Interpretable Melanoma Detection for a Clinical Environment

Elliot Naylor

Submitted in accordance with the requirements for the degree of
Doctor of Philosophy

University of South Wales
Faculty of Mathematics and Computing

Date

The candidate confirms that the work submitted is his/her own and that appropriate credit has been given where reference has been made to the work of others.

Dedication here.

Acknowledgements

Abstract

Contents

1 Background **2**
1.1 Aim . . . 5
1.2 Objectives . . . 5
1.3 Contributions to knowledge . . . 5

2 Literature Review **8**
2.1 Introduction . . . 8
2.2 Discussion . . . 8
2.2.1 Datasets . . . 9
2.3 Previous Works on Automating ABCD Rules . . . 9
2.3.1 Segmentation . . . 9
2.3.2 Handcrafted Features . . . 10
2.4 Conclusion . . . 14

3 NHS Data Suitability for Melanoma Detection Using Machine Learning Algorithms **15**
3.1 Introduction . . . 15
3.2 Background . . . 15
3.3 Other Public Datasets . . . 16
3.4 Data Biases . . . 16
3.5 Image Criteria . . . 17

4 Data Preparation **19**
4.1 Introduction . . . 19
4.2 Related Works . . . 19
4.3 Skin Lesion Augmentation . . . 19
4.3.1 Hair Removal (Dull-Razor) . . . 19
4.3.2 Specular Removal . . . 20
4.4 Skin Lesion Segmentation . . . 20
4.4.1 SegNet a Semantic Pixel Wise Segmentation . . . 20
4.5 Border extraction . . . 21

4.5.1 U-Otsu Threshold 21
4.5.2 LBPC segmentation 22
4.5.3 Semantic Pixel-Wise Segmentation 23
4.6 Experimental results 24

5 Improved Melanoma Detection using Asymmetry 25
5.1 Introduction 25
5.2 Related Works 25
5.3 Asymmetry Analysis Approach 26
5.4 Results 27

6 Dermoscopic features extraction 30
6.1 30

7 Combined ABCD Rules and Dermoscopic Structures using Bayesian Network 31

8 Conclusion 32

9 Future Work 34

10 Tables 40

11 Appendix 41

List of Figures

2.1 Images of two skin lesions from the PH^2 dataset showing the asymmetry calculated from moments. 11
2.2 Images of two skin lesions split into 8 sections using moments, each border is measured for irregularity. 12

4.1 Demonstrating the Semantic Pixel-Wise Segmentation (SegNet) results showing the a) original image, b) expert ground-truth and c) SegNet results. . . 21
4.2 Otsu thresholding alongside ground-truth mask, where grey Otsu and white is SegNet. The bar chart shows the histogram with an otsu threshold of 138. 22
4.3 Local Binary Pattern Clustering (LBPC) showing the a) original image, b) ground-truth, and c) LBPC. LBPC successfully exaggerates the border cut-off on the skin lesions with regular and irregular borders 23
4.4 Demonstrating the Semantic Pixel-Wise Segmentation (SegNet) results showing the a) original image, b) expert ground-truth and c) SegNet results. . . 24

5.1 This diagram shows the skin lesion split relating to superpixels instead of averaging squares. 28
5.2 This diagram shows the difference between averaging squares and using superpixels, with the threshold of 10 implying curves and 50 being square. The horizontal colour difference is improved, making it more likely to be seen asymmetrical. The vertical comparison is roughly the same, except for removing a false positive of 40. 29

List of Tables

1.1 Total dermoscopy score (TDS) is a scoring system used with ABCD rules to support clinicians when diagnosing melanoma [11]. Each rule is multiplied by weights and the sum of the combined values is the final score, all together: [(Asymmetry × 1.3) + (Border × 0.1) + (Colour × 0.5) + (Dermoscopic structure × 0.5)]. 3

Chapter 1

Background

Skin cancer is considered amongst the most severe public health concerns, with mortality rates of 2,353 per 100,000 within the United Kingdom (UK) in 2018 [44]. Skin cancers can be categorised between melanoma and non-melanoma, whereas melanoma is the most dangerous because it is unpredictable. When left untreated and after growing sufficiently, it can spread to other regions of the body (known as metastatic melanoma), which once progressed is challenging to treat effectively with a 10% survival over ten years in the US [9]. Furthermore, it is beneficial to catch melanoma early because it is the most easily treatable form of cancer, with 86% of cases being preventable [44]. However, melanoma can remain dormant from anywhere between 6 months to 10 years before maturing and becoming a danger to the patient [44]. Another danger of melanoma is its similarity to non-melanoma skin cancers, such as a mimic called seborrhoeic keratosis (SK), which frequently leads to misdiagnoses [17]. There are features unique to SK called fissures, ridges, and hairpin vessels [24]. Problematically these features require trained specialists to recognise them needing more than ten years of experience to have an accuracy of 86% compared to 62% or 56% (3-5 years of experience) [25]. However, because of the cost of training new doctors, there are limited available. Dermatologists primarily treat skin conditions (biopsies) and confirm diagnoses submitted by GP. General practitioners (GP) are the first to diagnose skin conditions and sometimes have limited experience diagnosing them (especially dermatological features). This project aims to improve the accuracy of GP observations by providing tools for the automatic classification of skin lesions. An automatic system should be cost-effective and advantageous to doctors.

Diagnostic procedures are instructions developed by doctors to simplify diagnosing conditions. Various methods have been developed to diagnose skin lesions and have greatly improved GP accuracy within clinical environments [27, 45]. Considering melanoma is the most dangerous skin condition, most procedures were developed specifically for early detection. Some include ABCD rules, 2-point checklist, 7-point checklist, and CASH. The most preferred of these techniques are ABCD rules and CASH because they have a higher

Criteria	Methodology	Score	Weight
Asymmetry	Measuring asymmetry involves first finding the centroid and splitting it twice with a 90-degree axis. Each side is subtracted with its opposite half to measure the asymmetry of shape, colour, and dermoscopic structures. If both sides are asymmetrical then the score is 2, one side asymmetrical is a score of 1, and otherwise, the score is 0.	0 - 2	× 1.3
Border	border is found by finding the centroid and drawing lines through it with a 45-degree angle, splitting the skin lesion into eight segments. Border segments might be irregular with convexity, sharp corners, or edges. Irregular segments are incremented by 1, reaching 8 for each segment.	0 - 8	× 1.3
Colour	The area of the skin lesion is up to 6 colours (white, red, light brown, dark brown, blue-grey, black). The score is increased by 1 for each visible colour, reaching a total of 6.	1 - 6	× 0.5
Dermoscopic Structures	Dermoscopic structures are measured by finding structureless areas, pigment networks, atypical networks, dots, and globules. Each visible structure adds a score of 1, reaching a total of 5.	1 - 5	× 0.5

Table 1.1: Total dermoscopy score (TDS) is a scoring system used with ABCD rules to support clinicians when diagnosing melanoma [11]. Each rule is multiplied by weights and the sum of the combined values is the final score, all together: [(Asymmetry × 1.3) + (Border × 0.1) + (Colour × 0.5) + (Dermoscopic structure × 0.5)].

sensitivity [45] and ABCD rules are generally the most preferred because it is easy to learn and is rapidly calculated [27]. There are several variations of ABCD rules, but, is originally measured using asymmetry, border, colour, and diameter. Diameter is sometimes replaced with dermatological structures because many features (i.e. blue-black signs, pigment networks, pseudopods, streaks, or milia-like cysts [37]) improve the classification accuracy between melanoma and the mimic seborrheic keratosis [11]. Furthermore, automatically measuring diameter is often difficult because it is dependent on the photo apparatus and the distance from the skin lesion, which is rarely consistent in research. Table 1 describes each rule in more detail, including a scoring system called total dermoscopy score (TDS), where each rule is assigned a score and combined to reach a result of either malignant, suspicious or benign. The criteria is: [(Asymmetry × 1.3) + (Border × 0.1) + (Colour × 0.5) + (Dermoscopic structure × 0.5)]. Each rule is calculated using the following descriptions in Table 1 and multiplied by their weight and then added together to reach a final score where [< 4.76 = benign, > 4.76 or < 5.45 = suspicious, > 5.45 = melanoma]. The disadvantage is the subjectivity of GP observations relating to their experience. So, it would be beneficial to automate the techniques using algorithms to standardise results and improve GP accuracy.

Computer-aided diagnostic (CAD) frameworks are a collection of algorithms designed to guide decision-making processes within clinical environments [13]. A paper written by Andre

Estava demonstrates a deep convolutional neural network (DCNN) that has comparable accuracy to that of dermatologists, trained using 129,450 clinical images consisting of 2,032 different diseases [7]. DCNN generates a collection of artificial neurons organised into layers, where each neuron receives input from a previous layer to perform a computation. The collection of layers is a network, which (once trained) ultimately measures the relationship between input parameters based on provided data. It is important to note that the accuracy is proportionate to the number of images and data quality for training that network. Unfortunately, these image samples are frequently private and unavailable to many institutions. Without adequate image data to test the capabilities of machine learning models, there is no method for measuring these biases and is therefore unsafe to use within clinical environments. Secondly, these approaches will often produce a parallel diagnosis, meaning that results are not always explainable[20]. There are many valuable techniques, but even the best techniques are inadequate for doctors without catering to interpretability. Other techniques are interpretable by considering diagnostic procedures, such as ABCD rules, many of which are described by Ali [5]. Techniques based on diagnostic procedures can be more easily tested for biases and provide further insight to GPs with the means to learn from and understand results. Techniques include support vector machine (SVM), a supervised machine learning that uses regression analysis to categorise labelled data into two or more groups. The advantage means less data for training is required, and the model is interpretable.

Explainable AI (XAI) techniques have recently gained attention because the European general data protection regulation (GDPR and ISO/IEC 27001) stated that these approaches, commonly referred to as "black box" approaches, are challenging to utilise in medical environments. Since then, there has been significant progress in making neural network architectures more interpretable. A wide range of techniques [15, 34, 31] have since been developed, demonstrating that it is possible to make neural network techniques interpretable. However, the problem is instead the current scepticism on whether these techniques are trustworthy [43, 33], and they can produce realistic but incorrect results [16]. Some other interpretable techniques do not utilise neural networks. For example, Javier López-Labraca et al. [21] described an interpretable technique using multiple SVM models with colour and three dermoscopic structures (i.e., pigment networks, globules, and streaks). Bayesian fusion combines each model to calculate a diagnosis. Bayesian probability is a type of probability theory that uses probability distribution to estimate the values of unobserved variables. Bayesian fusion has a comparable accuracy to neural network techniques [40]. Overall, results should be partially interpretable for use within clinical environments.

Doctors will often only have access to a patient for a short time before moving to another. CAD frameworks are beneficial because they speed up the process, can improve accuracy [13], and ensure the gathering of relevant data (ABCD rules). Furthermore, it could take days for a second opinion from another doctor, where an automatic system immediately provides it. Automated systems should also provide adequate explanations that can be understood quickly and easily by doctors[20]. One method is to provide visual

explanations. Many authors [48, 18, 3] describe different ways to measure ABCD rules, including the asymmetry of skin lesions using bi-fold. Automated versions of the procedure use the centroid and moments of inertia to fold the skin lesion horizontally and vertically along the centroid. The overhung area on both axes is subtracted from the final score to measure asymmetrical or symmetrical. This technique produces an adequate visualisation that can provide GPs with an interpretable result. There is a range of other examples for ABCD rules, including border [18, 48, 4], colour [35, 41, 18], and dermoscopic structures [21] that use a range of interpretable algorithms that produce interpretable results.

Overall, many advanced machine learning techniques using neural networks lack the interpretability required within clinical environments. Furthermore, public datasets lack rarer skin conditions, making finding biases challenging. Automating the ABCD rules can solve this by using a technique that GPs are familiar with, and by using statistical models to extract relevant features (relating to the ABCD rules). This is followed by summarising rules using Bayesian fusion and calculating the significance of individual features.

1.1 Aim

- Develop an interpretable CAD framework based on the ABCD rules to diagnose skin lesions automatically. The goal is to utilise statistical models to extract each ABCD rule (asymmetry, border, colour, and Dermoscopic structure). Each rule will be trained using individual SVM models and are combined using Bayesian Fusion.

1.2 Objectives

- Develop and validate skin lesion segmentation and border cut-off approach for improved irregularity detection of ABCD rules using SegNet and LBPC.
- Develop and validate melanoma classification based on the diagnostic procedure ABCD rules (asymmetry, border, colour, and dermoscopic structures) for improved interpretability to doctors using various statistical techniques and SVM models.
- Develop and validate combining ABCD rules for the probabilistic analysis of the most dependent features using Bayesian fusion. This could include meta-data for gender, age, touch, feeling, and location on the body.

1.3 Contributions to knowledge

1. **Developing and validating a novel skin lesion segmentation approach for accurate border cut-off segmentation to improve border irregularity analysis using SegNet and LBPC.**

SegNet is highly accurate at finding the area for the segmentation of skin lesions but is inaccurate for measuring border irregularities because the border cut-off between skin and skin lesion is insufficient. Border irregularity detection necessitates an accurate cut-off for more reliable results, which SegNet does not provide. LBPC solves this problem by exaggerating the cut-off and improving the accuracy of border irregularity detection. However, the disadvantage of LBPC is its inaccuracy when finding the skin lesion area. By combining SegNet and LBPC, detecting the skin lesion area using SegNet, followed by adjusting the border with LBPC; retaining the accuracy of SegNet while improving the border cut-off accuracy. Experimental testing utilising the PH^2 dataset containing expert segmentation data will determine the benefits of segmentation.

2. **Developing and validating a novel asymmetry analysis approach for improved irregular asymmetry detection in skin lesions using moment-based texture analysis for improved bi-fold analysis and superpixels for improved asymmetry colour comparisons.**

   The disadvantage of asymmetry measuring techniques for skin lesions is rotational moments for creating bi-folds. Current bi-folds solely consider the silhouette of the skin lesion, with no consideration towards colour or texture. Furthermore, recent techniques have measured asymmetrical irregularities based on colour and texture. Producing a bi-fold based on the shape, colour, and texture using moment-based texture analysis should improve the accuracy of asymmetry detection. In addition, utilising superpixels to measure colour asymmetry to avoid merging important features improves accuracy. Both techniques will be validated using the PH^2 asymmetrical score.

3. **Developing and validating a novel interpretable melanoma classifier for improved interpretability of ABCD rules (asymmetry, border, colour, and dermoscopic structures) using feature extraction, support vector machines (SVM), and Bayesian fusion.**

   The disadvantage of many neural network-oriented techniques is their lack of adequate interpretability, making them challenging to utilise in clinical environments. However, ABCD rules (asymmetry, border, colour, and dermoscopic structures) are a diagnostic procedure that most doctors are familiar with; therefore, developing a system automating this procedure is beneficial. Feature extraction techniques aim to separate the data essential for each ABCD rule and train an SVM model from the extracted features. For example, bi-folds measure asymmetry, which can be modified to train an SVM model. Repeating this for border, colour, and dermoscopic structures ensures that each rule is independent. Finally, combining the Bayesian fusion results measures the probabilistic significance between ABCD rules and combines them into benign or malignant. Techniques will be validated using the PH^2 dataset for testing

ABCD rules and ISIC 2018 datasets for diagnosis.

4. **Developing and validating a novel interpretable melanoma classifier with meta-data including age, gender, feeling, and location on the body to improve classification accuracy between melanoma and seborrhoeic keratosis (SK) using Bayesian probability for a modifiable probabilistic analysis.**

   Seborrhoeic keratosis (SK) is a melanoma mimic because it sometimes shares clinical features with melanoma. Moreover, differentiating between the two with entirely image data can lead to inaccuracies. Including meta-data age, gender, feeling, and location on the body should improve accuracy because SK appears more frequently on the head or back of old male patients. Bayesian probability networks are considered highly modifiable and can generate results with incomplete input, meaning meta-data is only inputted when necessary, benefiting doctors and improving the diagnosis. The associated organisation has a vast amount of valuable meta-data alongside image data of skin lesions; a private dataset will be created from these results and used to validate results.

Chapter 2

Literature Review

2.1 Introduction

This chapter reviews statistical and machine learning algorithms for the automatic classification of ABCD rules and discussion on techniques beneficial for use within clinical environments.

2.2 Discussion

When doctors diagnose conditions using CAD, they should rationalize and build an explanation to prove their diagnosis, building criteria that other doctors understand. Currently, many techniques [7] called "black box" approaches produce parallel diagnosis that lacks an explanation. These are insufficient for use within some clinical environments for this reason. Instead, it would be beneficial for doctors to follow procedures they are familiar with, such as diagnostic procedures including ABCD rules. The reviewed techniques aim to automate the ABCD rules using various statistical and machine-learning techniques. Many are interpretable and suitable for clinical environments.

Hybrid machine learning techniques are recently gaining traction, an example by Ali combines results from both Gaussian naive Bayes (GNB) and a CNN [4] for border irregularity detection. The CNN ensures high-accuracy classification by finding the relationship between each component, and the GNB is interpretable. Results are combined using an ensemble approach, making a prediction probability. Such techniques are promising for use within clinical environments.

There is a lack of literature describing adequate visual representations for doctors, and it is understandable as there is still little evidence proving that CAD systems improve doctors decision making-processes [32]. It would be beneficial to create literature describing a catalogue of different visualisations that benefit doctors. Putting all this information together, alongside a questionnaire, might provide further insight into the visualisations

that might be most useful to doctors.

2.2.1 Datasets

One fundamental problem is the overutilisation of private or privately annotated datasets, making a direct comparison of algorithms difficult; hence, this review has no direct comparisons. Some are between benign and malignant [23, 18, 4, 3] while others utilise private or never mention any datasets [18, 35, 41, 29, 48]. None compare their rules, likely because of subjectivity depending on the dermatologists that labelled them. More datasets should be public to assess individual rules and reach objective measurements. Until then, results conform with malignant, suspicious, or benign.

2.3 Previous Works on Automating ABCD Rules

Many CAD frameworks follow a methodology for the classification of skin lesions. These are listed below:

1. Segmentation – Image segmentation is the process of partitioning an image into multiple segments for more accessible analysis. These areas can be separated manually by a dermatologist (known as the ground truth) or separated automatically using statistical or machine learning algorithms.

2. Feature Extraction - Gathering features through filtering, morphology and other statistical approaches. ABCD rules include asymmetry, border, colour, and dermoscopic structures.

3. Combination - Combining the extracted features before using Principal Component Analysis (PCA) or after classification using Bayesian Fusion. Others combine the results using the Total Dermoscopy Score (TDS).

4. Classification – Measuring the results from the features and components through classification. Containing the final diagnosis of the type of skin lesion (Naveus, SK, or Melanoma)

2.3.1 Segmentation

Yading Yuan and Yeh Chi Lo describe a fully convolutional network (FCN) with an accuracy of 91.7% with the PH^2 dataset [47]. FCN is a variation of a CNN using 1x1 convolutions instead of dense layers. Essentially, an FCN forms a more complex function (generating a more complex neural network), whereas the CNN forms a less complex function, likely to degrade essential features. Therefore, more data is needed to train an FCN effectively than a CNN. After the convolution layers, transposed convolution layers (or deconvolution) and

other layers (un-pooling) up-sample the input feature map to the size of the input image. Then, the network, trained from ground truth (human-generated segmentation mask) and the original images, can automatically generate segmentation masks based on textures and colours of the skin lesion provided. There are dozens of examples of this, such as SegNet [8], which is another transposed CNN not designed initially for skin lesions but is effective at segmenting skin lesions.

E. Meskini et al. proposed using Otsu binarisation - a threshold technique that is effective at locating the border of a skin lesion after segmenting using Segnet [23]. Researchers proposed that when analysing the skin lesion border using ABCD rules, the original SegNet methods were ineffective because the ground truth is subjective - ineffective at finding the border cut-off between the skin lesion and skin. While SegNet has a 91.7% with the PH^2 dataset, the data is not effective at finding the precise border cut-off required for accurate border classification using ABCD rules. Therefore, researchers proposed the Otsu threshold to find the skin lesion border after segmenting using SegNet. Fan proposes another technique that uses a saliency-based segmentation approach to capture the area, followed by an Otsu threshold [14] to find the border cut-off from the skin lesion with a precision of 96.78% validated using the PH^2 dataset.

Pedro M.M. Pereira et al. proposed local binary pattern clustering (LBPC) to exaggerate the border, producing accurate results when classifying ABCD rules than ground-truth borders in the PH^2 dataset [28]. Local binary patterns (LBP) are texture descriptors calculated by comparing the centre pixel (of each pixel in the grey scaled image) with the eight neighbouring pixels as 'i', and converting it to a binary using the equation: $[if centroid > neighbour_i = 0, otherwise = 1]$. These eight neighbouring values produce a binary of 01101100 (decimal of 108) and change the centroid to 108. Next, the described process repeats on each other pixel in the image. Finally, the newly filtered image subtracted from the original grey-scaled image creates a segmentation mask with an accurate border cut-off. Finally, Pereira describes classification methods using SVM or FNN presenting the extracted border with an accuracy of 79% and 77% (respectively) with the MED-NODE dataset.

2.3.2 Handcrafted Features

Handcrafted features are the extraction of particular features using statistical algorithms—the benefit of separating data into components is a more accessible breakdown, improving explainability. In addition, this might instantiate trust for use within a clinical environment and prove more helpful to doctors.

Asymmetry

Asymmetry can be measured using the bi-fold technique, which involves drawing a line down the middle of the skin lesion and comparing the two halves to confirm whether the

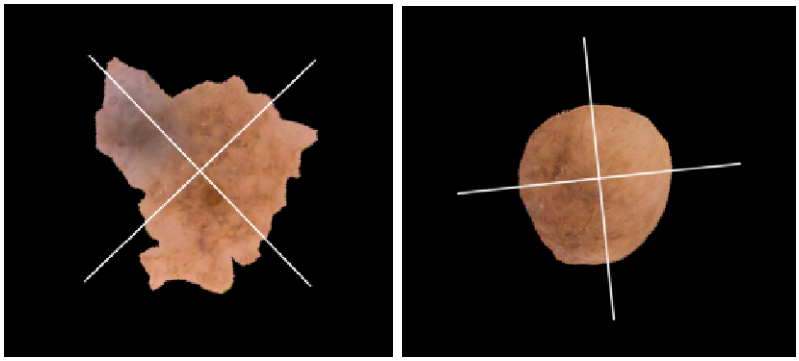

Figure 2.1: Images of two skin lesions from the PH^2 dataset showing the asymmetry calculated from moments.

sides match, on both the horizontal and vertical axes, as shown in 2.3.2. If the two sides are greatly different, it could be a warning sign of melanoma. Asymmetry can be measured using the shape [48], colour[18] and texture [3].

Measuring the asymmetrical shape requires a precise border cut-off. Ihab S. Zaqout [48] describes a technique using the centroid and rotation of the skin lesion using moments of inertia. By Folding the skin lesion on both vertical and horizontal axes subtracting the opposite half. Pixels that cannot subtract are summed and compared with a threshold considering the skin lesion asymmetrical if the combined sum is more than the threshold.

Reda Kasmi and Karim Mokrani [18] describe creating a grid of 20x20 pixels of the skin lesion image and converting it into the LAB colour space. Next, each block's average colour is compared with a perpendicular block (vertical and horizontal axes) using the three-dimensional Euclidean luminance distance, a-axis, and b-axis. If more than half of the colour comparisons are over the threshold, that axis is considered colour asymmetrical. Blocks that have no symmetrical pair are ignored. Finally, luminance calculated separately prevents brightness problems. This technique has an accuracy of 94% with a private dataset.

Measuring similarities in texture can be achieved by using SIFT-based similarity and projection profiles [3]. SIFT is scale-invariant and helpful for texture components with varying texture quality. First, the skin lesion is split vertically and horizontally across the centre into four halves, comparing texture components on the symmetrical halves and measuring the similarity. Lastly, the projection profile in the x and y directions generates histograms. These results train a decision tree and have an 80% accuracy of the ISIC 2018 with 204 images privately annotated for ABCD rules and combined.

Border

Estimating border irregularities involves splitting the skin lesion into eight equal sections (through the centroid), where each section with tight corners and convexity is considered

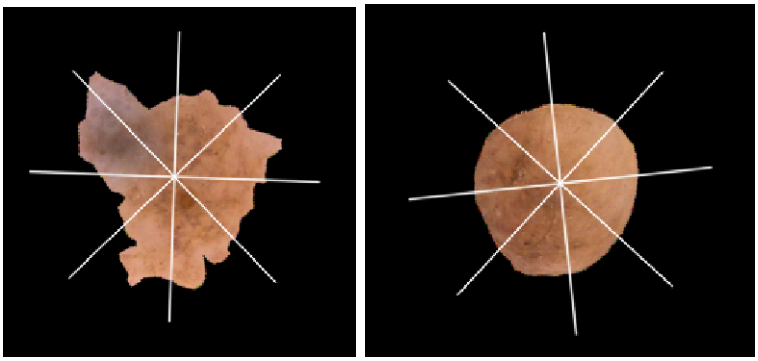

Figure 2.2: Images of two skin lesions split into 8 sections using moments, each border is measured for irregularity.

irregular. Each irregular section of the border adds a score of 1 ranging from 0 to a total of 8, as shown in figure 2.3.2.

Border irregularity contours were found by splitting the skin lesion into eight segments around the centre, and then calculating a fitting error for each. If the error is larger than 0.05 (x contour), that area is considered irregular [18].

Abder Rahman H. Ali et al. calculate the compactness of each border by first calculating the contour around the area of the lesion containing x and y positions. Next, measure the space between each position to estimate the compactness. The tighter the curves and corners, the more contour positions, revealing irregular borders within a segment, combining all of these scores creates the irregularity index [48].

Fractal dimensions (FDs) is a statistical index measuring the detail in a pattern changing with the image scale index. One technique called box-counting increases values if there are more corners and edges around the border. The higher the value demonstrates the level of border irregularity. Ali describes using machine learning alongside Zernike moments, and convexity measurements for a high-accuracy border irregularity classification [4]. However, results are ambiguous because the output is either “irregular” or “regular” border (not relating to the TDS). Thus, conforming to the TDS and splitting the border into eight sections would make it more interpretable and useful o doctors. However, a hybrid GNB and CNN approach are combined to allow interpretability through GNB.

Colour

Colour refers to the shades of pigment within the area of a skin lesion, not referring to abnormalities relating to bruises, crust, and grazes. Melanoma usually contains more than two colours compared with benign lesions, singular in colour. Skin lesions can consist of one or many colours: white, red, light brown, dark brown, blue-grey and black.

Finding colour variations has been achieved by calculating the normalised standard

deviation of the red, green, and blue components [35]. The normalisation process improves the recognition of normal skin pigmentation, which would show pigmentation levels, making comparisons easier between different skin lesions.

Arthur Tenenhaus, et al. utilise joint learning using Kohonen map, and k-means clustering [41]. Five random pixels create a 5x5 Kohonen map represented by 25 neurons in a neural network for each skin lesion in the dataset. Colour variations on a 25-dimensional vector find the proportions of pixels projected onto each of the 25 neurons. Next, K-means classifies the skin lesions set by the number of colours found by dermatologists. Only four colours were present in the dataset in this scenario, while seven could be. Eventually, the colour components are represented as a 42-dimensional vector and are passed into a KL-PLS based classifier to detect variations in colour at 66% using a private dataset.

Reda Kasmi et al. locate the number of colour variations by converting the image into the LAB colour space matching the colour ranges that can be perceived by human eyes [26], measuring the average colour distribution of the dataset and assigning each colour as a threshold range. Next, the Euclidean distance between each colour threshold is compared with each pixel colour [18], finding the closest matching colour of the six colours. Finally, removing the areas of colour with less than 5% prevents the classification of dots. This approach uses a colour range of white, light brown and dark brown. However, there is a static threshold value for the other colours, which would be unlikely to cover the ranges of the colours, including red, blue-grey, and black.

Dermoscopic structures

Dermoscopic structure refers to structures on the skin lesion, including pigment networks, structureless areas, dots, globules, streaks, white structures, and 22 others (not including sub-types). Variations of pigment networks are more commonly found in melanoma [6] and are therefore a valuable feature for automatic classification. Similarity other features such as milia-like cysts, a sub-type called milia-like cysts (MLC) called cloudy MLC appears more frequently on melanoma than SK, with a specificity of 99.1% specificity [37].

Javier López-Labraca et al. [21] describes a statistical approach to classifying melanoma using dermoscopic structures through Gabor filtering, support vector machines, and Bayesian fusion. This technique uses a form of soft segmentation to find the area of these dermoscopic features. Firstly the structures are located using Gabor filtering using different values to find fissures and globules. Each structure is then compared with a trained SVM model to check the similarity of the detected features. The results from the model are then combined using Bayesian fusion to reach a result of malignant or benign. Finally, training a CNN model alongside an SVM improves the retractability of dermoscopic structures; compared to a standalone CNN model.

Combining ABCD Rules

This section describes combining features from the ABCD rules into a classification between malignant, suspicious or benign after considering all clinical features. Again, meta-data and texture can potentially improve the results.

Maryam Ramezani et al. proposed a method to extract features from ABCD rules storing them in vectors and extracting the texture as a GLCM. First, these 187 features are shrunk to 13 using PCA [29]. Next, the data trains an SVM to classify skin lesions into benign or malignant with an accuracy of 82.2% on macroscopic images using a private dataset.

Other methods output TDS [48, 49], which combines them using: [(Asymmetry x 1.3) + (Border x 0.1) + (Colour x 0.5) + (Diameter x 0.5)]. A statistical model for each ABCD rule outputs a score in the same format. The benefit is interpretability because it follows the diagnostic procedure. The technique achieved an accuracy of 90% using a private dataset.

2.4 Conclusion

Many techniques utilise ABCD rules to produce an automatic and interpretable diagnosis. Interestingly, many focus on detecting and classifying asymmetry, border, and colour (ABC) or dermoscopic structures, but neither combine the whole ABCD rules into a single framework. Despite dermoscopic structures providing a means of diagnosing problematic forms of melanoma, including mimics (seborrhoeic keratosis) [17], and non-pigmented melanomas. Thus, it would be valuable to combine both into a single system for possibly higher accuracy.

Despite various valuable features, asymmetry rarely utilises techniques other than statistical models. For example, researchers highly focused on border irregularity and dermoscopic structures, leading to hybrid machine-learning models for their assessment. However, asymmetry still utilises statistical approaches to measure and combine shape, colour, and texture. It would be beneficial to transform this data and process it using an SVM, improving accuracy.

Utilising external data, including feeling, touch, age, and location on the body, are helpful to doctors when diagnosing skin conditions, but is not mentioned in any of the discussed techniques. It would be beneficial to implement this data into the decision-making process.

Chapter 3

NHS Data Suitability for Melanoma Detection Using Machine Learning Algorithms

3.1 Introduction

The use of machine learning algorithms for the detection of melanoma is a promising and evolving field with detection accuracies often beating that of a dermotologists[7]. However, the effectiveness of such depends heavily on the quality of the datasets used to develop them[38]. The goal of this chapter is to describe the data extraction process from the National Health Service (NHS) and highlight biases, preprocessing, and other potential issues involved in the training of machine learning algorithms for the detection of melanoma.

3.2 Background

There are requirements for the project including the use of macroscopic images instead of dermoscopic images. Macroscopic is described as viewing with a naked eye or by taking a standard picture with no special lenses. Dermoscopic is when you use a specialized tool called a dermoscope.

Dermoscopy improves the diagnostic accuracy of dermatologists for melanoma when compared with macroscopic examination[46] and is widely considered superior [42]. Dermoscopic images provide a detailed visualization of patterns and structures on the surface of the skin lesion that might not be visible to the naked eye[42]. Some of these structures are pigment network, asymmetry, irregular borders, and other features that support in the differentiation between benign and malignant lesions[42].

Another example, shows the diagnosis for BCC was 91% when using dermoscopy, compared to 57% when using close-up images [12]. Similarly, the sensitivity for SCC was

77% with dermoscopy, compared to 70% with close-up images [12]. These findings highlight the superior diagnostic performance of dermoscopy compared to macroscopic.

Dermoscopic examination is superior to macroscopic examination, however, the projects requirements specify using macroscopic. The logic behind this is that general practitioners are unlikely to recognize dermoscopic features, so there is no need to supply them with dermoscopes. This appears to be consistent with an authors findings showing that 92% of dermatologists correctly recognize at least four of size types of melanoma. In contrast, only 38% of non-dermatologists were able to recognize the same number of melanomas [38]. Therefore, 'the dataset' is created with macroscopic images for examination.

Considerations to clean the data removing hair and specular reflection to improve classification accuracy. This chapter will discuss the data transformation of NHS macroscopic images, including augmentation techniques to remove lighting, hair and other anomalous data from the images. All of which will support in improving the accuracy when classifying.

Ideally the metadata included would include Filename, Tags, Gender, DOB, Department, Consent, Diagnosis, Date Photographed. However, there has been difficulty accessing this data, and therefore only the diagnosis of the skin lesions are available.

The goal of this chapter is to measure the usefulness of NHS data, highlighting potential issues or biases in the data.

3.3 Other Public Datasets

Overall there are a range of public datasets including MEDLINE, PH2, ISIC, and others making a total of 21 open access datasets containing 106,950 skin lesion images[**Wen2022**]. Out of these datasets only the PH2 dataset has publicly accessible data on the ABCD rules with a total of 200 images.

3.4 Data Biases

The use of datasets is fundamental to the development and evaluation of machine learning algorithms, and the accuracy and effectiveness heavily weigh on the quality of the data used. Biases can arise from data collection procedures and pre-processing techniques. Not considering possible biases greatly affects machine learning algorithms using them and their effectiveness. Furthermore, careful consideration is essential to ensure the accuracy and reliability of the conclusions proposed in this document. Failure to consider all these factors could result in skewered conclusions that could undermine the validity of findings. For these reasons, it is essential to carefully identify and evaluate data before using and testing it.

The NHS datasets contain a wealth of information that can be utilized. However, some biases need consideration before creating a dataset. These biases include:

1. The diagnostic procedure dismisses skin lesions without recognizably suspicious features and do not reach the phase that photographs were captured. As such, there is a

lack of typical benign skin lesions within the dataset, and most have some undesirable features.

2. Dermatologists have diagnosed the large majority of skin lesions which have varying accuracy depending on their experience. There is no way of knowing how accurate this data is.

3. Dermatologists could diagnose during an in-person examination where patients can be asked questions in real-time and further tests can be made involving touch. Otherwise, dermatologists diagnose using previously saved images, which might be less accurate because it is lacking the insight that an in-person examination would provide.

4. Some skin lesions within the dataset are lacking metadata including their diagnosis. Such image samples should be avoided.

5. Diagnoses of skin lesions are written in plain text including questions marks where there is some uncertainty and the possibility of multiple diagnoses. Only diagnoses that are certain of their findings are used.

6. Photographs of the skin lesions may be captured on different body parts such as hands, legs, face, and others. Most pre-processing methods are designed to differentiate between skin and skin lesions, so it is important to avoid using these images. Otherwise, new pre-processing methods will have to be made and tested.

7. Seborrhoeic keratosis (SK) have similar features to that seen in malignant skin lesions. Therefore, there might be skin lesions diagnosed as melanoma that are SK. Furthermore, because of its similarity there are potentially many SK images. It will be vital to separate these.

Varied accuracy of

3.5 Image Criteria

The dataset includes skin lesions of Malignant Melanoma (MM), Seborrhoeic keratosis (SK), Atypical Naevi (AN), Typical Naevi (TN), Squamous Cell Carcinoma (SCC), Basal Cell Carcinoma (BCC).

Considering the data biases described in the previous section we can begin to piece together criteria that describe which images are appropriate for the creation of the 'dataset'. Firstly, the most glaring problems are the diagnoses, because some images are missing diagnostic data and some have a variety of diagnoses, the scenario in which the diagnoses were captured is not mentioned, SK is potentially misdiagnosed as malignant, and the dermatologist that originally diagnosed are not mentioned. There are many ways to combat these problem. For example during the creation of the PH2 dataset[22], several

dermatologists were asked to diagnose a skin lesion with appropriate metadata relating to the ABCD rules. If most dermatologists agree it is included in the study, otherwise it is removed. The goal of this is to minimize the incorrect data within the dataset. Alternatively, you might argue that removing less adequate records fails to prove whether the developed algorithms work in a real medical environment. Therefore, other methods were explored to validate the data.

In this project, we do not have the resources to re-diagnose skin lesions, so instead the focus is on gathering diagnostic results from the histopathology department. These results are more accurate than dermatologists[25] and counter the mentioned biases. Comparing this data with the diagnoses from the dermatology department provides insight into dermatology accuracy. Furthermore, histopathology serves as a ground-truth for training the algorithms and testing their capabilities.

As mentioned in the data biases section the skin lesion images are taken under various different conditions including angles, lighting, and distance from the skin lesion. While the variety of conditions will decrease the accuracy of results and hinder the detection of dermoscopic features, it is a requirement of the project.

These variations in image quality would decrease the accuracy of results. For this study the focus is on skin lesions captured using dermoscopes, avoiding the cases where skin lesions are captured under less-adequte conditions.

Use ANOVA to check whether features of Atypical and typical moles are similar.

Analyse the data and make a point about the method used for labelling and how it might be improved

Chapter 4

Data Preparation

4.1 Introduction

This section discusses segmentation techniques producing statistically significant border cut-off at the perimeter of the skin lesion. An accurate border cut-off is an essential criterion for melanoma detection [28, 19] using ABCD rules. Unfortunately, segmentation continues to be a challenging task because datasets regularly contain an estimated border and sometimes an inaccurate border cut-off.

4.2 Related Works

4.3 Skin Lesion Augmentation

Skin lesion augmentation is especially vital because of the use of macroscopic images instead of dermoscopic images. This means there are various artefacts including hair, specular reflections, rulers, varying sizes, and shapes of the skin lesion. All of these can obscure the skin lesion and affect the accuracy of segmentation[**Unver2019**] and in effect feature detection.

By augmenting the skin lesion images using specular reflection removal and hair removal, the accuracy of feature classification methods can be improved

4.3.1 Hair Removal (Dull-Razor)

Dull-Razor is an algorithm developed by Lee et al[**Lee1997**]. and is frequently implemented with

4.3.2 Specular Removal

4.4 Skin Lesion Segmentation

Segmentation plays a crucial role in melanoma detection because it separates melanoma from healthy tissue. Accurate segmentation is essential for various aspects of melanoma diagnosis and treatment, and classification[2].

One of the main challenges in melanoma detection is the visual similarity normal and infected regions. Others are the presence of artefacts such as bubbles, hair and clinical marks[2]. These factors lead to low accuracy rates in traditional approaches. However, segmentation techniques can help overcome these challenges by removing these areas and isolating the melanoma from the rest of the image.

A range of traditional segmentation techniques including SegNet, Unet methods have been shown to outperform other approaches in capturing the most significant melanoma characteristics. However, these techniques do not provide an effective border for the analysis of ABCD rules. Other techniques have been explored including active countouring-based segmenation[30], LBPC and others for border adjustment include u-otsu and edge-imfill.

4.4.1 SegNet a Semantic Pixel Wise Segmentation

SegNet is a deep learning architecture that is used for semantic image segmentation for melanoma detection. It was originally developed by[10] and has shown promising results in various segmenation tasks.

The idea of SegNet is to perform pixel-wise classification by assigning each pixel in an image to a specific class or category. This is achieved through a fully convolutional neural network (FCN) architecture, which allows for end to end learning and inference at the pixel level.

The performance of SegNet has been evaluated using various datasets including the PACAL VOC-2012 semantic image segmentation task, where it achieved state-of-the-art results of the mean intersection over union (mIOU) of 79.7% on the test dataset.

Semantic pixel-wise segmentation (SegNet) is a machine learning architecture utilizing a deep, fully convolutional neural network (DCNN). This network requires training from ground-truth and pre-segmented images for automatic segmentation. SegNet consists of encoding layers, decoding layers, and a pixel-wise classification layer. The encoder layers consist of 3x3 convolutions (including batch normalization and ReLU), pre-trained filters for classifying features. After some convolutions, the data is down-sampled using a 2x2 pooling layer. Next, decoding layers consist of up-sampling, followed by 3x3 convolutions. Finally, the pixel-wise classification uses a softmax layer to represent each pixel between 0 and 1 based on the previous layers, generating a segmentation mask.

Results in figure 4.1 are generated from the architecture using the ISIC 2018 dataset split into 80% training and 20% validation images. The accuracy of locating the lesions

is 85%. However, figure 4.1 represents the border cut-off between skin and skin lesion is accurate to the dataset but inadequate for using the ABCD rules. Finding the border cut-off is vital for measuring ABCD rules [28].

4.5 Border extraction

Box-counting method

4.5.1 U-Otsu Threshold

Otsu threshold is a versatile automatic image thresholding technique meant to separate each pixel between two classes of foreground or background. One of the benefits of this method is that it does not require any training data. The equation 4.1 (within-class variance) describes splitting weights of $w_0(t), w_1(t)$, which are the probabilities divided by the threshold t, between 0 to 255. Furthermore, σ_1^2 and σ_0^2 are variances of these two classes. The class probability w is computed from the histogram in figure 4.2, which is an intensity histogram describing the colour distribution in an image. Measuring the values above and below the generated thresholds splits the image into two classes.

$$\sigma_w^2(t) = w_0(t)\sigma_1^2(t) + w_1(t)\sigma_2^2(t) \tag{4.1}$$

The histogram is split into two segments with the threshold t of 138 and the corresponding pixel locations to the histogram segment the skin lesion into two classes. Image morphology closing is applied to fill gaps that the threshold missed. On other occasions, the segmentation missed the skin lesion because of a similar colour between the skin and skin lesion. It might be beneficial to combine otsu with SegNet to improve its accuracy while producing a border cut-off. Figure 4.2 describes the difference between otsu and SegNet.

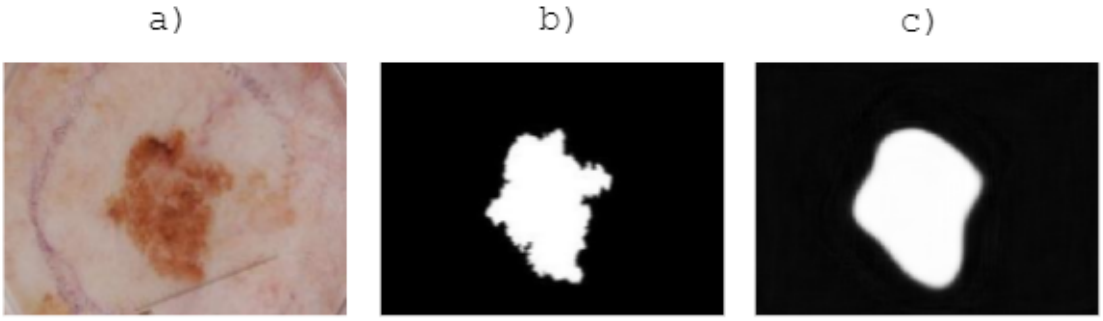

Figure 4.1: Demonstrating the Semantic Pixel-Wise Segmentation (SegNet) results showing the a) original image, b) expert ground-truth and c) SegNet results.

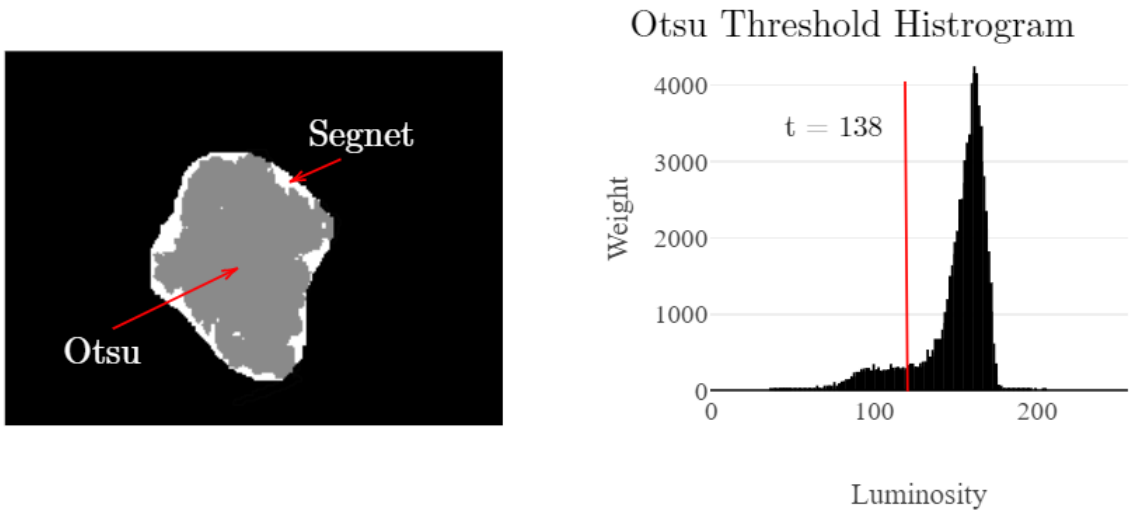

Figure 4.2: Otsu thresholding alongside ground-truth mask, where grey Otsu and white is SegNet. The bar chart shows the histogram with an otsu threshold of 138.

4.5.2 LBPC segmentation

Local Binary Patterns (LBP) is a texture descriptor commonly used for augmenting the image improving classification accuracy [28, 19]. First, equation 4.2 calculates each pixel, where p (equal to 8) is the number of neighbouring pixels compared to the centre of c, and radius of r from the centre. Next, shown in equation 4.3 each value is subtracted counter-clockwise with the centre value and compared to function S where each $gp - gc$, if more than or equal to 0, is equal to 1, and less than 0 is equal to 0. Next, add corresponding values equal to 1 of gp together, changing the centre value, ignoring values of 0. Next, applying a Gaussian kernel of 13-pixel iterations and a standard deviation of 3 removes smaller features that interfere with the segmentation. Finally, applying k-means with a value of 2 subtracts the greyscale and segments the skin lesion from the skin.

$$LBP(gp_x, gp_y) = \sum_{p=0}^{P-1} s\big(gp - gc\big)2^p \tag{4.2}$$

$$s\big(x\big) = \begin{cases} 1, & x \geq 0; \\ 0, & \text{otherwise.} \end{cases} \tag{4.3}$$

Figure 4.5.2 demonstrates the segmentation of two skin lesions, one with an irregular border and another with a regular border. LBPC is applied to both skin lesions, followed

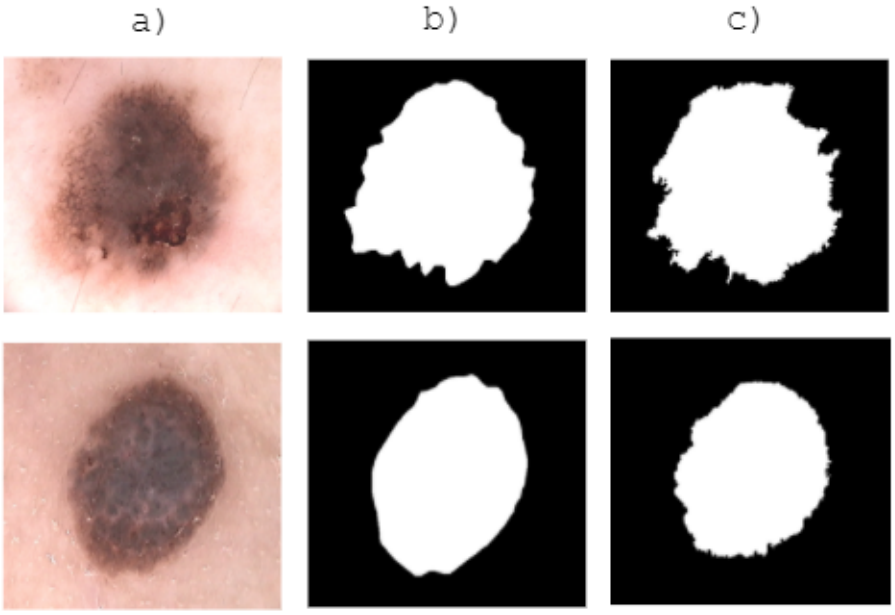

Figure 4.3: Local Binary Pattern Clustering (LBPC) showing the a) original image, b) ground-truth, and c) LBPC. LBPC successfully exaggerates the border cut-off on the skin lesions with regular and irregular borders

by Gaussian blurring and morphology closing to remove dots. The result is an improved border cut-off compared to the ground-truth in the Ph^2 dataset with more corners and ledges. This technique will improve accuracy for measuring border irregularity [28].

Validating LBPC is not expected because the goal is to exaggerate the border to improve the classification process of ABCD rules, which it does successfully [28, 19]. For example, the segmentation might not match dataset segmentations but is still essential to classifying ABCD rules. Furthermore, many datasets lack expert border segmentation, an accurate border cut-off between the skin and skin lesions, so comparisons are not always possible.

4.5.3 Semantic Pixel-Wise Segmentation

Semantic pixel-wise segmentation (SegNet) is a machine learning architecture utilising a deep, fully convolutional neural network (DCNN). This network requires training from ground-truth and pre-segmented images for automatic segmentation. SegNet consists of encoding layers, decoding layers, and a pixel-wise classification layer. The encoder layers consist of 3x3 convolutions (including batch normalisation and ReLU), pre-trained filters for classifying features. After some convolutions, the data is down-sampled using a 2x2 pooling layer. Next, decoding layers consist of up-sampling, followed by 3x3 convolutions.

Finally, the pixel-wise classification uses a softmax layer to represent each pixel between 0 and 1 based on the previous layers, generating a segmentation mask.

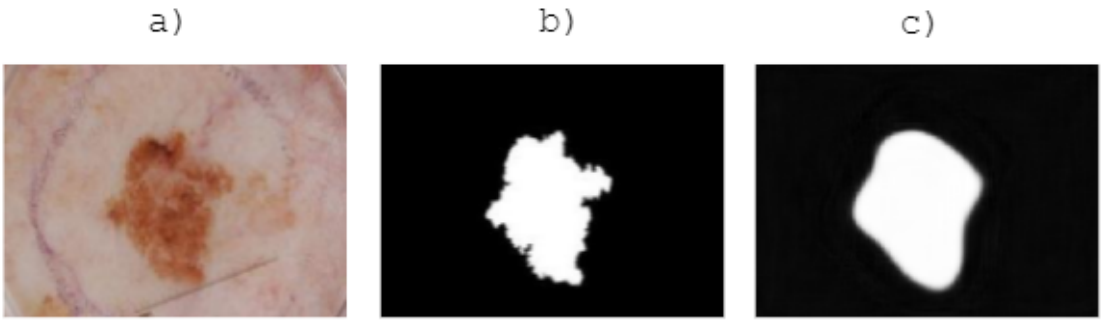

Figure 4.4: Demonstrating the Semantic Pixel-Wise Segmentation (SegNet) results showing the a) original image, b) expert ground-truth and c) SegNet results.

Results in figure 4.4 are generated from the architecture using the ISIC 2018 dataset split into 80% training and 20% validation images. The accuracy of locating the lesions is 85%. However, figure 4.4 represents the border cut-off between skin and skin lesion is accurate to the dataset but inadequate for using the ABCD rules. Finding the border cut-off is vital for measuring ABCD rules [28].

4.6 Experimental results

Both statistical models of LBPC and Otsu threshold generated accurate border cut-off as compared to the machine learning approach SegNet. Measuring border cut-off exaggerates irregular borders successfully, making it helpful in detecting border irregularities and possibly with other ABCD rules. It might be beneficial to combine SegNet and LBPC by using SegNet to find the skin lesions' location, followed by adjusting the border cut-off using LBPC.

Chapter 5

Improved Melanoma Detection using Asymmetry

5.1 Introduction

Melanoma is a type of malignant skin cancer that accounts for a significant proportion of cancer-related deaths around the world. In 2018 there were approximately 2,353 per 100,000 deaths in the United Kingdom (UK)[44]. Early detection is critical for improving the diagnosis and survival of patients. However, existing approaches including clinical examinations and dermoscopy, have limitations in terms of accuracy and cost-effectiveness[39]. Machine learning approaches have beaten dermatologists in terms of accuracy[7]. However, these approaches lack explainability implementing such techniques difficult for clinical environments[14]. One concern is the production of realistic, but incorrect results[16]. Another is the use of parallel processes, which describes the creation of an answer with little to no explanation. In this paper, we propose a combined asymmetry approach using shape, colour, and texture analysis alongside a detailed comparison. The technique itself can be used in conjunction with ABCD rules (Asymmetry, border, colour, dermoscopic features).

5.2 Related Works

Asymmetry analysis is a fundamental component in the early detection of melanoma because it often exhibits asymmetric shapes[3]. Bi-fold is a diagnostic procedure designed to recognize melanoma by drawing a line down the middle of the skin lesion and comparing the two halves to confirm whether the sides match (considering the difference in shape, colour, and texture). Using this horizontally and vertically calculates whether the skin lesion is possibly malignant with a score between 0 and 2. Calculating with Total Dermoscopy Score (TDS) alongside the other ABCD rules including asymmetry, border, colour, and diameter calculates the likelihood of malignancy. Dermatologists frequently use bi-fold due

to its simplicity, but it can be subjective to the original observer and time-consuming when managing large numbers of skin lesions. Therefore, automating techniques is beneficial to clinicians and can improve the objectivity of results.

Ihab S. Zaqout [48] describes a technique using the centroid and rotation of the skin lesion using moments of inertia. By Folding the skin lesion on both vertical and horizontal axes subtracting the opposite half. Pixels that cannot subtract are summed and compared with a threshold considering the skin lesion asymmetrical if the combined sum is more than the threshold.

Kasmi and Mokrani [18] create a grid of 20 by 20 pixels from the skin lesion image and convert it into the LAB colour space. They then compare the average colour of each block with a perpendicular block (vertical and horizontal axes) using the three-dimensional Euclidean luminance distance, a-axis, and b-axis. If more than half of the colour comparisons exceed the threshold, they consider that axis to be colour asymmetrical. They ignore blocks that have no symmetrical pair. Finally, they calculate luminance separately to prevent brightness problems. This technique achieves an accuracy of 94% with a private dataset.

Ali [3] uses SIFT-based similarity and projection profiles to measure similarities in texture. SIFT is scale-invariant and helpful for texture components with varying texture quality. First, they split the skin lesion vertically and horizontally across the centre into four halves, compare texture components on the symmetrical halves, and measure similarity. Lastly, they generate histograms for the projection profile in the x and y directions. These results train a decision tree and achieve an 80% accuracy of the ISIC 2018 with 204 images privately annotated for ABCD rules.

Prior studies have introduced techniques that measure distinct aspects of asymmetry, such as Ihab S. Zaqout [48] measurement of shape, Kasmi and Mokrani [18] measurement of colour, and Ali [3] measurement of texture. The new approach seeks to combine the following approaches into a more comprehensive analysis of asymmetry that takes into account multiple features of the skin lesion. The proposed novel technique updates colour measurement to improve accuracy using superpixels and an SVM model.

5.3 Asymmetry Analysis Approach

The proposed method has many similarities to Kasmi and Mokrani [18] colour comparison technique, except it is updated to improve accuracy using a number of techniques. Superpixels are k-means colour extraction techniques to separate areas of an image into their associated areas of colour by applying a soft border around an edge. The premise is the original method of splitting the lesion into a 20 by 20 grid without any flexibility splits areas of interest in half, making comparisons less adequate. Using superpixels allows for a softer border which improves colour separation and accuracy.

To initiate the classification of skin lesions a technique called bi-fold is applied involving folding the skin lesion in half vertically and horizontally and a comparison of their respective

dimensions. While the original technique was designed only to assess the lesions' shape, it's been utilized to account for colour and texture as well. The centre and orientation are determined by calculating its moments, where the centre is (m10 / m00, m01 / m00) and phi is 0.5tan(2m11)/(m20 -m02).

Next, the lesion is partitioned into a 20 by 20 grid centred on the mentioned centre point, and the average of each region is computed. This is followed by finding the matching region on the perpendicular area from the centre of the skin lesion and comparing the colour distance between the two. Distance is measured using the LAB colour space and a 2d Euclidean distance of A and B, removing L (luminosity) to eliminate light variation. Once compared, all compared regions are obtained, and they are plotted onto a graph. If over half of the values are above a threshold of 8, then the lesion is asymmetrical.

The weakest point in the technique is the use of a threshold, if more than half of the subtracted skin lesions are more than 8 it is considered asymmetrical. To demonstrate why this method is not suitable, look at the graph below using the PH$\hat{2}$ dataset. Essentially, the symmetrical skin lesion has a smaller area and the asymmetrical lesion has a larger area, but both remain in the same zone and therefore splitting the data only using a threshold has poor results. Furthermore, there are a lot of outlying data and the threshold does not adjust according to these values. See the graph below:

To improve the accuracy of the algorithm some changes need to be made based on the previous statements. First will be superpixels and next is k-means.

Superpixels is an algorithm for grouping pixels into a region with similar values. This one uses simple linear iterative clustering (SLIC) algorithm[1] to do so. The goal of using this instead of averaging specific areas as shown in [18], is to segment areas of similar interest values. This is to prevent areas from being cut in half.

The image in ... demonstrates the usual average and the new averages based on superpixels and the changes in values. areas that are lighter in colour appear to have a lower value and darker appear darker.

Using the thresholding method for classification we can already see the accuracy has been improved with a threshold of ....

5.4 Results

The goal of this experiment is to improve the accuracy of the asymmetry bi-fold technique described by Ihab S. Zaqout et al. [48]. Initially, the skin lesion is split into a 10x10 grid and converted into the LAB colourspace. Next, a line is drawn through the middle horizontally and vertically. Measuring the euclidean distance from the centroid, locating the closest opposite patch of colour finds the parallel square. Subtracting the squares generates a score for each value, the closer to 0, the more similar the colour. These are then removed from the list to prevent them from being selected a second time. If half the results are over a specific threshold, it is considered asymmetrical in colour, otherwise considered symmetrical.

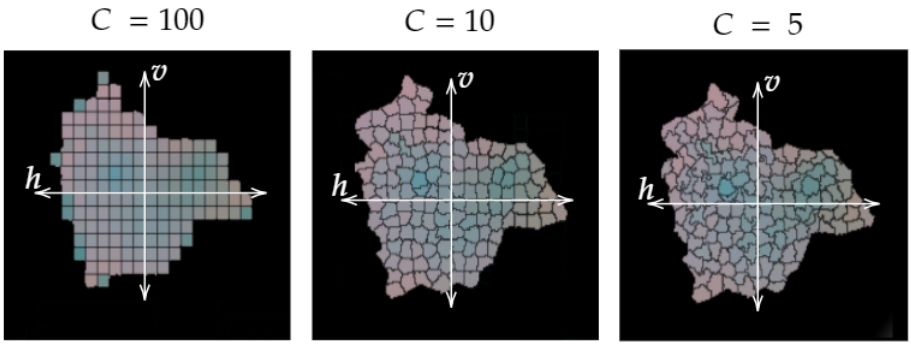

Figure 5.1: This diagram shows the skin lesion split relating to superpixels instead of averaging squares.

The aim is to make a 10x10 grid, but instead of averaging squares, superpixels reduce data redundancy in the grid, allowing for a less complex algorithm and improving accuracy. The clustering method k-means partition each pixel to its nearest most similar centroid relating to colour. Next, it generates a superpixel that represents the average colour of that area. The diagram 5.4 demonstrates different borders when changing the C for compactness, where 100 generates a square grid similar to the original technique. The border becomes more flexible as the compactness value decreases.

Each parallel square on the vertical and horizontal axes measures similarity using a three-dimensional euclidean distance in the LAB colour space. For example, the perceivable difference of colour to the human eye is a three-dimensional euclidean distance of 6 [26]. Using similar logic, a value of 20 is the threshold, where any value over that amount is considered asymmetrical in colour. Next, each square is compared with its closest parallel square and removed from an array after being compared. The next improvement is to generate a unique threshold for the significance of each square. For example, using superpixels with the compactness of 10 has an accuracy of 61% with the PH^2 dataset compared to the original 59.5%. This approach demonstrates that a flexible border that considers features is more effective than averaging squares.

There is a correlation in colour differences between the inner and outer edges because melanoma typically expands outwards, creating an abnormal border. This information specifies that the statistical model accuracy could be improved by increasing the threshold for the outer edges and decreasing for the inner.

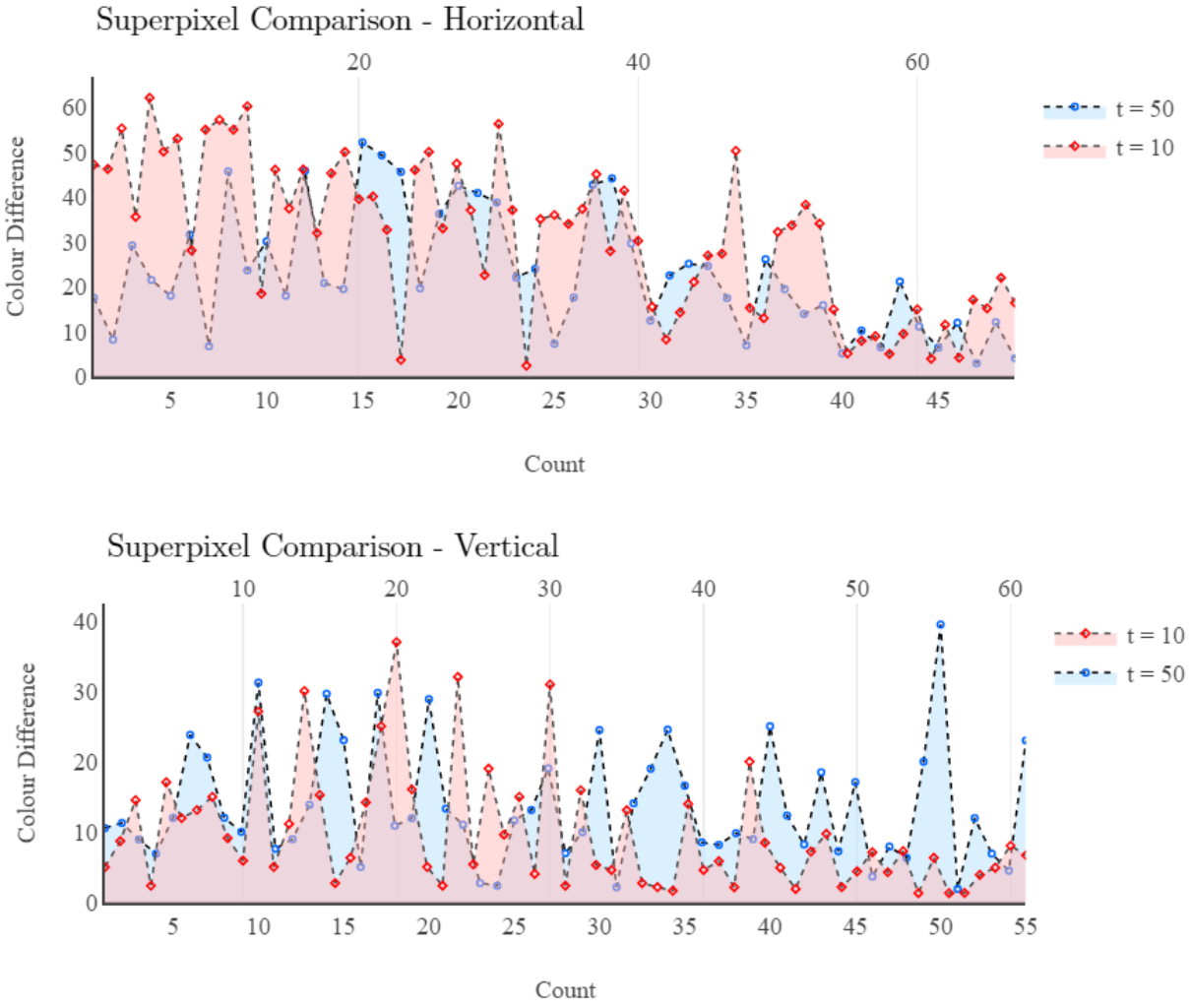

Figure 5.2: This diagram shows the difference between averaging squares and using superpixels, with the threshold of 10 implying curves and 50 being square. The horizontal colour difference is improved, making it more likely to be seen asymmetrical. The vertical comparison is roughly the same, except for removing a false positive of 40.

Chapter 6

Dermoscopic features extraction

6.1

Chapter 7

Combined ABCD Rules and Dermoscopic Structures using Bayesian Network

Chapter 8

Conclusion

Validating the automatic ABCD rules is challenging because public datasets are scarce and often lack sufficient data. For example, PH^2 contains 200 images on asymmetry, colour, and some dermoscopic structures but misses border irregularity. Therefore researchers aiming to measure borders use private or privately annotated datasets. Furthermore, many papers measuring asymmetry, colour and dermoscopic structures lack validation using public datasets despite PH^2 being available at the date of their publication. On the other hand, public datasets are crucial to comparing, validating, and reproducing algorithms. Therefore ABCD rules (apart from the border) will be validated using PH^2 datasets so that future researchers can replicate techniques. Furthermore, once rules are combined using Bayesian fusion, a type of probabilistic analysis, results can conform to the diagnosis between malignant and benign, validated from larger datasets, including ISIC 2019.

Finding the border cut-off is fundamental for the classification of melanoma using the ABCD rules [28]. Many valuable techniques use statistical models, including LBPC and Otsu, instead of transposed CNNs such as SegNet. Hybrid approaches using SegNet followed by Otsu to measure the border cut-off have been proven beneficial. However, using SegNet without a statistical model is worse when used with the ABCD rules than current methods such as LBPC and Otsu. Therefore, exploring other statistical segmentation techniques and hybrids would be beneficial. Furthermore, segmentation ground-truths do not always correspond to good classification accuracy with ABCD rules, which means even a low accuracy segmentation compared to datasets might have better accuracy when classifying the ABCD rules for border irregularity.

Statistical models for asymmetry, border, and colour extract relevant features for melanoma classification. The goal is to mimic the diagnostic procedure that clinicians are familiar with to produce results that they can utilise in a clinical environment. Extracting relevant features using box-counting and bi-folds ensures capturing relevant features and that the technique is retractable. However, accuracy is lacking in these techniques where superpixels improved asymmetry, changing the accuracy from 58.5% to 61% for the PH^2

dataset. Further improvements will be made after training an SVM model using the extracted features. Further implementation of convexity and Zernike moments for border irregularity will improve the accuracy. Furthermore, implementing a texture comparison for asymmetry measurements improve accuracy again.

Chapter 9

Future Work

Developing algorithms to extract features of ABCD rules is beneficial to GPs because it improves interpretability. Future work will involve extracting more features and training SVM models. For example, extracting more relevant asymmetry features will help classify asymmetry as there is currently no unification of shape, colour, and texture into a single classification model. The extracted features will be combined into a diagnosis between benign and malignant using a Bayesian probabilistic network. Bayesian probability is beneficial because its highly accurate [40] and modifiable and ability to classify with incomplete data. For example, asymmetry, border, and colour are sometimes enough to classify skin lesions. However, in some cases, dermoscopic structures or other meta-data, including age, gender, touch, feeling, and location on the body, are required for an accurate diagnosis. Furthermore, This might benefit GPs because it encourages considering a wide range of not always considered features.

Melanoma evolves from benign lesions at initially 30%-50%, and despite its significance, clinicians or computers are not yet able to reliably predict this change. AI trained on relevant images could predict melanoma before it occurs [36]. Data on skin lesion evolution is rare in public datasets. However, the associated organisation has taken images of the same skin lesion multiple times. It would be incredibly beneficial to assess the quality of these images, which could potentially lead to the development of a technique describing evolution. Considering evolution in machine learning techniques in the future would be incredibly beneficial to the early detection of melanoma but can only be achieved when there is more data.

Bibliography

[1] Radhakrishna Achanta et al. "SLIC superpixels compared to state-of-the-art superpixel methods". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 34 (11 2012). ISSN: 01628828. DOI: 10.1109/TPAMI.2012.120.

[2] Saleh Albahli et al. "Melanoma Lesion Detection and Segmentation Using YOLOv4-DarkNet and Active Contour". In: *IEEE Access* 8 (2020). ISSN: 21693536. DOI: 10.1109/ACCESS.2020.3035345.

[3] Abder Rahman Ali, Jingpeng Li, and Sally Jane O'Shea. "Towards the automatic detection of skin lesion shape asymmetry, color variegation and diameter in dermoscopic images". In: *PLoS ONE* 15 (6 2020). ISSN: 19326203. DOI: 10.1371/journal.pone.0234352.

[4] Abder-Rahman Ali et al. "A machine learning approach to automatic detection of irregularity in skin lesion border using dermoscopic images". In: *PeerJ Computer Science* 6 (2020), e268. ISSN: 2376-5992. DOI: 10.7717/peerj.cs.268/table-3.

[5] Abder Rahman H. Ali, Jingpeng Li, and Guang Yang. "Automating the ABCD Rule for Melanoma Detection: A Survey". In: *IEEE Access* 8 (2020), pp. 83333–83346. ISSN: 21693536. DOI: 10.1109/ACCESS.2020.2991034.

[6] Murali Anantha, Randy H. Moss, and William V. Stoecker. "Detection of pigment network in dermatoscopy images using texture analysis". In: *Computerized Medical Imaging and Graphics* 28 (5 2004). ISSN: 08956111. DOI: 10.1016/j.compmedimag.2004.04.002.

[7] Esteva Andre et al. "Dermatologist-level classification of skin cancer with deep neural networks". In: *Nature* 542 (7639 2017), pp. 115–118. ISSN: 1476-4687. DOI: 10.1038/nature21056LK-http://elinks.library.upenn.edu/sfx_local?sid=EMBASE&issn=14764687&id=doi:10.1038%2Fnature21056&atitle=Dermatologist-level+classification+of+skin+cancer+with+deep+neural+networks&stitle=Nature&title=Nature&volume=542&issue=7639&spage=115&epage=118&aulast=Esteva&aufirst=Andre&auinit=A.&aufull=Esteva+A.&coden=NATUA&isbn=&pages=115-118&date=2017&auinit1=A&auinitm=. URL: http://www.embase.

com/search/results?subaction=viewrecord&from=export&id=L614981551%0Ahttp://dx.doi.org/10.1038/nature21056.

[8] Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla. "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39 (12 2017), pp. 2481–2495. ISSN: 01628828. DOI: 10.1109/TPAMI.2016.2644615.

[9] Shailender Bhatia, Scott S. Tykodi, and John A. Thompson. *Treatment of metastatic melanoma: An overview*. 2009.

[10] Liang Chieh Chen et al. "DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 40 (4 2018). ISSN: 01628828. DOI: 10.1109/TPAMI.2017.2699184.

[11] Armand B. Cognetta et al. "The ABCD rule of dermatoscopy: High prospective value in the diagnosis of doubtful melanocytic skin lesions". In: *Journal of the American Academy of Dermatology* 30 (4 1994), pp. 551–559. ISSN: 01909622. DOI: 10.1016/S0190-9622(94)70061-3.

[12] A. Dascalu et al. "Non-melanoma skin cancer diagnosis: a comparison between dermoscopic and smartphone images by unified visual and sonification deep learning algorithms". In: *Journal of Cancer Research and Clinical Oncology* 148 (9 2022). ISSN: 14321335. DOI: 10.1007/s00432-021-03809-x.

[13] Vincent Dick et al. "Accuracy of Computer-Aided Diagnosis of Melanoma: A Meta-analysis". In: *JAMA Dermatology* 155 (11 2019), pp. 1291–1299. ISSN: 21686068. DOI: 10.1001/jamadermatol.2019.1375.

[14] Haidi Fan et al. "Automatic segmentation of dermoscopy images using saliency combined with Otsu threshold". In: *Computers in Biology and Medicine* 85 (2017), pp. 75–85. ISSN: 18790534. DOI: 10.1016/j.compbiomed.2017.03.025.

[15] Masaru Fuji et al. "Explainable AI through combination of deep tensor and knowledge graph". In: *Fujitsu Scientific and Technical Journal* 55 (2 2019), pp. 58–64. ISSN: 00162523.

[16] Amirata Ghorbani, Abubakar Abid, and James Zou. "Interpretation of Neural Networks Is Fragile". In: *Proceedings of the AAAI Conference on Artificial Intelligence* 33 (2019), pp. 3681–3688. ISSN: 2159-5399. DOI: 10.1609/aaai.v33i01.33013681.

[17] Leonid Izikson et al. "Prevalence of melanoma clinically resembling seborrheic keratosis: Analysis of 9204 cases". In: *Archives of Dermatology* 138 (12 2002). ISSN: 0003987X. DOI: 10.1001/archderm.138.12.1562.

[18] Reda Kasmi and Karim Mokrani. "Classification of malignant melanoma and benign skin lesions: Implementation of automatic ABCD rule". In: *IET Image Processing* 10 (6 2016), pp. 448–455. ISSN: 17519659. DOI: 10.1049/iet-ipr.2015.0385.

[19] Sertan Kaya et al. "Abrupt skin lesion border cutoff measurement for malignancy detection in dermoscopy images". In: *BMC Bioinformatics* 17 (2016). ISSN: 14712105. DOI: 10.1186/s12859-016-1221-4.

[20] Zachary C. Lipton. "The mythos of model interpretability". In: *Communications of the ACM* 61 (10 2018). ISSN: 15577317. DOI: 10.1145/3233231.

[21] Javier López-Labraca et al. "Enriched dermoscopic-structure-based cad system for melanoma diagnosis". In: *Multimedia Tools and Applications* 77 (10 2018), pp. 12171–12202. ISSN: 15737721. DOI: 10.1007/s11042-017-4879-3.

[22] Teresa Mendonca et al. "PH2 - A dermoscopic image database for research and benchmarking". In: 2013. DOI: 10.1109/EMBC.2013.6610779.

[23] E. Meskini et al. "A new algorithm for skin lesion border detection in dermoscopy images". In: *Journal of Biomedical Physics and Engineering* 8 (1 2018), pp. 109–118. ISSN: 22517200. DOI: 10.22086/jbpe.v0i0.444.

[24] Akane Minagawa. "Dermoscopy–pathology relationship in seborrheic keratosis". In: *Journal of Dermatology* 44 (5 2017), pp. 518–524. ISSN: 13468138. DOI: 10.1111/1346-8138.13657.

[25] C. A. Morton and R. M. Mackie. "Clinical accuracy of the diagnosis of cutaneous malignant melanoma". In: *British Journal of Dermatology* 138 (2 1998). ISSN: 00070963. DOI: 10.1046/j.1365-2133.1998.02075.x.

[26] Nikolaos E. Myridis. "Ultra-realistic Imaging: Advanced Techniques in Analogue and Digital Colour Holography, by Hans Bjelkhagen and David Brotherton-Ratcliffe". In: *Contemporary Physics* 55 (3 2014), pp. 247–248. ISSN: 0010-7514. DOI: 10.1080/00107514.2014.907348.

[27] Franz Nachbar et al. "The ABCD rule of dermatoscopy". In: *Journal of the American Academy of Dermatology* 30 (4 Apr. 1994), pp. 551–559. ISSN: 01909622. DOI: 10.1016/s0190-9622(94)70061-3.

[28] Pedro M.M. Pereira et al. "Skin lesion classification enhancement using border-line features – The melanoma vs nevus problem". In: *Biomedical Signal Processing and Control* 57 (2020). ISSN: 17468108. DOI: 10.1016/j.bspc.2019.101765.

[29] Maryam Ramezani, Alireza Karimian, and Payman Moallem. "Automatic Detection of Malignant Melanoma using Macroscopic Images". In: *Journal of Medical Signals and Sensors* 4 (4 2014), pp. 281–290. ISSN: 22287477. DOI: 10.4103/2228-7477.144052.

[30] Farhan Riaz et al. "Active Contours Based Segmentation and Lesion Periphery Analysis for Characterization of Skin Lesions in Dermoscopy Images". In: *IEEE Journal of Biomedical and Health Informatics* 23 (2 2019). ISSN: 21682208. DOI: 10.1109/JBHI.2018.2832455.

[31] Marco Ribeiro, Sameer Singh, and Carlos Guestrin. ""Why Should I Trust You?": Explaining the Predictions of Any Classifier". In: *ArXiv* (2016), pp. 97–101. DOI: 10.18653/v1/n16-3020.

[32] Lavinia Ferrante di Ruffano et al. "Computer-assisted diagnosis techniques (dermoscopy and spectroscopy-based) for diagnosing skin cancer in adults". In: *Cochrane Database of Systematic Reviews* (12 2018). ISSN: 1469493X. DOI: 10.1002/14651858.CD013186.

[33] Wojciech Samek and Klaus Robert Müller. "Towards Explainable Artificial Intelligence". In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 11700 LNCS (2019), pp. 5–22. ISSN: 16113349. DOI: 10.1007/978-3-030-28954-6_1.

[34] Ramprasaath R. Selvaraju et al. "Grad-cam: Why did you say that? visual explanations from deep networks via gradient-based localization". In: *Revista do Hospital das Cl??nicas* 17 (2016), pp. 331–336. ISSN: 00418781. URL: http://arxiv.org/abs/1610.02391.

[35] Zhishun She, Y. Liu, and A. Damatoa. "Combination of features from skin pattern and ABCD analysis for lesion classification". In: *Skin Research and Technology* 13 (1 2007), pp. 25–33. ISSN: 0909752X. DOI: 10.1111/j.1600-0846.2007.00181.x.

[36] Wiebke Sondermann et al. "Prediction of melanoma evolution in melanocytic nevi via artificial intelligence: A call for prospective data". In: *European Journal of Cancer* 119 (2019), pp. 30–34. ISSN: 18790852. DOI: 10.1016/j.ejca.2019.07.009.

[37] S. M. Stricklin et al. "Cloudy and starry milia-like cysts: How well do they distinguish seborrheic keratoses from malignant melanomas?" In: *Journal of the European Academy of Dermatology and Venereology* 25 (10 2011), pp. 1222–1224. ISSN: 09269959. DOI: 10.1111/j.1468-3083.2010.03920.x.

[38] Ki Hyun Tae et al. "Data Cleaning for Accurate, Fair, and Robust Models". In: 2019. DOI: 10.1145/3329486.3329493.

[39] Abdulrahman Takiddin et al. *Artificial intelligence for skin cancer detection: Scoping review*. 2021. DOI: 10.2196/22934.

[40] Maen Takruri and Abubakar Abubakar. "Bayesian decision fusion for enhancing melanoma recognition accuracy". In: *2017 International Conference on Electrical and Computing Technologies and Applications, ICECTA 2017* 2018-Janua (2017), pp. 1–4. DOI: 10.1109/ICECTA.2017.8252063.

[41] Arthur Tenenhaus et al. "Detection of melanoma from dermoscopic images of naevi acquired under uncontrolled conditions". In: *Skin Research and Technology* 16 (1 2010), pp. 85–97. ISSN: 0909752X. DOI: 10.1111/j.1600-0846.2009.00385.x.

[42] B.H. Thiers. "Dermoscopy compared with naked eye examination for the diagnosis of primary melanoma: a meta-analysis of studies performed in a clinical setting". In: *Yearbook of Dermatology and Dermatologic Surgery* 2009 (2009), pp. 378–379. ISSN: 00933619. DOI: 10.1016/s0093-3619(08)79154-8.

[43] Erico Tjoa and Cuntai Guan Fellow. "A Survey on Explainable Artificial Intelligence (XAI): Towards Medical XAI". In: *arXiv* (2019). ISSN: 23318422. DOI: 10.1109/tnnls.2020.3027314. URL: http://arxiv.org/abs/1907.07374.

[44] Cancer Research UK. *Melanoma skin cancer statistics — Cancer Research UK*. 2019. URL: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/melanoma-skin-cancer#heading-Three.

[45] Ezgi Unlu, Bengu N. Akay, and Cengizhan Erdem. "Comparison of dermatoscopic diagnostic algorithms based on calculation: The ABCD rule of dermatoscopy, the seven-point checklist, the three-point checklist and the CASH algorithm in dermatoscopic evaluation of melanocytic lesions". In: *Journal of Dermatology* 41 (7 2014), pp. 598–603. ISSN: 13468138. DOI: 10.1111/1346-8138.12491.

[46] Zachary J. Wolner et al. *Enhancing Skin Cancer Diagnosis with Dermoscopy*. 2017. DOI: 10.1016/j.det.2017.06.003.

[47] Yading Yuan and Yeh Chi Lo. "Improving Dermoscopic Image Segmentation With Enhanced Convolutional-Deconvolutional Networks". In: *IEEE Journal of Biomedical and Health Informatics* 23 (2 2019), pp. 519–526. ISSN: 21682194. DOI: 10.1109/JBHI.2017.2787487. URL: http://arxiv.org/abs/1703.05165%0Ahttp://dx.doi.org/10.1109/JBHI.2017.2787487.

[48] Ihab S. Zaqout. "Diagnosis of Skin Lesions Based on Dermoscopic Images Using Image Processing Techniques". In: *International Journal of Signal Processing, Image Processing and Pattern Recognition* 9 (9 2016), pp. 189–204. ISSN: 20054254. DOI: 10.14257/ijsip.2016.9.9.18.

[49] Yongfeng Zhang and Xu Chen. "Explainable Recommendation: A Survey and New Perspectives". In: (2018). URL: http://arxiv.org/abs/1804.11192.

()

Chapter 10

Tables

Chapter 11

Appendix