflects the labeling made by the dermatologist. Fig.6.22 shows a box-and-whisker plot depicting the distribution of fractal dimension values for the regular and irregular skin lesion borders used in training the machine learning algorithms (classifiers). As can be noticed, the irregular skin lesion borders tend to move towards higher fractal dimension values (i.e. the more irregular the skin lesion border the higher the fractal dimension). Another plot is drawn in Fig.6.23

that depicts the distribution of convexity values, where irregular skin lesion borders tend to move away from the value 1 (less convex). Fig.6.24 depicts the relationship between the fractal dimension and convexity which shows that irregular borders (label:0) tend to have larger fractal dimension values and smaller convexity values, whilst regular borders (label:1) tend to have smaller fractal dimension values and larger convexity values.

Table 6.8: Border irregularity measures for the images presented in Fig.6.18. FD: Fractal Dimension, C: Convexity, ZM: Zernike Moment, L: Label (regular: 1; irregular: 0). Images 1.a, 2.a, and 3.a refer to the first image in row *a* from the left, middle image, and last image, respectively. This order applies to images in row *b*

Image	FD	C	ZM 1	ZM 2	ZM 3	ZM 4	ZM 5	ZM 6	ZM 7	L
1.a	1.0394	0.9573	0.3183	0.0004	0.0017	0.0027	0.0007	0.0051	0.0028	1
2.a	1.1486	0.9354	0.3183	0.0001	0.0010	0.0028	0.0004	0.0047	0.0015	1
3.a	1.0679	0.9482	0.3183	0.0003	0.0003	0.0028	0.0006	0.0039	0.0003	1
1.b	1.2348	0.9208	0.3183	0.0006	0.0031	0.0017	0.0013	0.0048	0.0051	0
2.b	1.1481	0.9006	0.3183	0.0010	0.0004	0.0014	0.0021	0.0013	0.0005	0
3.b	1.2510	0.7747	0.3183	0.0001	0.0024	0.0023	0.0001	0.0057	0.0040	0

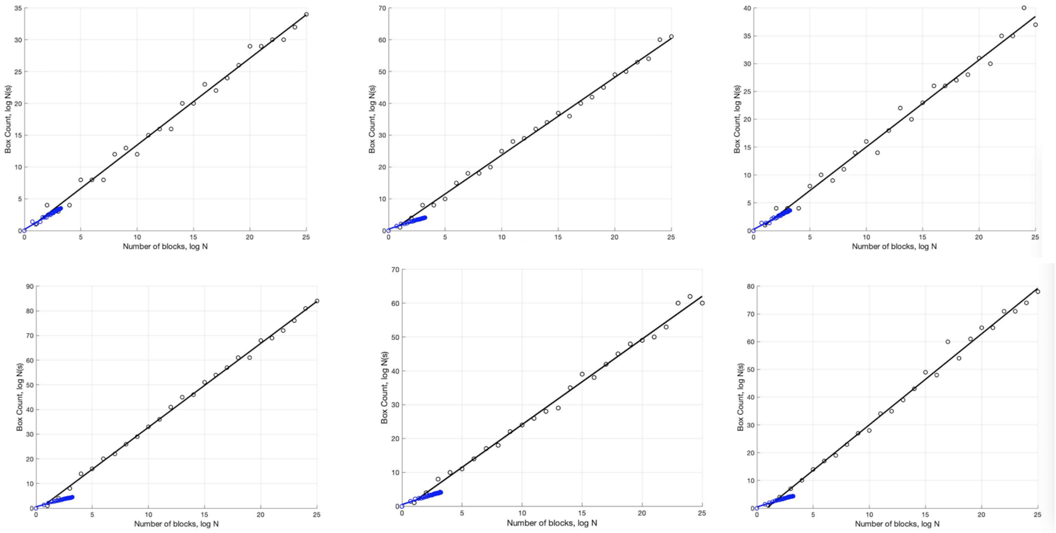


Figure 6.19: The log-log plots corresponding to the skin lesion borders shown in Fig.6.18 where the fractal dimension values are determined from the slope of the plot

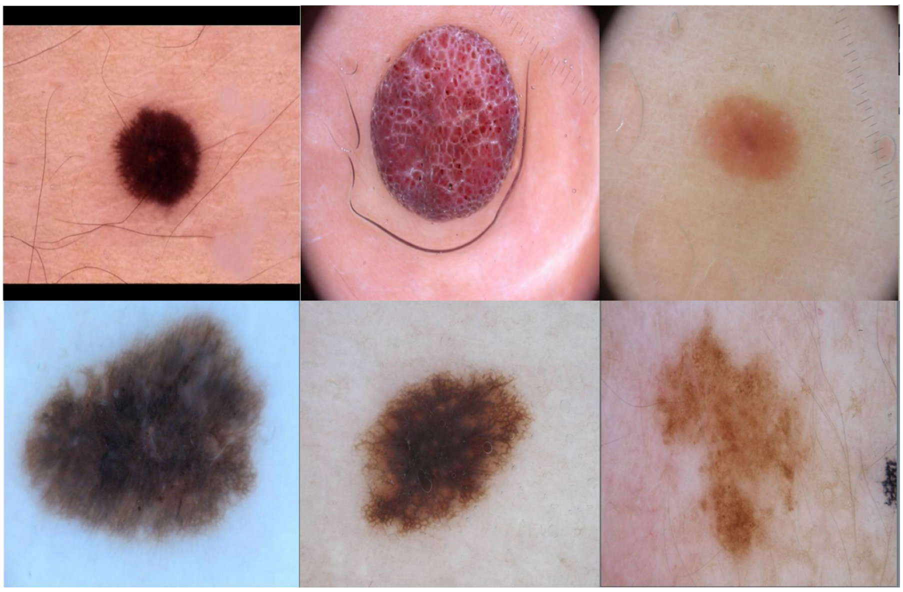


Figure 6.20: The original images corresponding to the skin lesion borders shown in Fig.6.18

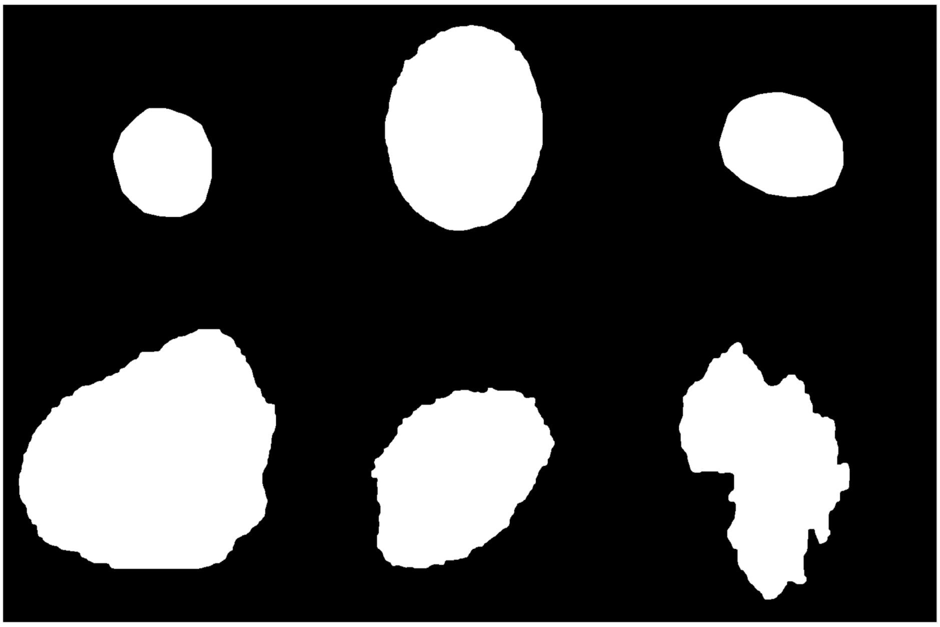


Figure 6.21: The smoothed segmented images corresponding to the skin lesion borders shown in Fig.6.18

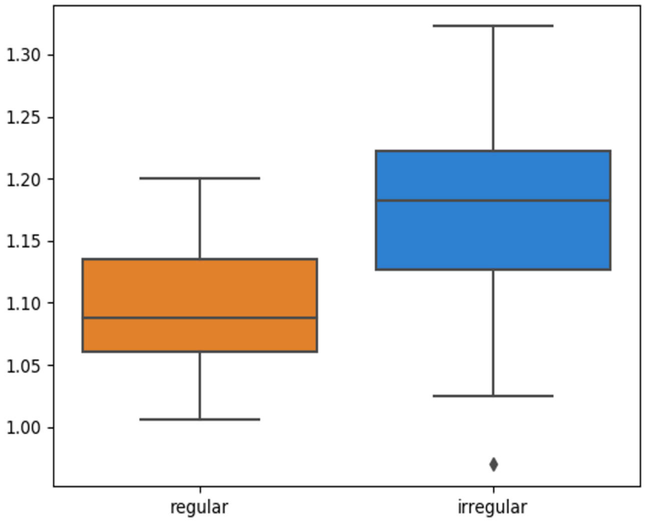


Figure 6.22: Box-and-whisker plot representing the fractal dimension distribution of the skin lesion borders (regular and irregular) in the training data

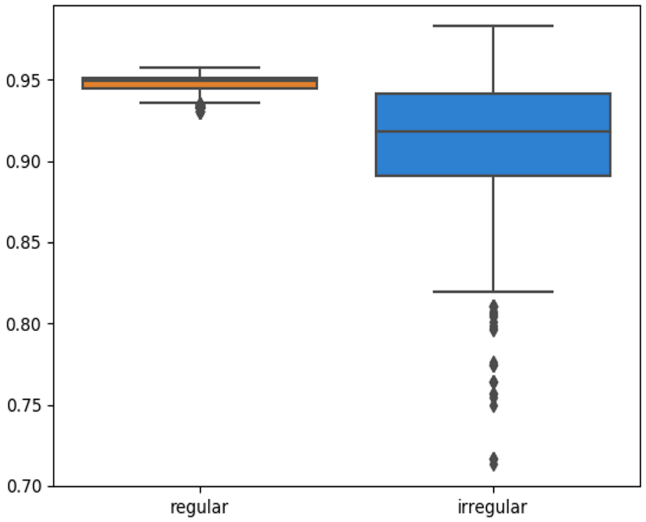


Figure 6.23: Box-and-whisker plot representing the convexity distribution of the skin lesion borders (regular and irregular) in the training data

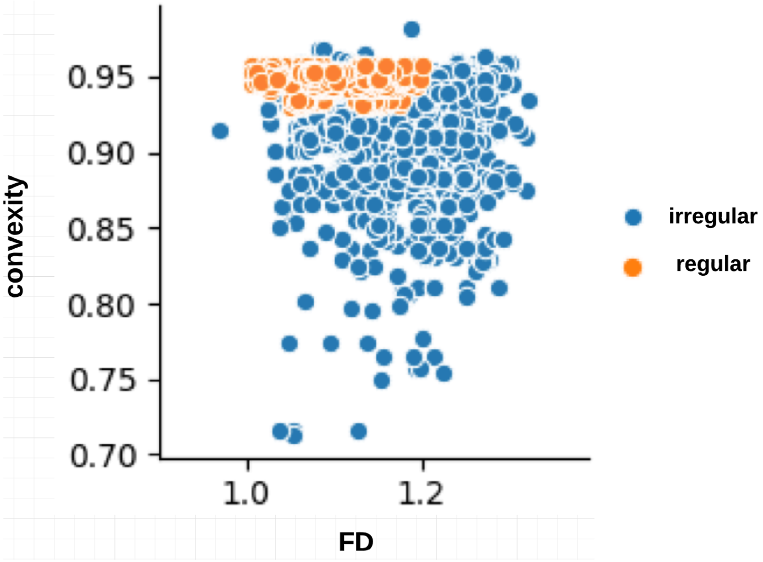


Figure 6.24: Relationship between fractal dimension and convexity values for regular (label:1) and irregular (label:0) skin lesion borders in the training dataset

The smoothed segmented images, skin lesion border images, and irregularity measures of the training data are used to train the CNN, which is composed of 5 convolutional layers, 5 max-pooling layers, and 2 dense layers. The convolutional layers use the ReLU activation function, the first dense layer uses the ReLU activation function, and the last dense layer uses the Sigmoid activation function. Adam is used as an optimization algorithm where the learning rate is set to 0.001. The CNN model was trained for 1 epoch on a Tesla P100 GPU; training for more epochs didn't improve the training accuracy. Gaussian naive Bayes is trained on the irregularity measures on an Intel(R) Core(TM) i7-4770HQ CPU @ 2.20GHz.

The proposed approach (Fig.6.17) was applied on 47 randomly selected test images extracted from the "ISIC 2018: Skin Lesion Analysis Towards Melanoma Detection" datasets, provided that those images were not used in the training phase of the approach. For evaluation purposes to compare the results with a groundt ruth, a dermatologist was asked to label the test images (the algorithms did not see before), resulting in 40 images being labeled as irregular and 7 images as regular. Fig.6.25 shows some samples of test images, smoothed segmented results and their extracted skin lesion borders.

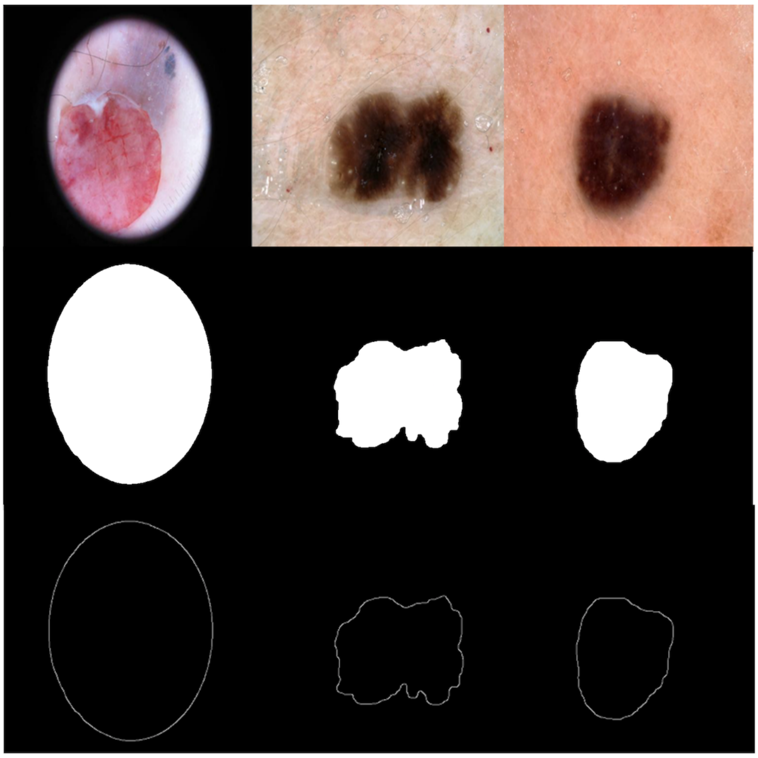


Figure 6.25: Samples of test images, smoothed segmented output, and their extracted skin lesion borders

The test samples prediction probabilities of the two classes (irregular and regular) obtained using the training model of both the CNN and Gaussian naive Bayes are combined together (i.e. ensemble), resulting in a total prediction probability P calculated based on the following equation:

$$P = \frac{CNN_{p_1} \times GnB_{p_1} \; CNN_{p_2} \times GnB_{p_2}}{2} \tag{6.13}$$

where CNN_{p_1} and GnB_{p_1} are the prediction probabilities of the *first* class (i.e. irregular) resulting from the CNN and Gaussian naive Bayes, respectively. CNN_{p_2} and GnB_{p_2} are the prediction probabilities of the *second* class (i.e. regular) resulting from the CNN and Gaussian naive Bayes, respectively. After obtaining the *prediction probability* of each test sample using Eq.6.13, a *threshold* is generated on those prediction probabilities to decide the final prediction (irregular or regular) according to Eq.6.14 that takes into account all the prediction probabilities including the peak (maximum) probability.

$$threshold = \frac{max(P)\ mean(P)}{2} \tag{6.14}$$

where $max(P)$ is the maximum prediction probability value amongst all test prediction probabilities, and $mean(P)$ is the mean (average) value of all test prediction probabilities. The final decision is eventually obtained using Eq.6.15.

$$Decision = \begin{cases} regular, & P_i < threshold) \\ irregular, & P_i > threshold) \end{cases} \tag{6.15}$$

where P_i is the prediction probability of test sample i.

The proposed approach resulted in 93.6% accuracy, where all the regular borders were predicted correctly, and 3 irregular borders were misclassified as regular. The elapsed time for training the CNN for 1 epoch and testing it on a Tesla P100 GPU evaluated to 7.1 minutes and 9.48 seconds, respectively. Training and testing the Gaussian naive Bayes on an Intel(R) Core(TM) i7-4770HQ CPU @ 2.20GHz together took 0.042 seconds. To understand the approach performance further (from different angles other than only accuracy), a confusion matrix has been generated as shown in Fig.6.26, in addition to finding the sensitivity, specificity, and F-score values, which resulted in 100%, 92.5%, and 96.1%, respectively.

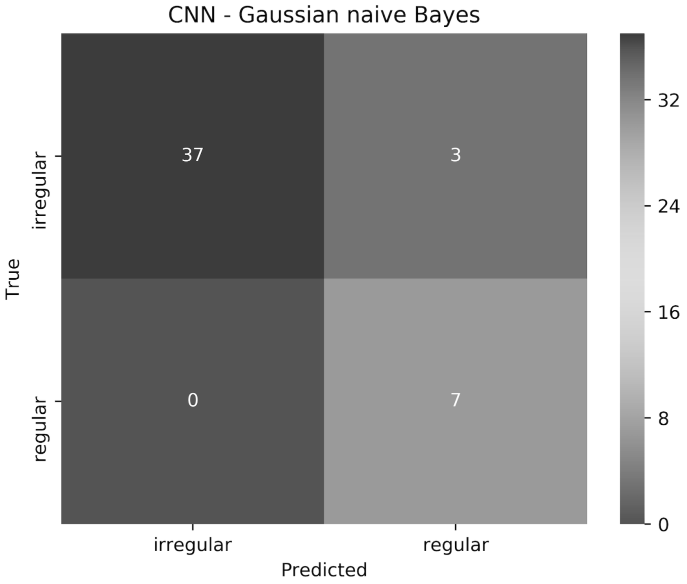


Figure 6.26: The confusion matrix of the test results obtained by the CNN - Gaussian naive Bayes ensemble

False Positive (FP) in this work is considered more important than False Negative (FN). In other words, if a patient had an irregular skin lesion border but was told (diagnosed) to have a regular border (FP), this might be life-threatening as opposed to classifying the patient to have an irregular skin lesion border while having a regular border (FN), which simply would send the patient for further investigation (i.e. biopsy). The approach works well in reducing FP and FN, with 3 misclassifications and 0 misclassifications, respectively. It should be emphasized that when evaluated separately, Gaussian naive Bayes and CNN result in 87.2% and 85.1% accuracy, respectively.

6.7 Summary

As border detection is considered an apparent prerequisite to determining skin lesion border irregularity, the chapter started by explaining FuzzEdge, a proposed edge detection approach that uses fuzzy concepts to represent an image; after defining the concepts, a filtering process was carried out to determine the final concept of each filtered pixel. A robust irregularity measure was also proposed which combines fractal dimension, convexity, and zernike moments. To automate the process of determining skin lesion border irregularity, skin lesions were first segmented (gradual focusing, U-Net), borders detected (FuzzEdge, Canny edge detector), border irregularity measures extracted, and classifiers used (F-MLP and CNN-Gaussian Naive Bayes ensemble) to learn the measures to objectively determine the presence of border irregularity in the skin lesion. In F-MLP, a fuzzy gradient descent was developed which considers the membership degrees of neurons, minimizing thereby the effects of ambiguous neurons on the neural network learning process; a type-II fuzzy sigmoid activation function was also proposed which allows the representation of the range (lowest and highest) of performance the fuzzy neural network is able to attain, where the fuzzy neural network with the highest performance (highest accuracy) could be utilized in the prediction process. F-MLP was trained on the irregularity measures and showed to provide better prediction accuracy than most of the state-of-the-art classification methods in general and its standard neural network counterpart in particular. As for the CNN-Gaussian Naive Bayes ensemble, the irregularity measures along with the segmented image and border image were used to train the CNN, while the measures alone were used to train a Gaussian naive Bayes. The models generated from both networks were eventually combined to form the ensemble. An outstanding performance was achieved using this approach.

Chapter 7

Conclusions and Future Work

In this chapter, a summary of the overall work presented in the thesis is provided in section 7.1, and recommendations of future work are given in section 7.2.

7.1 Summary

Due to the rapid increase in melanoma incidence and to the fact that the 5-year survival rate for early stage melanoma is significantly higher than melanomas that have spread to other parts of the body (94% vs. 17%), the early detection of melanoma has become a continuing public health priority. The subjectivity inherent in dermoscopic diagnosis in general and in melanoma indicators (i.e. ABCD features) in particular deem it necessary to develop an objective tool that could serve as an adjunct in the clinical assessment of skin lesions. The aim of this thesis is to develop objective measures and automated methods to evaluate the ABCD rule features in dermoscopy images.

To tackle the challenges associated with skin lesion segmentation, such as weak edges and the presence of artifacts, two segmentation approaches were proposed: gradual focusing, which takes the ambiguous pixels into consideration by revealing and affecting them to the appropriate clusters in a fuzzy clustering setting, and Otsu-II (an improvement of Otsu's method) coupled with pre-processing and post-processing stages; the two approaches outperformed U-Net. A comparative analysis was made between U-Net as a supervised deep learning based approach and an unsupervised deep learning based approach for skin lesion segmentation in dermoscopy images. The unsupervised approach was able to detect fine structures in skin lesions better than U-Net. However, U-Net showed to provide better accuracy in terms of Dice coefficient and Jaccard index (77.7% vs. 40% and 67.2% vs. 30.4%, respectively). An adaptation to the unsupervised approach caused the Dice and Jaccard index values to improve to 54.3% and 44%, respectively.

Objective measures involving statistical and geometrical measures have been proposed to extract asymmetry, border irregularity, color variegation, and diameter from skin lesions. The extracted asymmetry measures were used to train a decision tree which was then utilized for predicting the asymmetry of new skin lesion images; the approach was able to predict asymmetry with 80% accuracy. The border irregularity measure along with the segmented image and border image (two edge detectors have been used to detect lesion borders: FuzzEdge, a proposed approach that uses the fuzzy representation of image pixels, and the traditional Canny edge detector) were used to train a CNN, while the measure alone was used to train a Gaussian naive Bayes. The models generated from both networks were combined (ensemble) to test new images, and a threshold was created to determine the final classification decision from the test predictions. Results showed that the approach achieves outstanding accuracy, sensitivity, specificity, and F-score results, reducing concurrently false positives (FPs) and false negatives (FNs). The irregularity measure was also used to train a proposed F-MLP which contained a fuzzy gradient descent that considers the membership degrees of features/neurons, minimizing thereby the effects of ambiguous features/neurons on the neural network learning process. A type-II fuzzy sigmoid activation function was also used which allows to represent the range (lowest and highest) of performance the fuzzy neural network is able to attain, where the fuzzy neural network with the highest performance (highest accuracy) could be utilized in the prediction process. F-MLP showed to provide better prediction accuracy than most of the state-of-the-art classification methods in general and its standard neural network counterpart in particular. To determine color variegation, RGB images were converted to the CIELab color space, and a color palette which represents the image dominant colors was formed. If the pixel color belonged to the range of color shades of some suspicious color, or if more than 5% of the skin lesion pixels had their Minkowski distance to any of the suspicious colors less than a specified threshold, such pixels were considered to belong to that suspicious color. Feret's diameter was finally used to measure the diameter of the skin lesion, and spatial calibration was used to convert the diameter from pixels into millimeters.

Combining the work contributed throughout my PhD would produce a computer-aided diagnosis (CAD) system that is able to automatically and objectively evaluate the ABCD rule, and which could serve as an objective second opinion to the ABCD rule in general and its individual features in particular.

7.2 Future work

Envisioning the work proposed in this thesis as a core system to a larger CAD system, several items of future research can be identified. Sections 7.2.1 and 7.2.2 reflect my posters published in [220] and [221], respectively, and sections 7.2.3 - 7.2.5 emerge from some challenges encountered throughout my work.

7.2.1 Using deep learning to deblur skin lesion images

With the advent of teledermatology, it is crucial that clinical and dermoscopic images received remotely are of sufficiently high quality in order to correctly triage and manage referrals. Some of these images may be captured using a smartphone camera, or equivalent, and are prone to blurriness due to various artifacts (e.g. object motion, camera shake). Image deblurring techniques can be utilised to recover a latent image with more clarity. A variety of image deblurring algorithms have been proposed in literature, however such methods require intensive (sometimes heuristic) parameter tuning and expensive computation. A conditional Generative Adversarial Network (cGAN) [222] was used in [220] to create crisper photographs of blurry skin cancer dermoscopic images; preliminary results showed that cGAN was effective in deblurring images. As future work, cGAN can be trained using skin cancer images (as opposed to the cGAN model pre-trained on the GOPRO dataset used in the poster), and different low-pass filters can be applied to generate artificially blurred images; a Gaussian filter was used in the poster.

7.2.2 Neural networks and genomic data for melanoma detection

Melanoma is a heterogeneous disease. Its progression is driven by different patterns of oncogenic mutations. Driver genes, such as BRAF and NRAS can affect melanoma behaviour and their identification can have an impact on clinical management decisions. Using genomic data as the training input, the accuracy of a neural network in predicting whether the tissue source is melanoma or non-melanoma skin cancer has been assessed in [221], where it has been shown that the neural network was able to distinguish between melanoma and carcinoma. To build on top of this work, larger datasets can be used to help validate the performance of neural networks further, in addition to combining genomic data with other data sources such as skin lesion images and patient history.

7.2.3 Skin lesion features

The same rationale used in detecting the ABCD features in the thesis can be extended to tackle new features such as change (evolution) of the skin lesion, and attempt to automate other rules that aid in the early detection of melanoma such as the 7-point checklist for instance. In particular, as depicted in Fig.2.4, it has been shown that when tackling the ABCD features the proposed solutions were always based on some stage of the dermoscopic image analysis standard pipeline stages (segmentation, feature extraction, classification). It is thus hypothesized that working on solutions pertaining new skin lesion features would be based on one or more of those stages.

7.2.4 Low quality images

Due to the availability of public dermoscopic image datasets in general and the scarcity of skin lesion digital image datasets (i.e. taken using a mobile phone camera) in particular, the work in this thesis was mainly carried on dermoscopic images. With the emergence of many apps that help in organizing and archiving skin lesion images, large datasets of skin lesion images taken under lower quality settings, as opposed to those taken using a dermoscope, are now available. However, accessing such images is very complex due to privacy and regulations surrounding such images. It would be a big service to the research community if there would be some sort of collaboration between academia, startups, and regulatory bodies in making such datasets available for research purposes, especially with the challenges they bring up and the novel solutions they motivate.

7.2.5 Skin lesion features datasets

One of the challenges faced in my work was the unavailability of datasets pertaining skin lesion features like asymmetry and border irregularity. Although building such datasets is labour intensive as they require pre-processing and annotation by specialists, they however serve very well the research community especially with the rise in building machine learning and deep learning based approaches, which require the availability of sufficiently sized datasets relevant to the problem domain.

Bibliography

[1] NIH, “Skin cancer treatment (PDQ®)–patient version,” https://www.cancer.gov/types/skin/patient/skin-treatment-pdq,” Accessed on June, 10 2020.

[2] K. Wolff and H. Pehamberger, “Malignes melanom,” Füherkennung und prognose Wiener klinische Wochenschrift, 97(10), 1985.

[3] J. Han, G. Colditz, and D. Hunter,“Risk factors for skin cancers: a nested case control study within the nurses’ health study,” Int J Epidemiol, pp.1514–21, 2006.

[4] D. Parkin, D. Mesher, and P. Sasieni,“Cancers attributable to solar (ultraviolet) radiation exposure in the UK in 2010,” Br J Cancer, 105, pp.69-69, 2011.

[5] D. Eedy,“Dermatology: a specialty in crisis,” Clinical Medicine, 15(6), pp.509–10, 2015.

[6] JK. Schofield, D. Grindlay, and HC. Williams,“Skin conditions in the UK: a health care needs assessment,” Nottingham: Centre of Evidence Based Dermatology, University of Nottingham, 2009.

[7] I. Leigh,“Progress in skin cancer: the UK experience,” Br J Dermatol, pp.443-5, 171, 2014.

[8] British Association of Dermatologists, “Clinical services,” http://www.bad.org.uk/healthcare-professionals/clinical-services/, Accessed on June, 10 2020.

[9] G. Burg, “Das Melanom,” Serie Gesundheit: Piper/VCH, 1993.

[10] K. He, J. Sun, and X. Tang, “Guided image filtering,” IEEE Transactions on Pattern Analysis and Machine Intelligence, 35, pp.1397–1409, 2013.

[11] M. Hintz-Madsen, “A probabilistic framework for classification of dermatoscopic images,” Technical University of Denmark, 1998.

[12] NIH Consensus conference, “Diagnosis and treatment of early melanoma,” JAMA, 268(10), 1993.

[13] G. Argenziano, G. Fabbrocini, P. Carli, V. De Giorgi, E. Sammarco, M. Delfino, “Epiluminescence microscopy for the diagnosis of doubtful melanocytic skin lesions. Comparison of the ABCD rule of dermoscopy and a new 7-point checklist based on pattern analysis,” Arch Dermatol, 134(12), pp.1563-70, 1998.

[14] H. Soyer, J. Smolle, H. Kerl, and H. Stettner, “Early diagnosis of malignant melanoma by surface microscopy,” Lancet, 2(803), 1987.

[15] W. Stolz, O. Falco, P. Bliek, M. Kandthaler, W. Burgdorf, A. Cognetta, “Color Atlas of Dermoscopy, 2nd enlarged and completely revised edition” Berlin: Blackwell publishing, 2002.

[16] G1., Argenziano, HP. Soyer, S. Chimenti, R. Talamini, R. Corona, F. Sera, M. Binder, L. Cerroni, G. De Rosa, G. Ferrara, R. Hofmann-Wellenhof, M. Landthaler, SW. Menzies, H. Pehamberger, D. Piccolo, HS. Rabinovitz, R. Schiffner, S. Staibano, W. Stolz, I. Bartenjev, A. Blum, R. Braun, H. Cabo, P. Carli, V. De Giorgi, MG. Fleming, JM. Grichnik, CM. Grin, AC. Halpern, R. Johr, B. Katz, RO. Kenet, H. Kittler, J. Kreusch, J. Malvehy, G. Mazzocchetti, M. Oliviero, F. Ozdemir, K. Peris, R. Perotti, A. Perusquia, MA. Pizzichetta, S. Puig, B. Rao, P. Rubegni, T. Saida, M. Scalvenzi, S. Seidenari, I. Stanganelli, M. Tanaka, K. Westerhoff, IH. Wolf, O. Braun-Falco, H. Kerl, T. Nishikawa, K. Wolff, AW. Kopf, “Dermoscopy of pigmented skin lesions: results of a consensus meeting via the internet,” J Am Acad Derm, 48(5), pp.679–93, 2003.

[17] H. Pehamberger, A. Steiner, K. Wolff, “In vivo epiluminescence microscopy of pigmented skin lesions. I. Pattern analysis of pigmented skin lesions,” J. Am. Acad. Dermatol, 17, pp.571—583, 1987.

[18] H. Pehamberger, A. Steiner, K. Wolff, “In vivo epiluminescence microscopy of pigmented skin lesions. I. Pattern analysis of pigmented skin lesions,” J. Am. Acad. Dermatol, 17, pp.571—583, 1987.

[19] H.P. Soyer, G. Argenziano, R. Hofmann-Wellenhof, R. Johr , “Color atlas of melanocytic lesions of the skin,” Springer Science and Business Media, 2007.

[20] SW. Menzies, C. Ingvar, KA. Crotty, WH. McCarthy, “Frequency and morphologic characteristics of invasive melanomas lacking specific surface microscopic features,” Arch Dermatol, 1996.

[21] E. Stockfleth, “Cancer of the skin e-book,” Elsevier Health Science, 2001.

[22] J. Scharcanski and M. Emre Celebi, “Computer vision techniques for the diagnosis of skin cancer,” Springer Science & Business Media, 2013.

[23] JS. Henning, SW. Dusza, SQ. Wang, AA. Marghoob, HS. Rabinovitz, D. Polsky , and AW. Kopf, “The CASH (color, architecture, symmetry, and homogeneity) algorithm for dermoscopy,” J Am Acad Dermatol, 2007.

[24] A.A. Marghoob and Braun, “An atlas of dermoscopy, 2^{nd} edition,” CRC Press, 2012.

[25] P. Schmid-Saugeona, J. Guillodb, and J. Thirana, “Towards a computer-aided diagnosis system for pigmented skin lesions,” Comput Med Imaging Graph, 27(1), pp.65–78, 2003.

[26] D. Piccolo, G. Crisman, S. Schoinas, D. Altamura, and K. Peris, “Computer-automated ABCD versus dermatologists with different degrees of experience in dermoscopy,” Eur J Dermatol, pp.477-481, 2014.

[27] RJ. Friedman, DS. Rigel, and AW. Kopf, “Early detection of malignant melanoma: the role of physician examination and self-examination of the skin,” CA-Cancer J Clin, 35, pp.130–51, 1985.

[28] TW. McGover and MS. Litaker, “Clinical predictors of malignant pigmented lesions: a comparison of the Glasgow seven point check list and the American Cancer Society’s ABCDs of pigmented lesions,” J Dermatol Surg Oneal, 1992.

[29] RL. Barnhill, GC. Roush, MS. Ernstoff , and JM. Kirkwood, “Interclinician agreement on the recognition of selected gross morphologic features of pigmented lesions. Studies of melanocytic nevi V,” J Am Acad Dermatol, pp.185-190, 1992.

[30] N. Cascinelli, M. Ferrario, T. Tonelli, and E. Leo, “A possible new tool for clinical diagnosis of melanoma: the computer,” J Am Acad Dermatol, pp.361-367, 1987.

[31] R. White, OS. Rigel, RJ. Friedman, and E. Leo, “Computer applications in the diagnosis and prognosis of malignant melanom,” Dermatol Clin, 9, pp.695–702, 1991.

[32] N.R. Abbasi, H.M. Shaw, D.S. Rigel, R.J. Friedman, W.H. McCarthy, I Osman, A.W. Kopf, and D. Polsky, “Early diagnosis of cutaneous melanoma: revisiting the abcd criteria,” JAMA, 292, pp.2771–2776, 2004.

[33] R.P. Braun, H.S. Rabinovitz, M. Oliviero, A.W. Kopf, and J.H. Saurat, “Dermoscopy of pigmented skin lesions,” Journal American Academy of Dermatology, 52, pp.109–121, 2005.

[34] V. Ng and D. Cheung, “Measuring asymmetries of skin lesions,” In Proc. IEEE Int. Conf. systems, man, and cybernetics (SMC),Piscataway, NJ: IEEE Press, 5, pp.4211–6, 1997.

[35] V. Ng, B. Fung, and T. Lee, “Measuring asymmetries of skin lesions,” In Proc. IEEE Int. Conf. systems, man, and cybernetics (SMC),Piscataway, NJ: IEEE Press, 5, pp.4211–6, 1997.

[36] W. Stoecker, W. Li, and R. Moss, “Automatic detection of asymmetry in skin tumors,” Computerized Med. Imag. Graph., 16(3), pp.191–197, 1992.

[37] S. Seidenari, G. Pellacani, and A. Giannetti, “Digital videomicroscopy and image analysis with automatic classification for detection of thin melanomas,” Melanoma Research, 9, pp.163–171, 1999.

[38] L. Andreassi, R. Perotti, P. Rubegni, et. al., “Digital dermoscopy analysis for the differentiation of atypical nevi and early melanoma,” Arch Dermatol, 135, pp.1459–1465, 1999.

[39] M. d′Amico, M. Ferri, and I. Stanganelli, “Qualitative asymmetry measure for melanoma detection,” In ISBI2004: Proceedings of the IEEE International Symposiumon Biomedical Images. Arlington VA, pp.1155–1158, 2004.

[40] P. Frosini and C. Landi, “Size theory as a topological tool for computer vision,” Pattern Rec. and Image Analysis, pp.596–603, 1999.

[41] L. Ma, A. Guo, S. Zou, and W. Xu, “Irregularity and asymmetry analysis of skin lesions based on multi-scale local fractal distributions,” 2nd International Congress on Image and Signal Processing, 2009.

[42] K. Clawson, P. Morrow, B. Scotney, D. McKenna, and O. Dolan, “Determination of optimal axes for skin lesion asymmetry quantification,” In IEEE International Conference on Image Processing (ICIP 2007), 2, pp.453–456, 2007.

[43] A. Tenenhaus, A. Nkengne, J. Horn, C. Serruys, A. Giron, and B. Fertil, “Detection of melanoma from dermoscopic images of naevi acquired under uncontrolled conditions,”, Skin Research and Technology, 16(1), pp."85–97, 2010.

[44] A. Tenenhaus, A. Giron, E. Viennet, M. Bera, G Saporta and B. Fertil, “Kernel logistic PLS: A tool for supervised nonlinear dimensionality reduction and binary classification,” Computational Statistics & Data Analysis, 51, pp.4083–4100, 2007.

[45] Z. She, Y. Liu, and A. Damatoa, “Combination of features from skin pattern and ABCD analysis for lesion classification,” Skin Res Technol., 13(1), pp.25-33, 2007.

[46] I. Zaqout, “Diagnosis of skin lesions based on dermoscopic images using image processing techniques,” International Journal of Signal Processing, Image Processing and Pattern Recognition, 9(9), pp.189–204, 2016.

[47] R. Kasmi and K. Mokrani, “Classification of malignant melanoma and benign skin lesions: implementation of automatic ABCD rule,” IET Image Proc., 10(6), pp.448–455, 2016.

[48] J. Golston, W. Stoecker, R. Moss, and I. Dhillon, “Automatic detection of irregular borders in melanoma and other skin tumors,” Computerized Medical Imaging and Graphics, 16(3), pp.199-203, 1992.

[49] J. Golston, R. Moss, W. Stoecker, "Boundary detection in skin tumor images: an overall approach and a radial search algorithm," Pattern Recognition, 23, pp.1235-1247, 1990.

[50] C.T. Leondes, "General anatomy," CRC Press, 1997.

[51] M. Messadi, H. Cherifi, and A. Bessaid, "Segmentation and ABCD rule extraction for skin tumors classification," Journal of Convergence Information Technology (JCIT), 9(2), 2014.

[52] V. Ng and T. Lee, "Measuring border irregularities of skin lesions using fractal dimensions," SPIE Photonics China, Electronic Imaging and Multimedia Systems, 2898, pp.64-72, 1996.

[53] T. Lee, V. Ng, D. McLean, A. Coidman, R. Gallagher, J. Sale, "A multi-stage segmentation method for images of skin lesions," Proc. of IEEE Pacific Rim Conference on Commnnicaions, Compu1ers, and Signal Processing, pp.602-605, 1995.

[54] E. Claridge, J. Smith, P. Hall, "Evaluation of border irregularity in pigmented skin lesions against a consensus of expert clinicians," In: Berry E, Hogg D, Mardia K, Smith M, editors. Proc. medical image understanding and analysis (MIUA98). Leeds, UK: BMVA, pp.85-8, 1998.

[55] T. Lee, M. Atkins, R. Gallagher, C. MacAulay, A. Coldman, D. McLean, "Describing the structural shape of melanocytic lesions," Hanson KM, editor. Proc. SPIE, 3661, pp.1170-9, 1999.

[56] Lee, T and Ng, V. and Gallagher, R. and Coldman, A. and McLean, D., "Describing the structural shape of melanocytic lesions," Hanson KM, editor. Proc. SPIE, 3661, pp.1170-9, 1999.

[57] T. Lee and M. Atkins, "A new approach to measure border irregularity for melanocytic lesions," SPIE, 3979, pp.668-675, 2000.

[58] P. Rosin, "Irregularity index: a new border irregularity measure for cutaneous melanocytic lesions," Medical Image Analysis, 7(1), pp.47-64, 2003.

[59] T. Lee, D. McLean, and M. Stella, "Classification of pathological shapes using convexity measures," Pattern Recognition Letters, 30, pp.570-578, 2009.

[60] T. Lee and E. Claridge, "Predictive power of irregular border shapes for malignant melanomas," Skin Research and Technology, 11(1), pp.1-8, 2005.

[61] B. Aribisala and E. Claridge, "A border irregularity measure using a modified conditional entropy method as a malignant melanoma predictor," International Conference Image Analysis and Recognition, pp.914-921, 2005.

[62] L. Ma, B. Qin, W. Xu, and L. Zhu, "Multi-scale descriptors for contour irregularity of skin lesion using wavelet decomposition," In Proceedings of the 3rd International Conference on Biomedical Engineering and Informatics, 1, pp.414-418, 2010.

[63] J. Jaworek-Korjakowska and R. Tadeusiewicz, "Determination of border irregularity in dermoscopic color images of pigmented skin lesions," Conf Proc IEEE Eng Med Biol Soc, 2015.

[64] G. Grammatikopoulos, A. Hatzigaidas, A. Papastergiou, P. Lazaridis, Z. Zaharis, D. Kampitaki, and G. Tryfon, "Simple matlab tool for automated malignant melanoma diagnosis," WSEAS Trans. on Information Science and Applications, 3(4), pp.460-465, 2007.

[65] G. Argenziano, et al., "Interactive Atlas of Dermoscopy," Edra Medical Publishing and New Media, Milano, 2000.

[66] S. Umbaugh, R. Moss, and W. Stoecker, "Automatic color segmentation of images with applications in detection of variegated coloring in skin tumors," IEEE Eng Med. Biol., 8, pp.43–52, 1989.

[67] U. Kalwa, C. Legner, T. Kong, and S. Pandey, "Skin cancer diagnostics with an all-inclusive smartphone application," Symmetry, 11(6), 2019.

[68] G. Grammatikopoulos, A. Hatzigaidas, A. Papastergiou, P. Lazaridis, Z. Zaharis, D. Kampitaki, and G. Tryfon, "Simple MATLAB tool for automated malignant melanoma diagnosis," WSEAS Trans. Inf. Sci. Appl., vol.3, no.4, pp.460–465, 2007.

[69] R. Garnavi, M. Aldeen, and J. Bailey, “Computer-aided diagnosis of melanoma using border and wavelet-based texture analysis,” In: IEEE Trans. Inf. Technol. Biomed., 16(6), 2012.

[70] L. Thomas, P. Tranchand, F. Berard, T. Secchi, C. Colin, and G. Moulin, “Semiological value of ABCDE criteria in the diagnosis of cutaneous pigmented tumors,” Dermatology, 197, 11–17, 1998.

[71] Z. She and PJ. Fish, “Boundary detection of skin lesion using a fast snake algorithm,” Proceedings of the 16th Biennial International EURASIP Conference, pp.295-297, 2002.

[72] I. Jolliffe, “Principal component analysis,” Springer, 2nd Edition, New York, 2002.

[73] J. Jaworek-Korjakowska, “Automatic detection of melanomas: an application based on the ABCD criteria,” Information Technologies in Biomedicine, pp.67-76, 2012.

[74] N. Otsu, “A threshold selection method from gray-level histograms,” IEEE Trans. Sys., Man., Cyber., 9(1), pp.62-66, 1979.

[75] C. Harris and M. Stephens, “A combined corner and edge detector,” Proceedings of the Alvey Vision Conference, pp.147-151, 1988.

[76] M. Anantha, R.H. Moss, W.V. Stoecker, “Detection of pigment network in dermatoscopy images using texture analysis,” Comput. Med. Imaging Graph, 28(5), pp.225-234, 2004.

[77] M. Ramezani, A. Karimian, P. Moallem, “Automatic detection of malignant melanoma using macroscopic images,” Journal of Medical Signals and Sensors, 4(4), 281-290, 2014.

[78] R.M. Haralick, K. Shanmugam, I. Dinstein, “Textural features for image classification,” IEEE Transactions on Systems, Man, and Cybernetics, pp.610-621, 1973.

[79] Q. Abbas, M.E. Celebi, I.F. García, “Hair removal methods: a comparative study for dermoscopy images,” Biomedical Signal Processing and Control, 6(4), pp.395-404, 2011.

[80] V. Caselles, R. Kimmel, and G. Sapiro, “Geodesic active contours,” International Journal of Computer Vision, 22, pp.61-79, 1997.

[81] N. Codella, D. Gutman, M. Celebi, B. Helba, MA. Marchetti, S. Dusza, A. Kalloo, K. Liopyris, N. Mishra, H. Kittler, A. Halpern, “Skin lesion analysis toward melanoma detection: a challenge at the 2017 international Ssmposium on biomedical imaging (ISBI), hosted by the international skin imaging collaboration (ISIC),” arXiv: 1710.05006 [cs.CV] Available: https://arxiv.org/abs/1710.05006, 2017.

[82] P. Tschandl, C. Rosendahl, H. Kittler, “The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions,” Sci. Data, doi:10.1038/sdata.2018.161, 5, 2018.

[83] J. Canny, “A computational approach to edge detection,” IEEE Trans. Pattern Anal. Machine Intell., 35(8), pp.679-698, 1986.

[84] A. Ali, J. Li, G. Yang, and S.J. O’Shea, “A machine learning approach to automatic detection of irregularity in skin lesion border using dermoscopic images,” PeerJ Computer Science, 2020.

[85] B. Cai, X. Xing, and X. Xu, “Edge Structure preserving smoothing via Relativity-of-Gaussian,” Proc. IEEE Int. Conf. Image Process. (ICIP), 250-254, 2017.

[86] A. Ali, J. Li, S.J. O’Shea, G. Yang, T. Trappenberg, and X. Ye, “A deep learning 835 based approach to skin lesion border extraction with a novel edge detector in dermoscopy images,” The International Joint Conference on Neural Networks - IEEE IJCNN, 2019.

[87] A. Ali, J. Li, S. Kanwal, G. Yang, A. Hussain, and S.J. O’Shea, “A novel fuzzy multilayer perceptron (F-MLP)for the detection of irregularity in skin lesion border using dermoscopic images,” Frontiers in Medicine, 2020.

[88] A. Ali, J. Li, and S.J. O’Shea, “Towards the automatic detection of skin lesion shape asymmetry, color variegation and diameter in dermoscopic images,” PLoS ONE, 2020.

[89] N. Codella, J. Cai, M. Abedini, R. Garnavi, A. Halpern, and J.R. Smith, “Deep learning, sparse coding, and SVM for melanoma recognition in dermoscopy images,” Machine Learning in Medical Imaging, Munich, Springer, pp.118–126, 2015.

[90] T. Yoshida and H. Iyatomi, "Alignment of major axis for automated melanoma diagnosis with deep learning approach," Proceedings of the Fuzzy System Symposium, 31, pp.379–382, 2015.

[91] M. Attia, M. Hossny, S. Nahavandi, and A. Yazdabadi, "Skin melanoma segmentation using recurrent and convolutional neural networks," Proceedings of the IEEE International Conference on Computer Vision, 2015.

[92] J. Kawahara, A. BenTaieb, and G. Hamarneh,"Deep features to classify skin lesions," IEEE ISBI, pp.1397–1400, 2016.

[93] J. Kawahara and G. Hamarneh,"Multi-resolution-tract CNN with hybrid pretrained and skin-lesion trained layers," Medical Image Computing and Computer-Assisted Intervention Workshop on Machine Learning in Medical Imaging (MICCAI MLMI), 2016.

[94] J. Premaladh and K.S. Ravichandran, "Novel approaches for diagnosing melanoma skin lesions through supervised and deep learning algorithms," Journal of Medical Systems, 2016.

[95] M. Jafari, M.H.E. Nasr-Esfahani, N. Karimi, S.M.R. Soroushmehr, S. Savami, and K. Najarian, "Extraction of skin lesions from non-dermoscopic images using deep learning," Eprint arXiv:1609.02374 [cs.CV]. Available: https://arxiv.org/abs/1609.02374, 2016.

[96] E. Nasr-Esfahani, S. Samavi, N. Karimi, S.M.R. Soroushmehr, M.H. Jafari, K. Ward, and K. Najarian, "Melanoma detection by analysis of clinical images using convolutional neural network," 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2016.

[97] S. Sabbaghi, M. Aldeen, and R. Garnavi, "A deep bag-of-features model for the classification of melanomas in dermoscopy images," IEEE Engineering in Medicine and Biology Society (EMBC), 2016.

[98] V. Pomponiu, H. Nejati, and N. Cheung, "Deepmole: Deep neural networks for skin mole lesion classification," IEEE International Conference on Image Processing (ICIP), 2016.

[99] T. Majtner, S. Yildirim-Yayilgan, and J. Yngve Hardeberg, “Combining deep learning and hand-crafted features for skin lesion classification,” 6th International Conference on Image Processing Theory, Tools and Applications (IPTA), 2016.

[100] F. Cıcero, A. Oliveira, and G. Botelho, “Deep learning and convolutional neural networks in the aid of the classification of melanoma,” SIBGRAPI, 2016.

[101] A. Menegola, M. Fornaciali, R. Pires, S. Avila, and E. Valle, “Towards automated melanoma screening: exploring transfer learning schemes,” arXiv:1609.01228, 2016.

[102] S. Demyanov, R. Chakravorty, M. Abedini, A. Halpern, and R. Garnavi, “Classification of dermoscopy patterns using deep convolutional neural networks,” IEEE Int Symp Biomedical Imaging, pp.364-368, 2016.

[103] M. Jafari, N. Karimi, E. Nasr-Esfahani, S. Samavi, S. Soroushmehr, K. Ward, and K. Najarian, “Skin lesion segmentation in clinical images using deep learning,” 23rd International Conference Pattern Recognition (ICPR), 2016.

[104] P. Sabouri and H. GholamHosseini, “Lesion border detection using deep learning,” IEEE Congress on Evolutionary Computation (CEC), pp.1416-1421, 2016.

[105] P. Salunkhe and V. Mehta, “Intelligent mirror: detecting skin cancer (melanoma) using convolutional neural network with augmented reality feedback,” International Journal of Computer Applications, vol.54, no.6, pp.4-7, 2016.

[106] E. Karabulut and T. Ibrikci, “Texture analysis of melanoma images for computer-aided diagnosis,” Annual Int’l Conference on Intelligent Computing, Computer Science Information Systems (ICCSIS-16), 2016.

[107] L. Yu, H. Chen, Q. Dou, J. Qin, and P.A. Heng, “Automated melanoma recognition in dermoscopy images via very deep residual networks,” IEEE Trans. Med. Imaging, vol.36, no.4, 2017.

[108] N. Codella, Q. Nguyen, S. Pankanti, D. Gutman, B. Halpern, and J. Smith, “Deep learning ensembles for melanoma recognition in dermoscopy images,” IBM Journal of Research and Development, vol.61, no.45, 2017.

[109] A. Esteva, B. Kuprel, R.A., Novoa, J. Ko, S.M. Swetter, H.M. Blau, and S. Thrun, “Dermatologist-level classification of skin cancer with deep neural networks,” Nature, 2017.

[110] A. Menegola, M. Fornaciali, R. Pires, F.V. Bittencourt, S. Avila, and E. Valle, “Knowledge transfer for melanoma screening with deep learning,” IEEE ISBI, 2017.

[111] AR. Lopez, X. Giro-i-Nieto, J. Burdick, and O. Marques, “Skin lesion classification from dermoscopic images using deep learning techniques,” IEEE International Conference on Biomedical Engineering (BioMed), pp.49–54, 2017.

[112] A. Kwasigroch, A. Mikołajczyk, and M. Grochowski, “Deep convolutional neural networks as a decision support tool in medical problems – malignant melanoma case study,” Polish Control Conference, KKA 2017: Trends in Advanced Intelligent Control, Optimization and Automation, pp.848-856, 2017.

[113] P. Mirunalini, A. Chandrabose, V. Gokul, and S.M. Jaisakthi, “Deep learning for skin lesion classification,” arXiv:1703.04364v1, 2017.

[114] M.S. Elmahdy, S.S. Abdeldayem, and I.A. Yassine, “Low quality dermal image classification using transfer learning,” IEEE EMBS International Conference on Biomedical & Health Informatics, 2017.

[115] M. Attia, M. Hossny, S. Nahavandi, and A. Yazdabadi, “Spatially aware melanoma segmentation using hybrid deep learning techniques,” arXiv:1702.07963, 2017.

[116] Y. Yuan, M. Chao, and Y.-C. Lo, “Automatic skin lesion segmentation using deep fully convolutional networks with jaccard distance,” IEEE Trans. Med. Imaging, vol.36, no.9, pp.1876-1886, 2017.

[117] X. Zhang, “Melanoma segmentation based on deep learning,” Computer Assisted Surgery, DOI: 10.1080/24699322.2017.1389405, pp.267-277, 2017.

[118] D.Raupov, O. Myakinin, I. Bratchenko, and V. Zakharov, “Deep learning on OCT images of skin cancer,” Frontiers in Optics, 2017.

[119] B. Bozorgtabar, S. Sedai, P.K. Roy, and R. Garnavi, “Skin lesion segmentation using deep convolution networks guided by local unsupervised learning,” IBM Journal of Research and Development, vol.61, 2017.

[120] H. Liao and J. Luo, “A deep multi-task learning approach to skin lesion classification,” The AAAI-17 Joint Workshop on Health Intelligence, 2017.

[121] J. Burdick, O. Marques, J. Weinthal, and B. Furht, “Rethinking skin lesion segmentation in a convolutional classifie,” Journal of Digital Imaging, 2017.

[122] Z. Yu, X. Jiang, T. Wang, and B. Lei, “Aggregating deep convolutional features for melanoma recognition in dermoscopy images,” Machine Learning in Medical Imaging. MLMI 2017. Lecture Notes in Computer Science, Springer Cham, vol.10541, 2017.

[123] A. Kwasigroch, A. Mikołajczyk, and M. Grochowski, “Deep neural networks approach to skin lesions classification — a comparative analysis,” 22nd International Conference on Methods and Models in Automation and Robotics (MMAR), 2017.

[124] S.V. Georgakopoulos1, K. Kottari1, K. Delibasis1, V.P. Plagianakos1, and I. Maglogiannis, “Detection of malignant melanomas in dermoscopic images using convolutional neural network with transfer learning,” Springer International Publishing AG, G. Boracchi et al. (Eds.): EANN 2017, CCIS 744, pp.404-414, 2017.

[125] J. Monika and H. Soyer, , “Automated diagnosis of melanoma,” Med J Aust, vol.207, no.8, pp.361-362, 2017.

[126] N. Codella, J. Cai, M. Abedini, R. Garnavi, A. Halpern, J.R. Smith,“Deep learning, sparse coding, and SVM for melanoma recognition in dermoscopy images,” Machine Learning in Medical Imaging, Munich, Springer, pp.118-126, 2015.

[127] M. Silveira and J. Nascimento, “Comparison of segmentation methods for melanoma diagnosis in dermoscopy images,” IEEE Journal of Selected Topics in Signal Processing, 3(1), 2009.

[128] M. Kass, A. Witkin, and D. Terzopoulos, “Snakes: active contour models,” International Journal of Computer Vision, pp.189–204, 1988.

[129] J. Shan, “A fully automatic segmentation method for breast ultrasound images,” PhD thesis, Utah State University, 2011.

[130] M. Jafari, N. Karimi, E. Nasr-Esfahani, S. Samavi, S. Soroushmehr, K. Ward, and K. Najarian, “Skin lesion segmentation in clinical images using deep learning,” 23rd International Conference Pattern Recognition (ICPR), 2016.

[131] D. Gutman, N. Codella, E. Celebi, B. Helba, M. Marchetti, N. Mishra, and A. Halpern, “Skin lesion analysis toward melanoma detection: a challenge at the international symposium on biomedical imaging (ISBI) 2016, hosted by the international skin imaging collaboration (ISIC),” eprint arXiv:1605.01397, 2016.

[132] Y. Yuan, M. Chao, and Y.-C. Lo, “Automatic skin lesion segmentation using deep fully convolutional networks with jaccard distance,” IEEE Trans. Med. Imaging, 30(9), pp.1876—1886, 2017.

[133] T. Mendonca, P.M. Ferreira, J.S. Marques, A.R. Marcal, and J. Rozeira, “H2 - a dermoscopic image database for research and benchmarking,” Conf Proc IEEE Eng Med Biol Soc., 201, pp.5437-40, 2013.

[134] M. Al-Masni, M. Al-Antari, M. Choi, S. Han, and T. Kim, ‘Skin lesion segmentation in dermoscopy images via deep full resolution convolutional networks,” Computer Methods and Programs in Biomedicine, pp.221–231, 2018.

[135] J. Long, E. Shelhamer, and T. Darrell, “Fully convolutional networks for semantic segmentation,” Eprint arXiv:1411.4038 [cs.CV]. Available: https://arxiv.org/abs/1411.4038, 2015.

[136] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” In MICCAI, 2015.

[137] V. Badrinarayanan, A. Kendall, and R. Cipolla, “Segnet: a deep convolutional encoder-decoder architecture for image segmentation,” arXiv preprint arXiv:1511.00561, 2015.

[138] V. Gurum Munirathnam, N. Yarlapati Ganesh, S. Little, and N. O'Connor, "A deep residual architecture for skin lesion segmentation," In: ISIC Skin Image Analysis Workshop and Challenge at MICCAI, 2018.

[139] S. Izadi, Z. Mirikharaji, J. Kawahara, and G. Hamarneh, "Generative adversarial networks to segment skin lesions," In IEEE 15th International Symposium on Biomedical Imaging, pp.881–884, 2018.

[140] P. Maji and S. Pal, "Maximum class separability for rough-fuzzy c-means based brain MR image segmentation," T. Rough Sets, 9(2), pp.114–134, 2008.

[141] H. Tizhoosh, "Image thresholding using type II fuzzy sets," Pattern Recognition, 38(12), pp.2363–2372, 2005.

[142] N. Sladoje, J. Lindblad, and I. Nyström, "Defuzzification of discrete objects by optimizing area and perimeter similarity," In: J. Kittler, M. Petrou, M. Nixon (Eds.), Proc. of 17th International Conference on Pattern Recognition (ICPR 2004), IEEE Comp. Society, Cambridge, UK, 3(12), pp.526–529, 2004.

[143] N. Sladoje, J. Lindbald, and I. Nyström, "Defuzzification of spatial fuzzy sets by feature distance minimization," Image and Vision Computing, 29(2-3), pp.127-141, 2011.

[144] N. Boujemaa, G. Stamon, J. Lemoine, and E. Petit, "Fuzzy ventricular endocardium detection with gradual focusing decision," 14th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 14, 1992.

[145] J.M. Sung, H.G. Ha, B.Y. Choi, and Y.H. Ha, "Image thresholding based on within-class standard deviation using standard deviation," In Proceedings of SPIE - The International Society for Optical Engineering, 2014.

[146] J. Wang, H. Jiang, Z. Yuan, M.-M. Cheng, X. Hu, and N. Zheng, "Salient object detection: A discriminative regional feature integration approach," Int. J. Comput. Vision, 123(2), pp.251–268, 2017.

[147] M. Jahanifar, N. Zamani Tajeddin, B. Mohammadzadeh Asl, and A. Gooya, “Supervised saliency map driven segmentation of lesions in dermoscopic images,” IEEE Journal of Biomedical and Health Informatics, 2018.

[148] K.I. Laws, “Rapid texture identification,” In Proc. 24th Annu. Tech. Symp., pp.376–381, 1980.

[149] C. Barata, M.E. Celebi, and J.S. Marques, “Improving dermoscopy image classification using color constancy,” IEEE J. Biomed. Health Informat. 19(3), pp.1146–1152, 2015.

[150] Q. Chen, D. Li, and C. Tang, “KNN matting,” IEEE transactions on patern analysis and machine intelligence, 9, pp.2175–2188, 2013.

[151] A. Ali, J. Li, and T. Trappenberg, “Supervised versus unsupervised deep learning based methods for skin lesion segmentation in dermoscopy images,” I32nd Canadian Conference on Artificial Intelligence (CanadianAI 2019), Kingston, Ontario, Canada, May 28-31, 2019.

[152] J. Zhu, J. Mao, and A. Yuille, “Learning from weakly supervised data by the expectation loss svm (e-svm) algorithm,” In NIPS, 2014.

[153] F. Chang, Y. Lin, and K. Hsu, “Multiple structured-instance learning for semantic segmentation with uncertain training data,” In CVPR, 2014.

[154] D. Pathak, P. Krahenbuhl, and T. Darrell, “Constrained convolutional neural networks for weakly supervised segmentation,” In ICCV, 2015.

[155] W. Shimoda and K. Yanai, “Distinct class-specific saliency maps for weakly supervised semantic segmentation,” In ECCV, 2016.

[156] A. Kanezaki, “Unsupervised image segmentation by backpropagation,” Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2018.

[157] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, “Slic superpixels compared to state-of-the-art superpixel methods,” IEEE Trans. on Pattern Analysis and Machine Intelligence, 14(11), 2012.

[158] N. Codella, D. Gutman, M. Emre Celebi, B. Helba, M. Marchetti, S. Dusza, A. Kalloo, K. Liopyris, N. Mishra, H. Kittler, A. Halpern, "Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (ISBI), Hosted by the International Skin Imaging Collaboration (ISIC)," arXiv:1710.05006, 2017.

[159] P. Jaccard, "Etude comparative de la distribution orale dansune portion des alpes et des jura," Bulletin de la Société Vaudoise des Sciences Naturelles, 37, 547–579, 1901.

[160] C. Narayana, E. Sreenivasa Reddy, and M. Seetharama Prasad, "Automatic image segmentation using ultrafuzziness," International Journal of Computers and Applications, doi:10.5120/7677-0977, 49(12), pp.6–13, 2012.

[161] R. Dice, "Measures of the amount of ecologic association between species," Ecology, doi:10.2307/1932409. JSTOR 1932409, 26(3), pp.297–302, 1945.

[162] A. Vedaldi and S. Soatto, "Quick shift and kernel methods for mode seeking," In Proc. ECCV, 2008.

[163] G. Klambauer, T. Unterthiner, A. Mayr, and S. Hochreiter, "Self-normalizing neural networks," arXiv:1706.02515, 2017.

[164] P. Felzenszwalb and D. Huttenlocher, "Efficient graph-based image segmentation," International Journal of Computer Vision (IJCV), 59(2), pp.167-181, 2004.

[165] F. Nachbar, W. Stolz, T. Merkle, A.B. Cognetta, T. Vogt, M. Landthaler, P. Bilek, O. Braun-Falco, and G. Plewig, "The abcd rule of dermatoscopy," Journal of the American Academy of Dermatology, 30(4), pp.551–559, 1994.

[166] D. Lowe, "Distinctive image features from scale-invariant keypoints," IJCV, 60(2), pp.91–110, 2004.

[167] M. Hu, "Visual pattern recognition by moment invariants," IRE Transactions on Information Theory, 8, pp.179–187, 1962.

[168] C. Li, "Content-based microscopic image analysis," Logos Verlag, 2016.

[169] I. Sethi, “Entropy nets: from decision trees to neural networks,” Proc. IEEE, 78(10), pp.1605–1613, 1990.

[170] C. Li, “Causal sensitivity analysis for decision trees,” UWSpace. http://hdl.handle.net/10012/9076, 2015.

[171] R. Quinlan, “Induction of decision trees,” Machine learning, 1(1), pp.81–106, 1986.

[172] L. Breiman, J. Friedman, C.J. Stone, and R.A. Olshen, “Classification and regression trees,” Taylor Francis, 1984.

[173] M. Keefe, D. Dick, R. Wakeel, “A study of the value of the seven-point checklist in distinguishing benign pigmented lesions from melanoma,” Clinical Experimental Dermatol, pp.167-171, 1990.

[174] R. Dellavalle, “United States skin disease needs assessment, an issue of dermatologic slinics - E-Book,” Elsevier Health Sciences, 2012.

[175] D. Rigel, R. Friedman, A. Kopf, and D. Polsky, “ABCDE-an evolving concept in the early detection of melanoma,” Arch Dermatol, 141, 2005.

[176] M. Celebi, R. Kingravi, H. Iyatomi, Y.. Aslandogan, W. Stoecker, R.Moss, J. Malters, J. Grichnik, A. Marghoob, H. Rabinovitz, and S. Menzies, “Border detection in dermoscopy images using statistical region merging,” Skin Research and Technology, 14(3), pp.347-353, 2008.

[177] B. Erkol, R. Moss, R. Stanley, W. Stoecker, and E. Hva-tum, “Automatic lesion boundary detection in dermoscopy images using gradient vector flow snakes,” Skin Res. and Technol., 11, pp.17-26, 2005.

[178] M. Celebi, H. Kingravi, J. Lee, “Fast and accurate border detection in dermoscopy images using statistical region merging,” In Proc. SPIE Medical Imaging, 2007.

[179] R. Nock and F. Nielsen, “Statistical region merging,” IEEE Trans. on Pattern Analysis and Machine Intelligence, 20(11), pp.1452-1458, 2004.

[180] M. Celebi, Y. Aslandogan, W. Stoecker, H. Iyatomi, H. Oka, and X. Chen, “Unsupervised border detection in dermoscopy images,” Skin Res. Technol., 11(4), pp.454-462, 2007.

[181] Y. Deng and B.S. Manjunath, “Unsupervised segmentation of color-texture regions in images and video,” IEEE Trans Pattern Analysis Machine Intelligence, 23, pp.800-810, 2001.

[182] P. Tzekis, A. Papastergiou, A. Hatzigaidas, Z. Zaharis, D. Kampitaki, P. Lazaridis, and M. Goula, “A simple algorithm for automated skin lesion border detection,” WSEAS Transactions on Signal Processing, 5(6), pp.229-240, 2009.

[183] Q. Abbas, “A aoftware approach for border detection using pigmented skin lesions,” IJCSNS International Journal of Computer Science and Network Security, 17(5), 2017.

[184] Z. Zhang, W. Stoecker, R. Moss, “Border detection on digitized skin tumor images,” IEEE Trans Med Imaging, 19(11). pp.1128-1143, 2000.

[185] M. Celebi, Y. Aslandogan, P. Bergstresser, “Unsupervised border detection of skin lesion images,” In Int. Conf. Information Technology: Coding and Computing, 2, pp.123-128, 2005.

[186] Q. Abbas, I. Garcia, and M. Rashid, “Automatic skin tumor border detection for digital dermoscopy using a novel digital image analysis scheme,” British Journal of Biomedical Science, 67, pp.177-183, 2010.

[187] S. Umbaugh, R. Moss, W. Stoecker, “An automatic color segmentation algorithm with application to identification of skin tumor borders,” Computerized Medical Imaging and Graphics, 16(3), pp.227-235, 1992.

[188] M. Celebi, Q. Wen, H. Iyatomi, K. Shimizu, H. Zhou, and G. Schaefer, “A state-of-the-art survey on lesion border detection in dermoscopy images,” In Dermoscopy Image Analysis, M. E. Celebi, T. Mendonca and J. S. Marques, Eds., Boca Raton, CRC Press, pp.97-129, 2015.

[189] P. Sabouri and H. GholamHosseini, “Lesion border detection using deep learning,” IEEE Congress on Evolutionary Computation (CEC), pp.1416–1421, 2016.

[190] A. Jadhav, A. Ghontale, and V. Shrivastava, “Segmentation and border detection of melanoma lesions using convolutional neural network and SVM,” Computational Intelligence: Theories, Applications and Future Directions, pp.97-108, 2018.

[191] C. Lee, Y. Kuo, and P. Yu, “Weighted fuzzy mean filters for image processing,” Fuzzy Sets Syst., pp.157-180, 1997.

[192] H.J. Zimmermann, “Fuzzy set theory and its applications,” Kluwer Academic Publishers, second edition, 1991.

[193] K. Hua, “A comparison of edge detection methods for segmentation of skin lesions in mobile-phone-quality images,” Bachelor of Science thesis, Edith Cowan University, 2007.

[194] N. Kanopoulos, N. Vasanthavada, and R.L. Baker, “Design of an Image Edge Detection Filter using the Sobel Operator,” IEEE Journal of SolidState Circuits, 23(2), pp.358-367, 1988.

[195] E. Claridge, P. Hall, J. Keefe, P. Allen, “Shape analysis for classification of malignant melanoma,” J Biomed Eng, 14(1), pp.229-234, 1992.

[196] A. Piantanelli, P. Maponi, L. Scalise, S. Serresi, A. Cialabrini, A. Basso, “Fractal characterisation of boundary irregularity in skin pigmented lesions,” In Medical Biological Engineering Computing, 43, pp.436–442, 2005.

[197] B.B. Mandelbrot, “The fractal geometry of nature,” W. H. Freeman and Co., 1982.

[198] S. Cross, A. McDonagh, T. Stephenson, D. Cotton, and J. Underwood, “Fractal and integer-dimensional geometric analysis of pigmented skin lesions,” Am J Dermatopathol, 17, pp.374-378, 1995.

[199] K. Falconer, “The fractal geometry of nature,” Chicester: John Wiley, 1990.

[200] S. Cross and D. Cotton, “The fractal dimension may be a useful morphometric discriminant in histopathology,” J Pathol, 166, pp.409–411, 1992.

[201] J. Feder, “Fractals,” J Pathol, 166, pp.409–411, 1992.

[202] A. Tahmasbi, F. Saki, and S.B. Shokouhi, “Classification of benign and malignant masses based on zernike moments,” J. Computers in Biology and Medicine, 41(8), pp.726–735, 2011.

[203] B. Oluleye, A. Leisa, J. Leng, and D. Dean, “Zernike moments and genetic algorithm: tutorial and application,” British Journal of Mathematics and Computer Science, 4(15), pp.2217–2236, 2014.

[204] T. Lee, D. McLean, and M. Stella, “Irregularity index: a new border irregularity measure for cutaneous melanocytic lesions,” Medical Image Analysis, 7(1), pp.47-64, 2003.

[205] T. Do, T. Hoang, V. Pomponiu, Y. Zhou, Z. Chen, N. Cheung, D. Koh, A. Tan, and S. Tan, “Accessible melanoma detection using smartphones and mobile image analysis,” IEEE Trans. Multimed, 20, pp.2849-2864, 2018.

[206] S. Pattanayak, “Pro deep learning with tensorFlow: a mathematical approach to advanced artificial intelligence in python, 1st edition,” Apress, 2017.

[207] K. O’Shea and R. Nash, “An introduction to convolutional neural networks,” arXiv:1511.08458 [cs.NE], 2015.

[208] I. Goodfellow, Y. Bengio, and A. Courville, “Deep learning,” Cambridge, Massachusetts: MIT Press, 2016.

[209] A. Géron, “Hands-on machine learning with Scikit-Learn and Tensorflow,” O’Reilly, 2017.

[210] B.A. Cohn, “In consideration of the E in the melanoma ABCDE mnemonic,” Arch Dermatol, 142, pp.528-529, 2006.

[211] C.I.E., “Recommendations on uniform color spaces, color difference equations, psychometric color terms,” Supplement No.2 to CIE publication No.15 (E.-1.3.1) 1971/(TC-1.3.), 1978.

[212] R.S. Berns, “Billmeyer and Saltzman’s Principles of Color Technology,” Wiley-Interscience; 3rd edition, 2000.

[213] G.S. Sebestyen, “Decision making processes in pattern recognition,” New York: Macmillan, 1962.

[214] J. MacQueen, “Some methods of classification and analysis of multivariate observations,” In L. M. Le Cam & J. Neyman (Eds.), Proceedings of the fifth Berkeley symposium on mathematical statistics and probability. Berkeley, CA: University of California Press, 1, pp.281-297, 1967.

[215] P.J. Groenen, U. Kaymak, and J. van Rosmalen, “Fuzzy clustering with minkowski distance functions,” In Advances in Fuzzy Clustering and its Applications (eds J. Valente de Oliveira and W. Pedrycz), 2007.

[216] S. Butenko, W. Chaovalitwongse, and P. Pardalos, “Clustering challenges in biological networks,” World Scientific, 2009.

[217] S. Seidenari, A. Fabiano, S.A. Jalbout, S. Bassoli, S. Borsari, C. Magnoni, A. Tomasi, M. Vinceti, and G. Ponti, “Relationship between histological and computer-based assessment of melanoma diameter and thickness in head versus trunk-limbs melanomas,” Head Neck Oncol., 2013.

[218] L.R. Feret, “Assoc. Internat. pour l‘Essai des Mat.,” Zurich, 2, Group D., 1931.

[219] W.H. Walton, “Feret’s statistical diameter as a measure of particle size,” Nature, London, 162, pp.329-330, 1948.

[220] A. Ali, S.J. O’Shea, J. Li, and L. Roche, “A deep learning approach to deblurring skin cancer images,” British Association of Dermatologists, 99th Annual Meeting, Liverpool, UK, July 2-4 2019.

[221] A. Ali, S.J. O’Shea, and J. Li, “Are neural networks effective in detecting melanoma using genomic data?,” 16th Annual Joint GenoMel, BioGenoMel, MELGEN Scientific Meeting, Athens, Greece, April, 8-10, 2019.

[222] M. Mirza and S. Osindero, “Conditional generative adversarial nets,” arXiv preprint arXiv:1411.1784, 2014.

[223] S. Morris, B. Cox, and N. Bosanquet, "Cost of skin cancer in England," The European Journal of Health Economics, pp.267-273, vol.10, no.3, 2009.

[224] K. Volkovova1, D. Bilanicova1, A. Bartonova, S. Letašiová, and M. Dusinska1, "Associations between environmental factors and incidence of cutaneous melanoma. Review," Environ Health, 2012.

[225] Cancer Research UK, "Gut microbes linked to immunotherapy response in melanoma patients," https://www.cancerresearchuk.org/about-us/cancer-news/press-release/2016-11-07-gut-microbes-linked-to-immunotherapy-response-in-melanoma-patients," Accessed on December, 14 2020.

[226] R. Walton, J. Arnot, and P. Mackie, "Skin cancer in Scotland: What scope is there for further public health action?," Scottish Public Health Network (ScotPHN), 2015.

[227] World Cancer Research Fund, "Skin cancer statistics ," https://www.wcrf.org/dietandcancer/cancer-trends/skin-cancer-statistics, Accessed on December, 14 2020.

[228] J. Verne, V. Poirier, L. Hounsome, J. Kinge, L. Vallejo-Torres, S. Morris, "Cost of skin cancer in England, including projections to 2020," http://www.ncin.org.uk/view?rid=726, Accessed on December, 14 2020.

[229] Sakurra, "UVB and UVA explained. Filtering of rays sun protection. Penetration into the human skin <https://depositphotos.com/188016496/stock-illustration-uvb-uva-explained-filtering-rays.html>," created 15 Mar, 2018 (copyright purchased from depositphotos.com).

[230] G. Guy Jr, S. Machlin, D. Ekwueme, and K. Yabroff, "Prevalence and costs of skin cancer treatment in the U.S., 2002–2006 and 2007–2011," Am J Prev Med., vol.48, no.2, pp.183-187, 2015.

[231] The King's Fund, "How can dermatology services meet current and future patient needs, while ensuring quality of care is not compromised and access is equitable across

the UK?," https://www.bad.org.uk/shared/get-file.ashx?id=2348itemtype=document, Accessed on December, 16 2020.

[232] The Guardian, "It took my patient six years to see me – a dermatologist," https://www.theguardian.com/healthcare-network/views-from-the-nhs-frontline/2015/oct/19/patient-six-years-see-dermatologist-nhs-crisis, Accessed on December, 16 2020.

[233] Dermatology Deserts, "Skin cancer reports," https://www.reddingdesign.com/deserted/, Accessed on December, 16 2020.

[234] D. Morton, "Immune response to postsurgical adjuvant active immunotherapy with Canvaxin polyvalent cancer vaccine: correlations with clinical course of patients with metastatic melanoma," Dev Biol (Basel), 116:209-17, pp.229-36, 2004.

[235] Nhumnoi, "Dermatoscope with clipping path <https://depositphotos.com/99802720/stock-photo-dermatoscope-with-clipping-path.html>," created 18 Feb, 2016 (copyright purchased from depositphotos.com).

[236] The Guardian, "GPs fail to refer a third of malignant melanoma cases for urgent tests," https://www.theguardian.com/society/2016/sep/21/gps-fail-refer-one-in-three-people-malignant-melanoma-urgent-tests, Accessed on December, 16 2020.

[237] H. Ahammer, T.T. DeVaney, "The influence of edge detection algorithms on the estimation of the fractal dimension of binary digital images," Chaos, vol.14, no.183, 2004.

[238] C.-H. Wei, Y. Li, P.J. Huang, "Mammogram retrieval through machine learning within BI-RADS standards," J. Biomed. Inform., vol.44, no.4, pp.607–614, 2011.

[239] W.-Y. Kim, Y.-S. Kim, "A region-based shape descriptor using Zernike moments," Signal Processing: Image Communication, Vol.16, Issues 1–2, pp.95-102, 2000.