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.

This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

The right of Elliot Naylor to be identified as Author of this work has been asserted by him/her in accordance with the Copyright, Designs and Patents Act 1988.

©Year
University of South Wales
and
Candidate Elliot Naylor

Dedication here.

Acknowledgements

Abstract

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 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[10]. 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[18]. 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 to 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. For the previously mentioned reasons, an automatic system should be cost-effective and advantageous to doctors.

1.1.1 Diagnostic Procedures (ABCD Rules, CASH, 7-Point Checklist, Texture)

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 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[12]. 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 1.1.1 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 $\times$ 1.3) + (Border $\times$ 0.1) + (Colour $\times$ 0.5) + (Dermoscopic structure $\times$ 0.5)]. Each rule is calculated using the following descriptions in Table 1.1.1 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.

1.1.2 UI Development and visualisation

The user interface (UI) being the method in which algorithms are visualised is arguably the most import section of this project because without the means of presenting algorithms in an explaianable way they are of little to no use to healthcare professionals. Providing explanations will improve trust from doctors and patients to these algorithms[8]. This is also crucial for the successful implementation of AI technologies within healthcare settings[**skar2017**].

Computer-aided diagnostic (CAD) frameworks are a collection of algorithms designed to guide decision-making processes within clinical environments[14]. 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

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	$\times$ 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	$\times$ 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	$\times$ 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	$\times$ 0.5

Table 1.1: Total dermoscopy score (TDS) is a scoring system used with ABCD rules to support clinicians when diagnosing melanoma[12]. Each rule is multiplied by weights and the sum of the combined values is the final score, all together: [(Asymmetry $\times$ 1.3) + (Border $\times$ 0.1) + (Colour $\times$ 0.5) + (Dermoscopic structure $\times$ 0.5)].

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[21]. 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.

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[14], 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[21]. One method is to provide visual explanations. Many authors[48, 19, 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[19, 48, 4], colour[35, 41, 19], and dermoscopic structures[22] that use a range of interpretable algorithms that produce interpretable results.

1.1.3 Conclusion

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.

1.2 Purpose and Research Questions

In this thesis machine learning algorithms were developed for the automatic classification of skin lesions with consideration for its use within clinical environments. Furthermore, the technqiues need to produce an explainable and meaningful response that can support doctors (dermotologists and general practioners) by providing insights and details that might have been missed. Concerning this, two main research questions:

- Are the produced algorithms accurate enough to be properly utilised in a clincal environment?
- Does the developed technique provide meaningful responses that can be properly utilised within a clinical environemnt?

1.3 Scope and Limitations (Use-Case)

Some limitations were set by the company partner based on where they plan to utilise the developed algorithms. In this case the technique is primarly going to be utilised by GP (general practitioners) that are unlikely to have the specific training and equipment for diagnosing skin lesions.

- The algorithms are developed for the detection of Benign Naevi (BN), Melanoma (MM), Seborrhoeic keratosis (SK), Atypical Naevi (AN), Typical Naevi (TN), Squamous Cell Carcinoma (SCC), Basal Cell Carcinoma (BCC). With a primary focus on SK consdering its difficulty to diagnose.
- Macroscopic images will be analysed instead of dermoscopes. This essentially means a standard camera will be used, which includes variations in lighting which might obscure features in skin lesions.
- Meta-data includes area on body, gender, Date of Birth (DOB), department, Diagnosis. The goal is to utilise this data for automatic detection.
- Techniques must be explainable, so that doctors can recognise incorrect responses.

1.4 Target Group

The work is primarly for general practioners (GPs) because it is usually the source of misdiagnosed skin lesions due to the lack of specific training compared to dermatologists. The goal is to improve the accruacy of techniques using a cheap device for capturing data, furthermore automatic classification should provide a more adequate means for dermotologists to search for cases.

1.5 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.

1.6 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.

1.7 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

Analysis of Explainability for the Detection of Melanoma

2.1 Introduction

This chapter holds an analysis on the most popular explainable AI (XAI) techniques called DeepSHAP and Gradcam. Techniques will be compared to measure whether they produce suitable results for use within clinical environments.

2.2 Discssion

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 utilize in medical environments. Since then, there has been significant progress in making neural network architectures more interpretable. A wide range of techniques[16, 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 often results[17]. Some other interpretable techniques do not utilise neural networks. For example, Javier López-Labraca et al.[22] 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.

2.3 Exisiting techniques

DeepSHAP (Shapley Additive exPlanations) is a game theoretic approach designed to explain models during training by visualising features related to the classificaiton. It does this by explaining individual predictions in machine learning models, measuring the contribution of each feature to the contribution to an outcome[**Aas2021**].

In the field of healthcare, DeepSHAP is applied to predict and explain non-communicable diseases (NCDs). In explanations for individual predictions and a case study detecting the progression of Alzheimers.

Although explainable algorithms have seen some use within healthcare, there is no evidence of it's current use within dermatology.

Chapter 3

Literature Review

3.1 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.

3.2 Discussion

When doctors utilize a clinical diagnostic tool they should be capable of rationalizing and building explanations based on the data provided from that tool. Currently, many techniques[7] called named "black box" approaches produce parallel diagnosis that lacks adequate explanations for clinical environments. These provide insufficient information for use within some clinical environments[7]. 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.

3.3 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)

3.3.1 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[9], 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[15] 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.

3.3.2 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 sides match, on both the horizontal and vertical axes, as shown in 3.3.2. If the two sides are greatly different, it could be a warning sign of melanoma. Asymmetry can be measured using the shape[48], colour[19] 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[19] 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

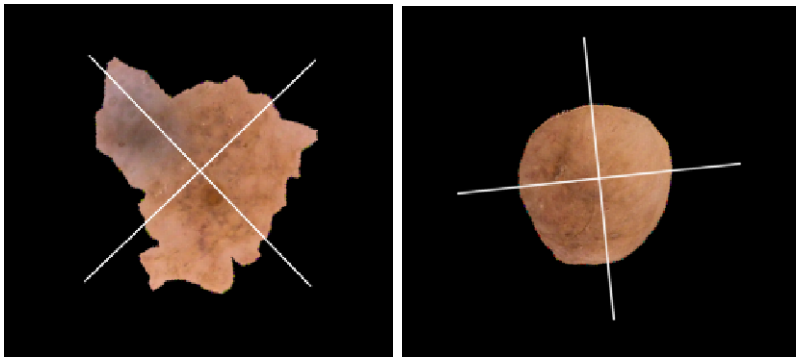

Figure 3.1: Images of two skin lesions from the PH^2 dataset showing the asymmetry calculated from moments.

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 irregular. Each irregular section of the border adds a score of 1 ranging from 0 to a total of 8, as shown in figure 3.3.2.

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[19].

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].

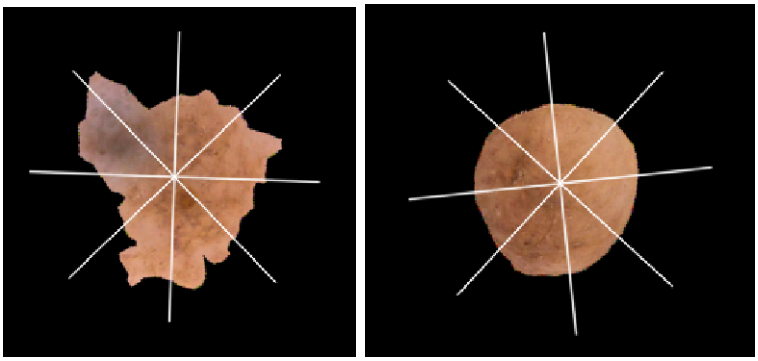

Figure 3.2: Images of two skin lesions split into 8 sections using moments, each border is measured for irregularity.

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 5 by 5 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[19], 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.[22] 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.

3.4 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)[18], 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 4

NHS Data Suitability for Melanoma Detection Using Machine Learning Algorithms

4.1 Introduction

This chapter contains an analysis of some popular datasets including ISIC 2019, and PH2. The goal is to identify any relationships between skin lesions and patients. Using this analysis 'the dataset' is created using NHS macroscopic images and metadata.

4.2 Other Datasets

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 ABCD rules with a total of 200 images. ISIC is the largest of these datasets, being a combination of many other datasets.

The datasets listed in table 4.2 include the number of images, classes, and metadata. Out of these datasets ISIC 2019, PH2 and 7-Point Criteria appear to be the most promising. PH2 is especially useful because it is the only dataset representing ABCD rules on asymmetry and colour. SKINL2 was considered, but ISIC 2019 was a much larger dataset with more metadata. SD-198 is publicly available but not accessible.

Although an ISIC 2020 dataset exists with a total of 44,108 images, its diagnosis is between benign and malignant and other metadata is on atypical melanocytic proliferation, café au lait macule, lentigo NOS, lichenoid keratosis, naevus, seborrhoeic keratosis, solar lentigo, and other/unknown. The metadata is very specific and doesn't match the requirements of the project, so ISIC 2019 is still a better candidate for analysis.

Name	Year	Image Type	Number	Classes	Metadata
ISIC 2019	2019	Dermoscopic	33,569	8	Age, anatomical site, gender, and diagnosis
PH2	2013	Dermoscopic	200	3	Asymmetry, colour, pigment network, dots/globules, streaks, regression areas, blue-whitish veil
MED-NODE	2015	Macroscopic	170	2	n/a
SD-198	2016	Dermoscopic & Macroscopic	6,584	198	anatomical site, symptoms, duration, morphology, and colour
SKINL2	2019	Macroscopic (unique tool)	376	8	Gender, age, and fototype
7-point Criteria	2018	Dermoscopic & Macroscopic	2000	2	Pigment network, regression, pigmentation, blue-whitish veil vascular structures, streaks, dots/globules

Table 4.1

Name	Year	Total Samples	Differences
HAM10000	2018	11,526	n/a
BCN 20000	2016	19,424	Includes nails and mucosa
MSK	2015, 2017	3918	Coloured stickers covering un-applicable skin lesions

Table 4.2

Overall ISIC 2019 and PH2 are the most suitable datasets. The PH2 dataset is utilized for an analysis of feature extraction techniques such as the detection of ABCD rules and dermoscopic structures. Then the entire technique is analysed using the ISIC 2019 dataset, which is the largest public dataset.

4.2.1 ISIC

The ISIC dataset is a collaborative effort of many institutions to support the development of automatic classification methods for melanoma detection. The most recent applicable dataset is ISIC 2019, which contains a total of 25,331 images for training and 8,238 for testing, making 33,569 images in total. Each image has corresponding metadata including sex, age, anatomical site, and diagnosis. These images are separated into classes melanoma (MM), melanocytic nevus (MV), basal cell carcinoma (BCC), actinic keratosis (AC), benign keratosis (BC), dermatofibroma (DF), vascular lesions (VL), and squamous cell carcinoma (SCC).

ISIC 2019 is a combined source of data from different hospital datasets including HAM10000, BCN 20000, and MSK described in more detail in 4.2.1. This is important to mention because each dataset has captured its images slightly differently. There are

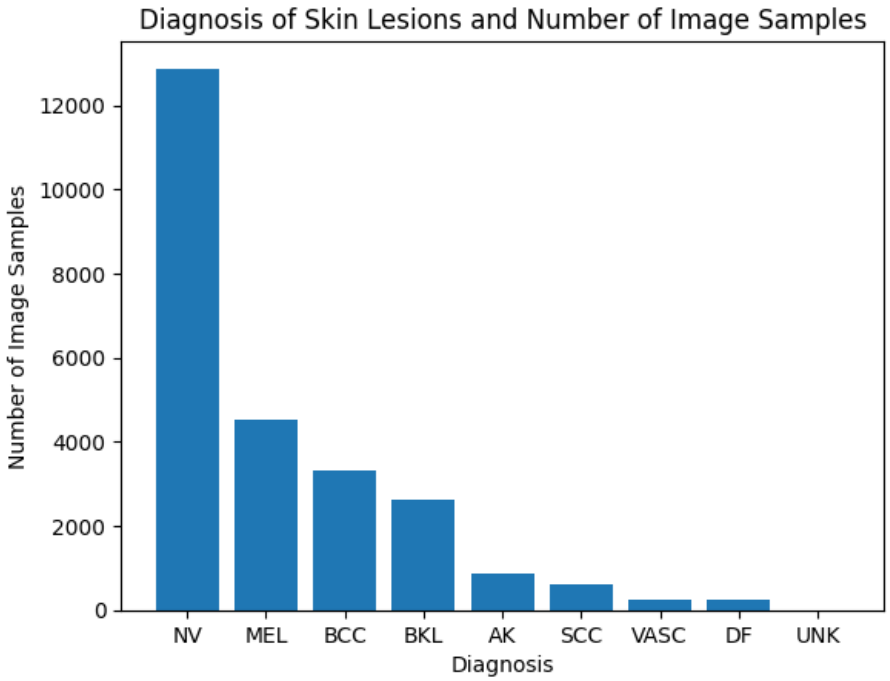

Figure 4.1: ISIC 2019 dataset showing the number of image samples and the diagnosis of those skin lesions dataset appears to be highly unbalanced with half being NV.

varying resolutions and styles in which the images are taken. MSK includes coloured stickers covering un-applicable skin lesions that are within the area of the image and the camera appears to be further away. BCN 20000 contains rarer types of images including nails and mucosa, which do not appear in the other datasets. HAM10000 does not appear to have any notable differences, but it is still captured with different tools and resolutions.

As demonstrated in figure 4.2.1 the dataset is highly in-balanced based on the diagnosis of the skin lesion with 12,875 NV and 4,522 MEL. There are only 867 AK, where AK has a very high variance compared to other skin lesions and the sample size is unlikely for adequate detection. Seborrhoeic keratosis (SK) is not in this dataset and AK a similar lesion is described instead. Others such as DF, and VASC only contain a handful of images. The difference in image samples makes the dataset primarily useful for testing between NV and MEL.

Metdata

In this dataset, the images are accompanied by metadata describing patient information. This includes patients age, sex, and anatomical location. Analysing this data provides

further insight into the influence of patient information on the classification process. Most of all we are looking for similar numbers for each category and some cases might be removed from the classification process to improve results.

Figure 4.2.1 demonstrates the number of samples relating to the patient's sex. There are more samples for each type of skin lesion except for NV which there are more female samples. They are all within a very close range and are unlikely to need rebalancing.

As demonstrated in figure 4.2.1 the dataset is unbalanced relating to age and type of skin lesion. Variation might be because skin lesions are more likely to develop in older people than younger ones. This might mean that many of the skin lesions are developed and there is going to be unlikely to find underdeveloped skin lesion samples.

In Figure 4.2.1 the age approximation is compared with the type of skin lesion. The black line (whisker) represents the minimum and maximum range of age, the box (quartile) shows the interquartile range (IQR), and the centre line in the middle represents the median. Some dots represent outliers in the data, that are outside the age range. Each class in the diagram is a diagnosis associated with the age of each patient. Interestingly, represented in this data SCC and AK appear to develop more in older adults and many of the younger patients were diagnosed with NV or DF. This is correct when regarding literature considering that SCC and AK develop in older adults[**empty citation**].

Figure 4.2.1 is similar to 4.2.1, except it compares the approximate age and location of the skin lesion. There are more older patients who have been diagnosed with skin lesions on their head/neck and younger for the lateral torso. It is concerning that there is a distinct lack of younger patients for palms/soles and head/neck, which could mean more developed skin lesions in these criteria.

Dermoscopic Structures

ISIC 2017 which shares some images with ISIC 2019 has some additional metadata relating to dermoscopic structure. This includes 2,694 segmentation masks of pigment networks, negative networks, globules, milia-like cysts, and streaks. While the original in ISIC 2017 only has metadata for dermoscopic structures, it was linked to ISIC 2019 using image file names to get their diagnosis.

The dermoscopic structures described in figure 4.2.1 show the number of image samples for each dermoscopic structure. Naturally pigmented networks have more than 1400 images which makes it ideal for training a SegNet algorithm. Other demroscopic structures are lacking and do not have enough data for any deep-learning algorithms.

As described in figure 4.2.1 certain dermoscopic structures are split almost evenly between melanoma and benign naevi, except for streaks and negative networks being more common. This demonstrates the importance of being able to detect the difference between typical and atypical pigmented networks. Furthermore, milia-like cysts appear more frequently in seborrhoeic keratosis and there is a lack of pigmented networks.

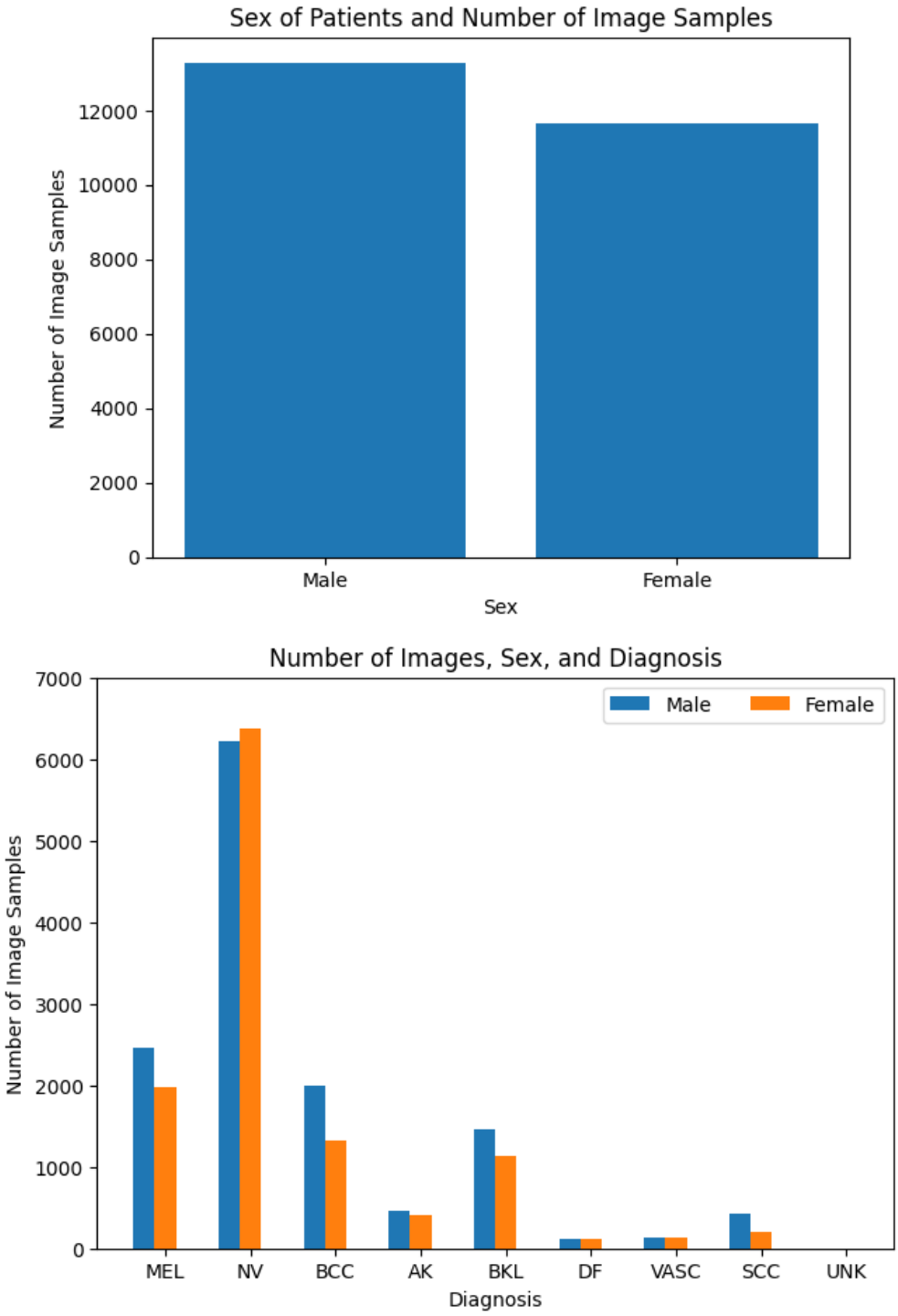

Figure 4.2: The dataset has more male than female patients except for NV which has more samples.

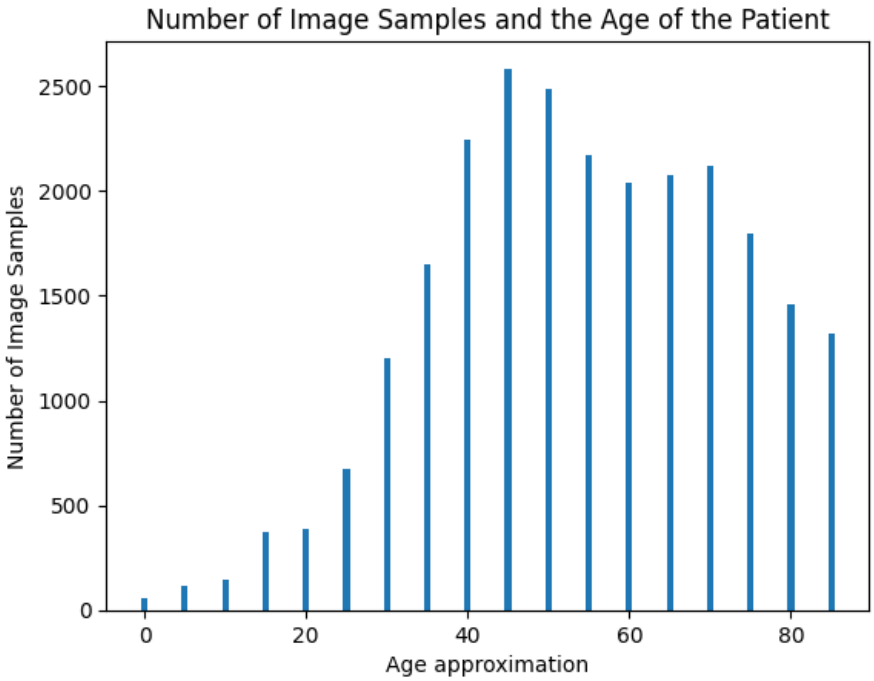

Figure 4.3: This shows the number of image samples compared to the age, the dataset is largely unbalanced regarding age where patients are between 40 and 75 years of age.

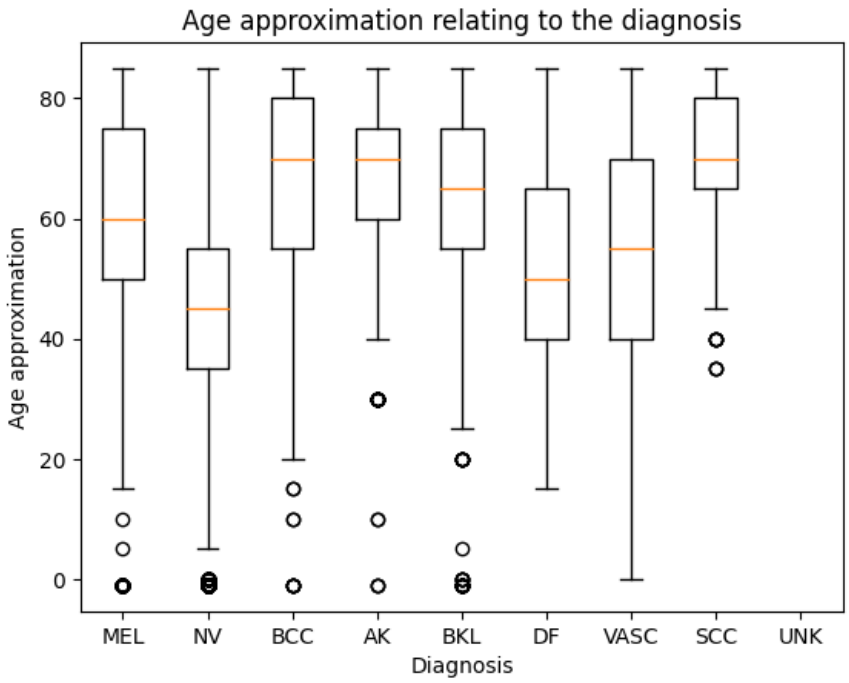

Figure 4.4: The approximate age range of patients and their diagnosis.

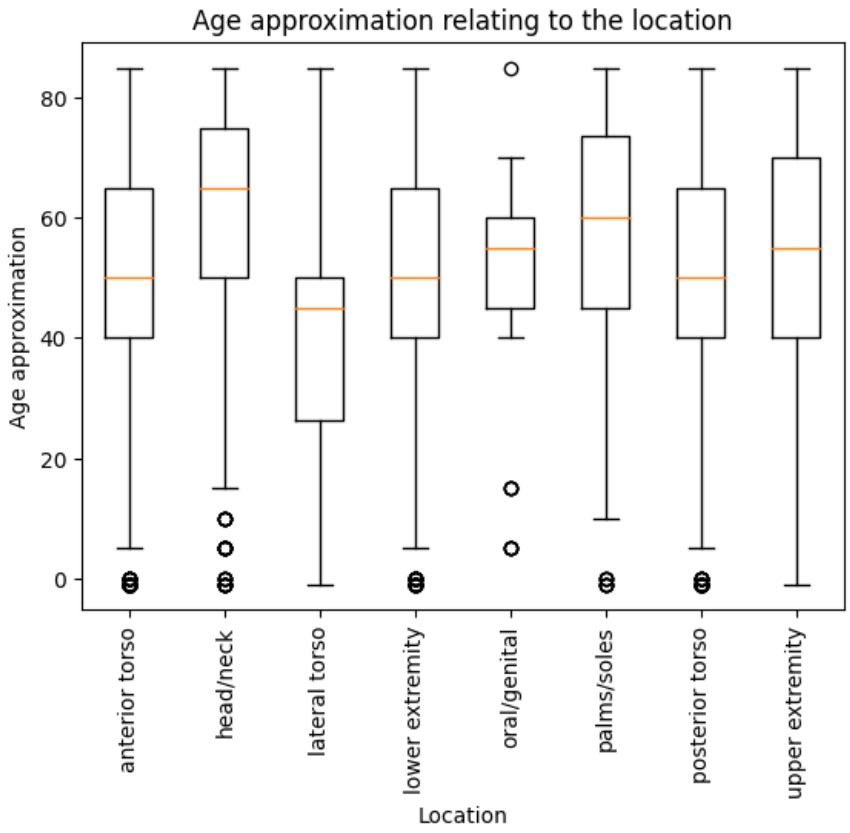

Figure 4.5: The dataset has more male than female patients except for NV which has more samples.

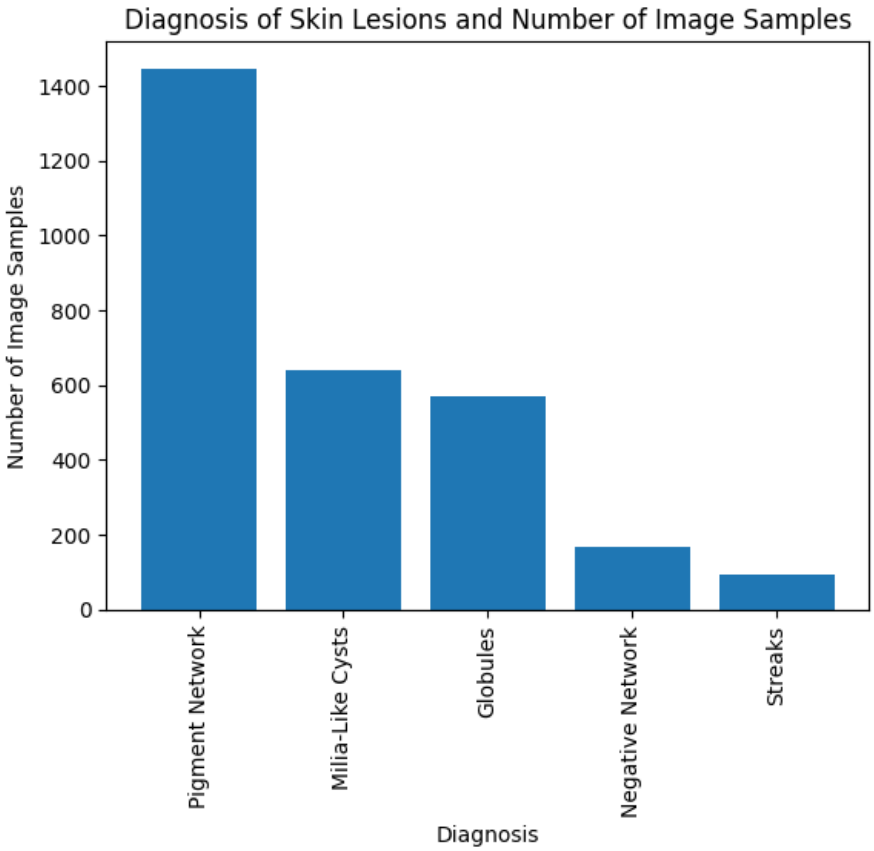

Figure 4.6: Number of images containing dermoscopic structures.

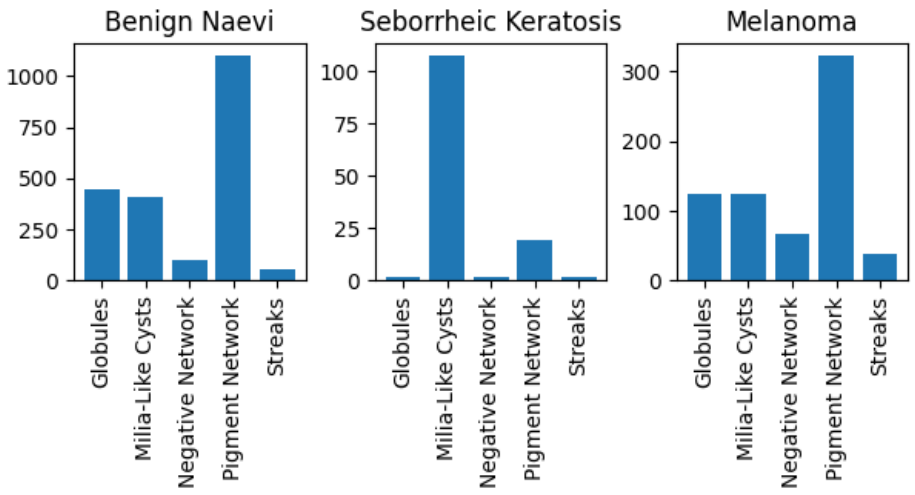

Figure 4.7: Number of dermosocpic structures relating

Summary

In summary, the ISIC dataset including data from 2017, 2018, and 2019 makes this dataset the largest public dataset for skin lesion analysis and melanoma detection. It contains a large collection of dermoscopic images with 8 different diagnoses. The dataset having 33,569 makes it ideal for a diverse range of research and development purposes including the evaluation of machine learning and deep learning models. It also contains 2694 images of dermoscopic structures labelled in the ISIC 2017 version of the dataset. Overall, this is the best dataset currently publicly available for the analysis of skin lesions.

4.2.2 PH2

The PH2 dataset is a collection of dermoscopic images that were made available in 2013 by Mendonca, et al[**empty citation**]. It consists of 200 images including 80 common nevus, 80 atypical nevus, and 40 Melanoma. Although the dataset is small it holds substantial metadata for describing features within the skin lesion, including asymmetry, colour, pigment network, dots/globules, streaks, regression areas, and blue-whitish veil. This is the only dataset that has such substantial data regarding smaller features. Each image has a segmentation mask of the skin lesion.

The images in figureph2-example-images are some examples of skin lesions from the PH2 dataset. All the images have a circular border and melanoma samples are too big to fit inside the area of the dermoscope in many cases. This results in a skewered border making it difficult to analyse asymmetry and border from ABCD rules.

As shown in figure 4.2.2 there are 200 image samples in total with 80 common nevus,

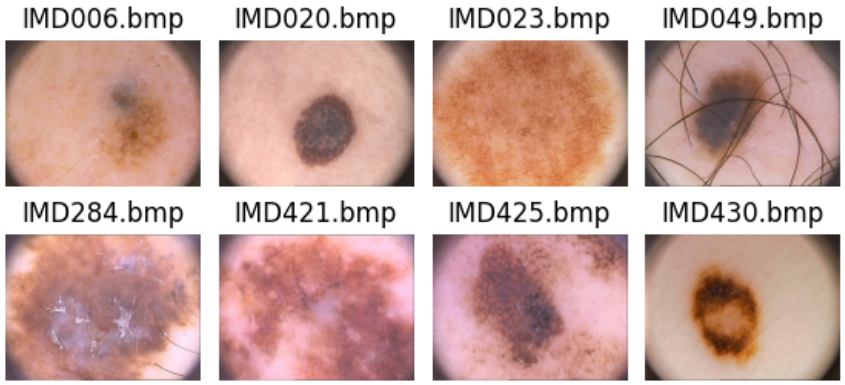

Figure 4.8: Example of images from the PH2 dataset. The first two are standard, the second two are atypical, and the last 4 are melanoma.

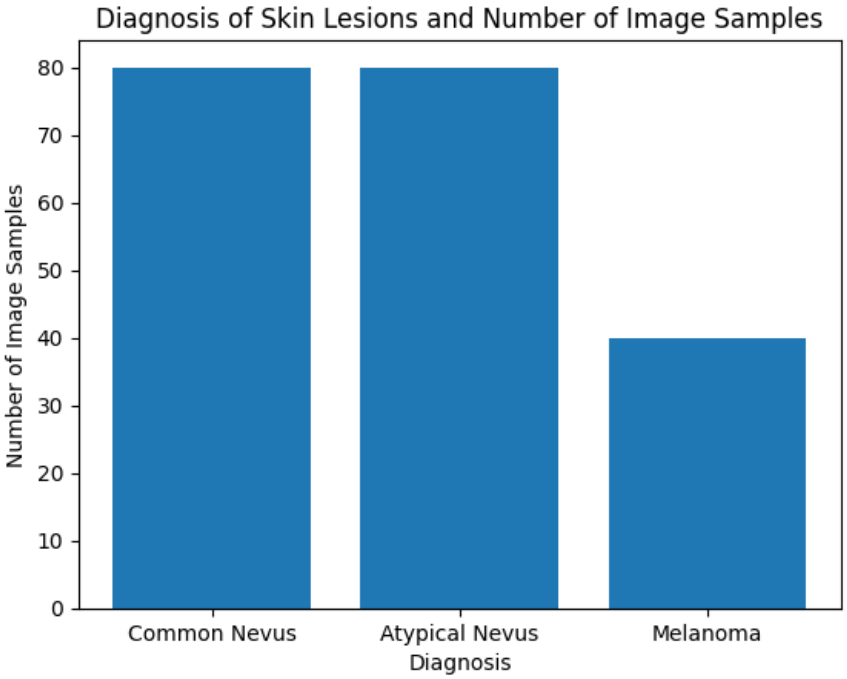

Figure 4.9: Number of image samples and diagnosis in the PH2 dataset.

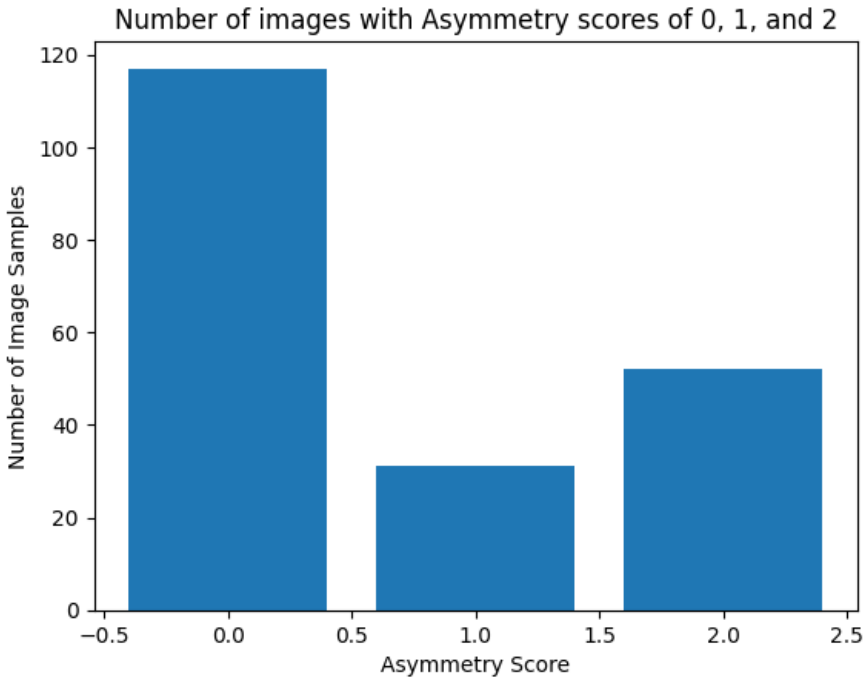

Figure 4.10: This shows the number of image samples and asymmetry score based on Total dermoscopy score (TDS).

80 atypical and 40 melanoma. The number of image samples is too small for most neural network techniques. The dataset is highly unbalanced with only 40 melanoma images and 160 naevus images. With such a skewed distribution of classes, the model might become overly biased towards the majority class (naevus), leading to poor performance in identifying melanoma. Another issue is the size of these classes, the scarcity of samples will likely result in overfitting when training models. For this reason, this is not a reliable candidate for training machine learning algorithms when considering that more diverse candidates such as ISIC exist.

Metadata

The benefit of this dataset is the rich metadata allowing for the analysis of specific features within an image and the relationship between them. This allows for the development of more sophisticated algorithms that provide further insight into the characteristics of melanoma and naevus.

Demonstrated in figure 4.2.2 demonstrates there is substantial data for measuring the asymmetry score based on the total dermoscopy score (TDS). Typically the algorithms

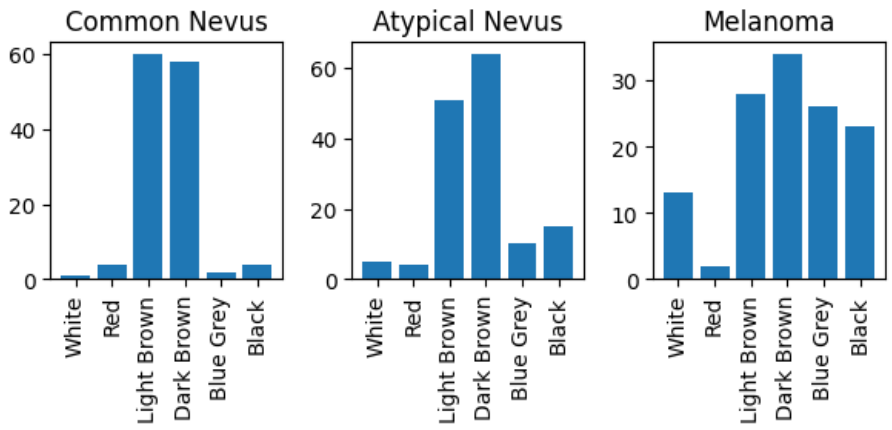

Figure 4.11: Number of colours in the PH2 dataset compared with the diagnosis. Colours are in order white, red, light brown, dark brown, blue-gray, and black.

used to measure asymmetry such as bi-fold do not require any training, making the smaller sample size ideal. However, there is a very small sample size for both TDS of 1 and 2, and it would be beneficial to have more data samples.

The observation in figure 4.2.2 demonstrates that Light and dark brown are commonly associated with typical naevus. On the other hand, white, blue, and black are more common in melanoma because they indicate structural and vascular irregularities. The scarcity of red samples demonstrates that it is uncommon in nevus and melanoma. Other lesions including BCC are more likely to contain red[**empty citation**], but these lesions are not included in this dataset.

There are many records of pigment networks, and dots/globules, but as seen in Figure 4.2.2 there are roughly 20 samples for each streaks, regression, and blue-whitish veil. The data is highly unbalanced, so it will be difficult to train a machine-learning algorithm for these features. There are more samples of pigment networks, and dots/globules because common. For this reason, typical and atypical features are a good indication of whether the skin lesion is melanoma.

Figure 4.2.2 shows dermoscopic structure labels relating to the diagnosis of the skin lesions. Common nevus have typical and present dermoscopic structures, atypical nevus have just as many absent with more present and atypical. Melanoma has more present and atypical types of skin lesions.

Figure 4.2.2 demonstrates that pigment networks are present in both nevus and melanoma. Furthermore, streaks, regression areas, and blue-whitish veils are more common in melanoma. Pigment networks and dots/globules use different labels of typical and atypical so it is understandable they do not change between lesion types.

There is little to no overlap between typical and atypical pigment networks between

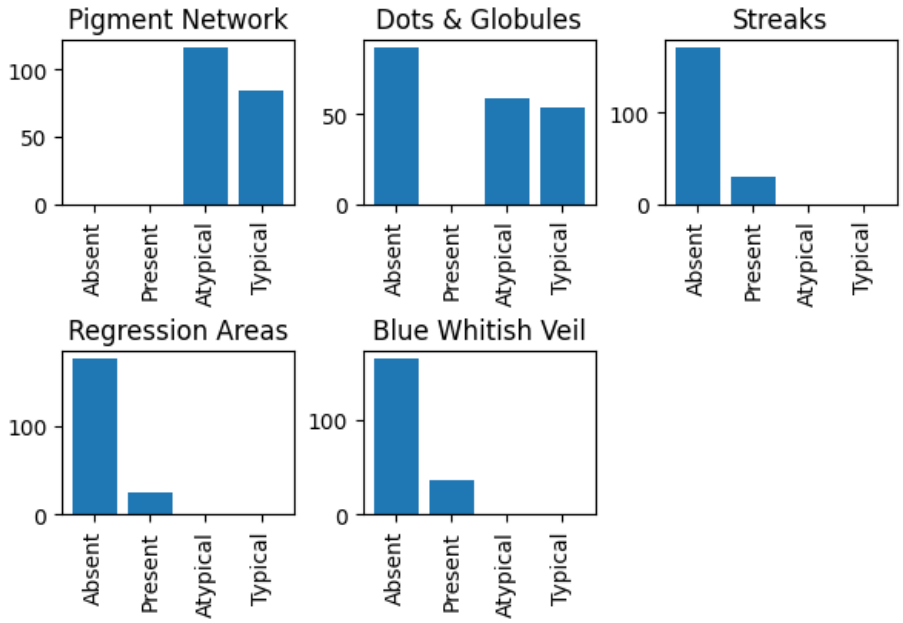

Figure 4.12: Dermosocpic structures and the number of images. These are labelled between absent, atypical, present, and Typical.

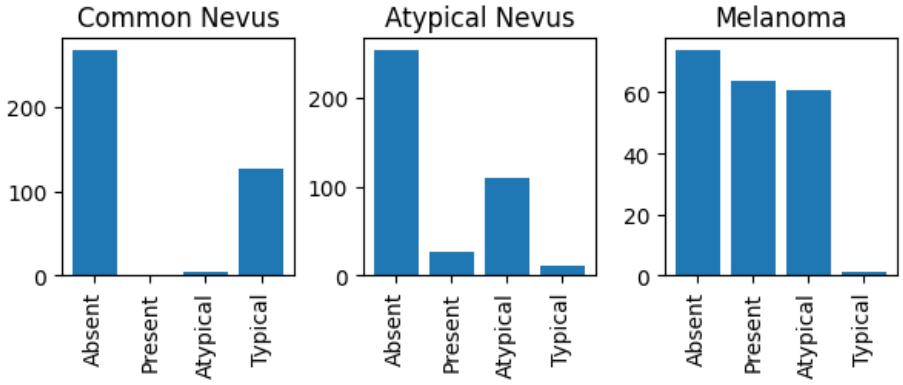

Figure 4.13: Shows the labels of dermoscopic structures, number of images, and diagnosis. These are labelled between absent, atypical, present, and Typical.

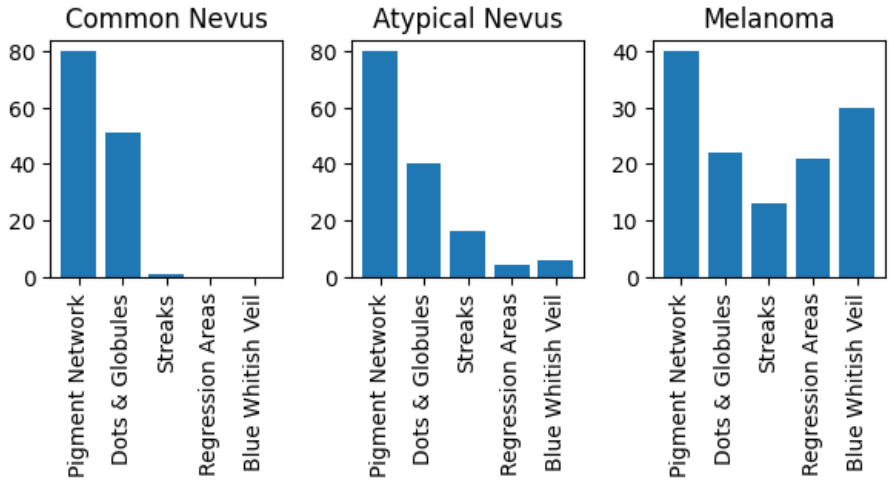

Figure 4.14: Shows the number of images based on the diagnosis and dermoscopic structures present, typical, and atypical.

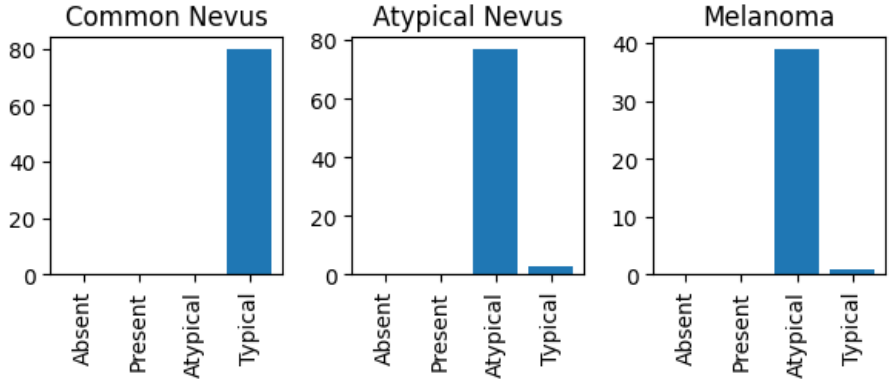

Figure 4.15: Pigment network data relating to the diagnosis.

melanoma and naevus. Shown in figure 4.2.2 common naevus are all typical pigment networks, while atypical naevus and melanoma are labelled atypical. This demonstrates that testing for pigment networks should be on type (typical and atypical) instead of whether they are present. Unusually, there is very little overlapping in the data. This is very unusual unless it was designed this way purposely, but it would have been more useful to have more samples without pigment networks.

Summary

In summary, the PH2 dataset is a valuable resource for researchers considering it is the only one of its kind to provide metadata relating to asymmetry, colour and dermoscopic structures. This has been used frequently in various studies to develop and evaluate algorithms for skin lesion analysis. Such datasets with substantial metadata are useful for producing explainable results. As explainable (XAI) becomes more common more datasets describing the data will be necessary. However, one of the downfalls of this dataset is the unusual labelling for dermoscopic structures where the pigment network and globules are labelled between typical and atypical and others are between present and absent. Furthermore, almost every image sample has a pigment network, it would be more useful to have some without this feature.

4.3 Creating 'the Dataset'

Whilst recognising the benefits of the ISIC 2019 and PH2 datasets we can begin to develop 'the dataset' using NHS data. The inclusion of NHS data brings real-world clinical cases into the mix, and techniques developed during this project can be tested in the scenario it is intended to be utilised. Requirements are decided below to highlight potential biases and issues, and then 2,000 images are chosen to create the dataset. This is followed by an analysis similar to the previous datasets and interesting findings from the data.

4.3.1 Requirements

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

There are requirements for this project, which include the use of macroscopic images instead of dermoscopic images. Macroscopic is described as viewing with the naked eye or by taking a picture with standard lenses. When referring to Dermoscopic images, are images captured with a specialised tool called a dermoscope that removes lighting variegation and improves the visual features within the skin lesion usually called dermoscopic features.

Although macroscopic images are used it is important to note that 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 networks, asymmetry, irregular borders, and other features that support 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[13]. Similarly, the sensitivity for SCC was 77% with dermoscopy, compared to 70% with close-up images[13]. These findings highlight the superior diagnostic performance of dermoscopy compared to macroscopic.

The dermoscopic examination is superior to macroscopic examination, however, the project use case specifies macroscopic. The logic behind this is that the tool is specifically designed for general practitioners who are unlikely to recognize dermoscopic features, so there is no need to supply them with dermoscopes. This appears to be consistent with an author's findings showing that 92% of dermatologists correctly recognize at least four 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.

Considering the use of macroscopic images there needs to be a more thorough clean-up of the data for it to be used effectively. This will include removing hair and specular reflections 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 improving the accuracy when classifying.

4.3.2 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.

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 does 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 and general practitioners have diagnosed the large majority of skin lesions which have varying accuracy depending on their experience. Images include metadata on the department and person capturing the image, so the doctors' experience can be measured.

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 they lack the insight that an in-person examination would provide.

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

5. Diagnoses of skin lesions are written in plain text including question 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) has 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 many SK images. It will be vital to separate these.

Following the potential issues 2,500 images were chosen from the NHS database. In the next section, further analysis of the images using the metadata will be conducted to remove image samples with uncertain diagnoses and to balance the dataset for better use with machine learning algorithms.

4.3.3 Data Transformation and Analysis

Each image holds metadata shown in table 4.3.3 as EXIF Tags within each image. After extracting the images the metadata included in the images are Filename, Tags (Capturing method, anatomical location), Gender, DOB, Department, Consent, Historical Diagnosis, Diagnosis, and Date Photographed. There was other metadata with each image, but they are potentially identifiable, so in order to protect patient confidentiality, they were removed.

The dataset includes skin lesions with diagnoses of Malignant Melanoma (MM), Seborrhoeic keratosis (SK), Atypical Naevi (AN), Typical Naevi (TN), Squamous Cell Carcinoma (SCC), Basal Cell Carcinoma (BCC). This is a generalised diagnosis, alongside these images are historical diagnoses of the skin lesion, written in plain text in some cases this question marks when the doctor is uncertain of a diagnosis and other times it includes a slash and a different diagnosis. Although it is not a specific format making it difficult to process, it includes a wealth of knowledge.

Historical Diagnosis of Skin lesions

The historical diagnosis is written text by the doctor that includes the possible diagnosis. The format is dependent on the doctor, so it changes dramatically. Generally, this is a '?' to show uncertainty and a '/' followed by another diagnosis. All of the variations in the format are shown in table 4.3.3.

Some historical diagnoses contain a '?' showing uncertainty from the doctor. Figure 4.3.3 describes the number of skin lesions between uncertain and certain. Interestingly AN, BN, and BCC appear to be certain with 117, 36, and 39 respectively. Followed by SK and SCC are roughly half of the images at 106 and 87. Most of all MM shows that more than half of the diagnoses at 259 are uncertain out of 184 that are certain. This in turn

Attribute	Description
Image ID	an integer representing the image ID - example: 123456.xxx
Doctor	a string representing the forename and surname of the doctor that made the diagnosis - example: JOHN SMITH
Department	a string representing the name of the department - example: DERMATOLOGY
Studio	a string representing the name of the studio used to capture an image of the skin lesion
Capture date	an integer representing the date the image was captured - example: 00:00:0000
Hospital ID	a string representing the hospital ID
Gender	a string representing the patient's gender
Date of Birth ID	an integer representing the patients date of birth - example: 00-00-0000
Surname	a string representing the patients surname
Forename	a string representing the patients forename
Initials	a string representing the patients initials
Patient ID	an integer representing the patients ID
Subject (Tags)	an array representing method the image capturing method (Dermo or Close-Up) and anatomical location
Creator	a string representing the forename and surname of the photographer - example: JOHN SMITH

Table 4.3: This table shows the metadata in each image and a description of each label. Rows highlighted in red are removed to protect patient confidentiality.

Image ID	Historical Diagnosis
998444.jpg	SEB K
549982.JPG	AK / SCC
824466.jpg	ATYPICAL MOLE / ? MM
879067.jpg	? ATYPICAL NAEVI
1028628.jpg	? MM / ? BCC
154414.jpg	1) ? SCC 2) SBCC 3) ? SPOT
586010.JPG	SUSPECTED N. MM

Table 4.4: Examples of historical diagnosis and doctors and some unique variations of labelling.

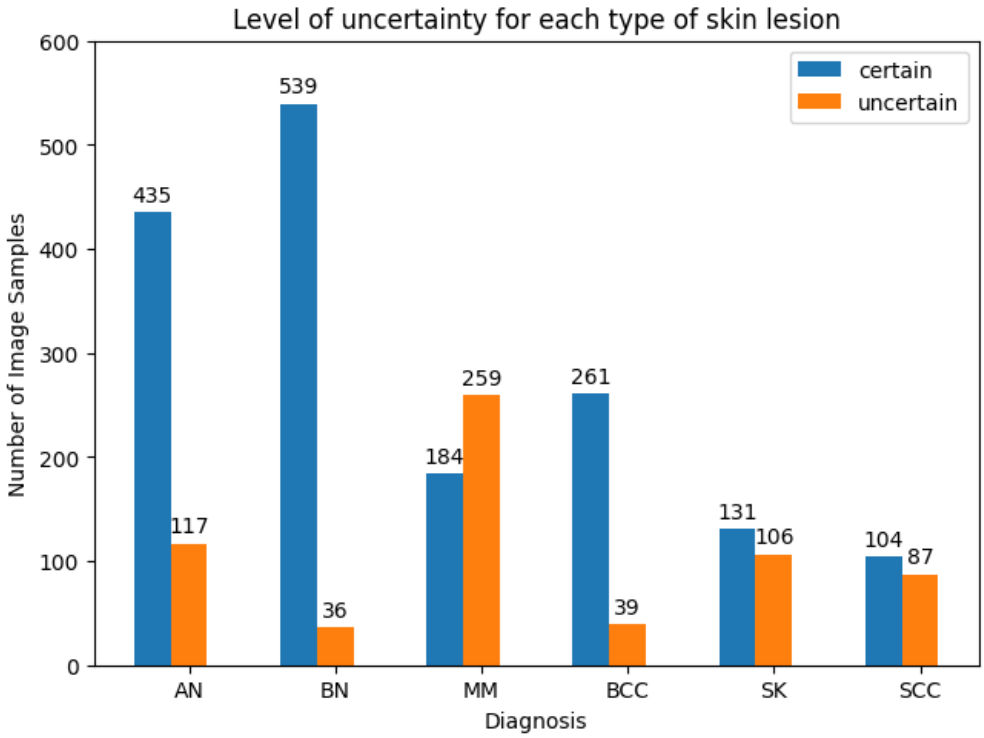

Figure 4.16: Number of image samples relating to the historical diagnosis. Labelled as uncertain if there is a '?' in the diagnosis.

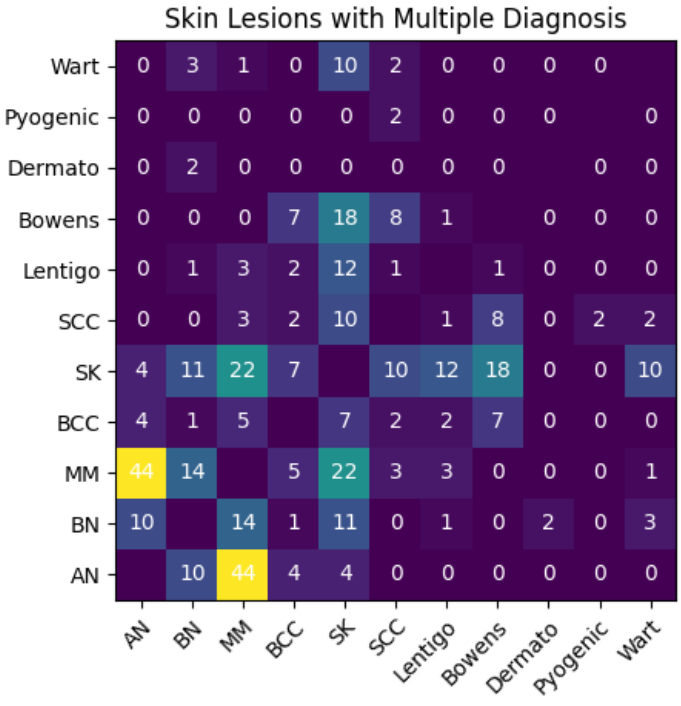

Figure 4.17: Number of skin lesion samples with multiple diagnoses in the historical diagnoses. Other types including lentigo, Bowen's disease, dermatofibroma, pyogenic granuloma, and wart are only associated with the main diagnoses because they are not specifically searched for.

Category	Organised sub-labels
Upper Extremity	Wrist, Elbow, Arm
Lower Extremity	Leg, Knee, Hip, Ankle
Lateral Torso	Axilla, Breast
Posterior Torso	Back, Shoulder
Anterior Torso	Chest, Abdomen, Trunk
Palms/Soles	Hand, Thumb, Foot
Oral/Genital	Groin, Genitalia, Sacrum, Buttocks, Sacrum
Head/Neck	Neck, Chin, Face, Temple, Head, Forehead

Table 4.5

demonstrates the type of skin lesions that doctors are having difficulty diagnosing where melanoma is especially difficult.

Multiple diagnoses are sometimes shown, figure 4.3.3 shows the number of skin lesions with multiple diagnoses mentioned in the historical diagnoses. Interestingly the most commonly associated are AN with MM at 44 and SK with MM at 22 images. Others are SK with MM, Bowen's disease, lentigo, warts, and SCC demonstrating that SK is associated with the widest range of skin lesions and the difficulty diagnosing it.

Anatomical location

Anatomical location has a total of 28 different descriptors for specific body parts including leg, wrist, neck, etc. To make comparisons easier with the ISIC dataset each label has been assigned to specific areas such as upper extremity, anterior torso, etc. All the locations are listed in the table 4.3.3.

4.3.4 Data Transformation and Augmentation

As mentioned in the data biases section the skin lesion images are taken under various 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.

One of the main challenges in melanoma detection is the visual similarity between 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.

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[**kasmi2023**].

Hair Removal

Hair artefacts in images can interfere with the recognition of handcrafted features and affect the performance of deep learning algorithms in melanoma detection[**kasmi2023**]. Applying morphological operations such as image sharpening and segmentation techniques can remove hair artefacts from dermoscopic images[**kasmi2023**].

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

Sharp-Razor[**kasmi2023**] is a technique for detecting hair and ruler marks to remove them from images. This uses a multiple-filter approach including grayscale plane modification, hair enhancement, segmentation using tri-directional gradients, and multiple filters for hair of varying widths. This technique is shown to outperform existing methods.

Specular Removal

Specular reflection removal techniques are effective in improving the accuracy of melanoma detection[**Shen2009**]. A technique was proposed utilizing a partial differential equation to iteratively erode the specular component, removing the specular reflection[**Shen2009**].

4.3.5 Conclusion

4.3.6 Dataset Statistics

The dataset has been analysed and modified accordingly, originally starting with a total of 2,500 and ending up with ...

As shown in figure 4.3.6 the image data and diagnosis of the skin lesion there are several differences in this dataset compared with the ones described so far. There has been more of an attempt to balance the data so there are more equal samples of each. Furthermore, benign naevi have been split into benign naevi (BN) and atypical naevi (AN). There are images of seborrheic keratosis, which is more than any other public dataset currently available.

The age variation of patients described in figure 4.3.6 demonstrates there are many younger patients included in the NHS dataset. This is substantially different from other datasets including ISIC that have mostly older patients. This demonstrates that there is an influx of younger patients regardless of them not being within the age group where melanoma usually develops. In both ISIC and NHS datasets the median the median is 60 years.

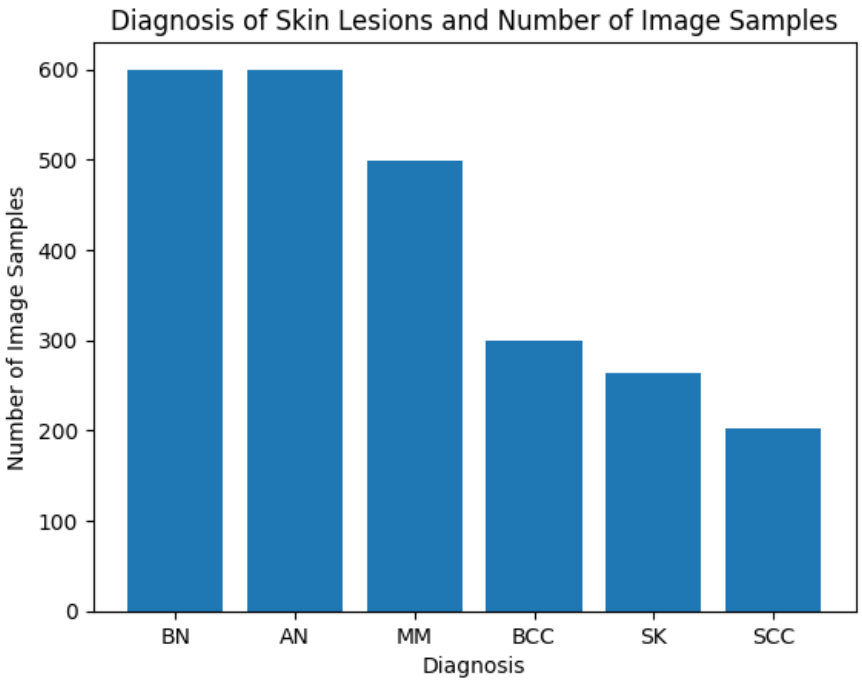

Figure 4.18: Number of image samples relating to the diagnosis of the image.

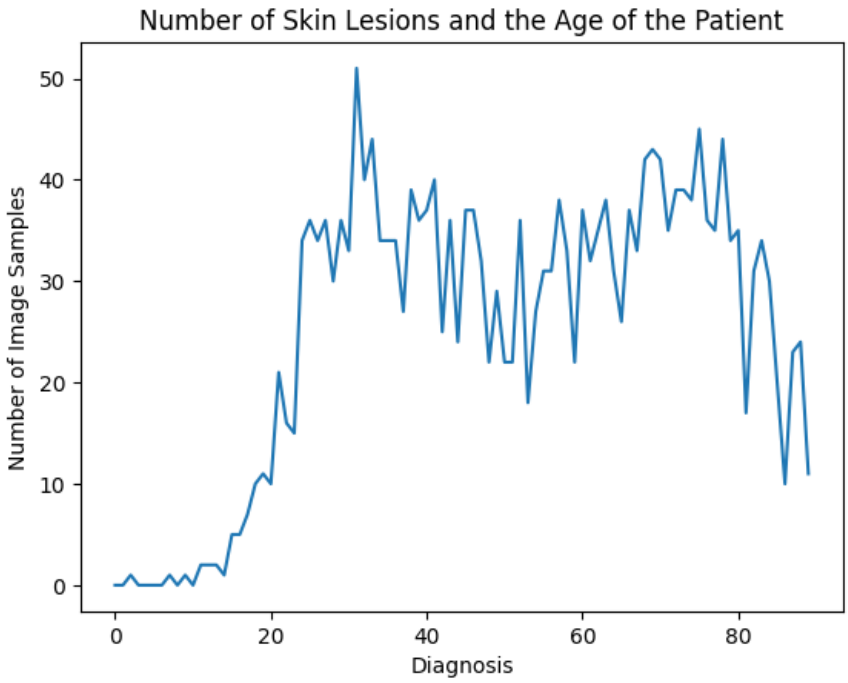

Figure 4.19: Age of patients and number of image samples.

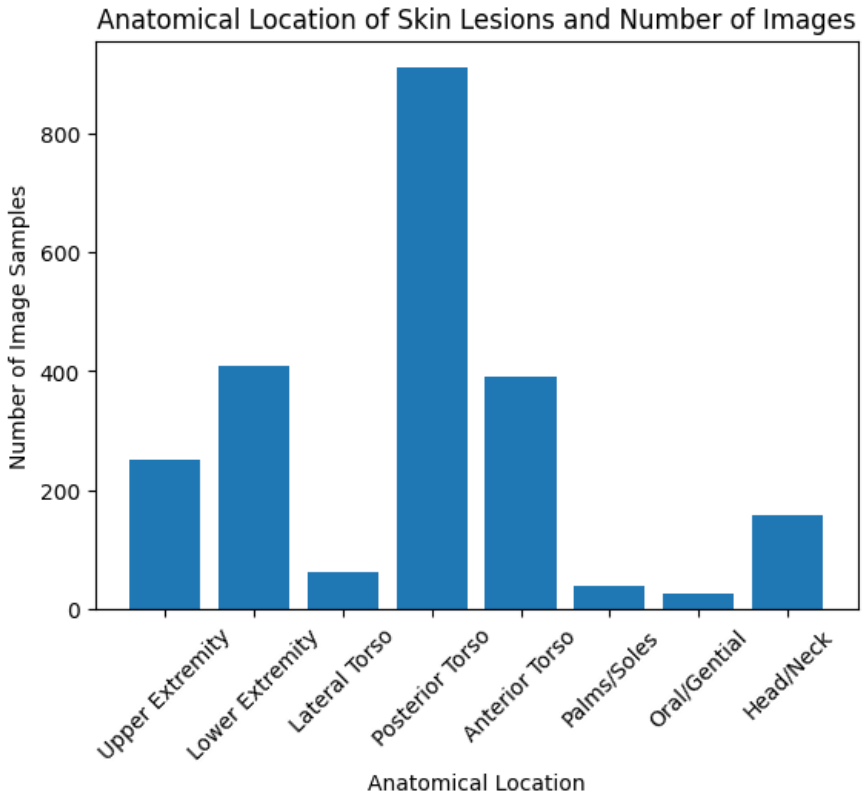

Figure 4.20: Number of image samples related to the location of the skin lesion.

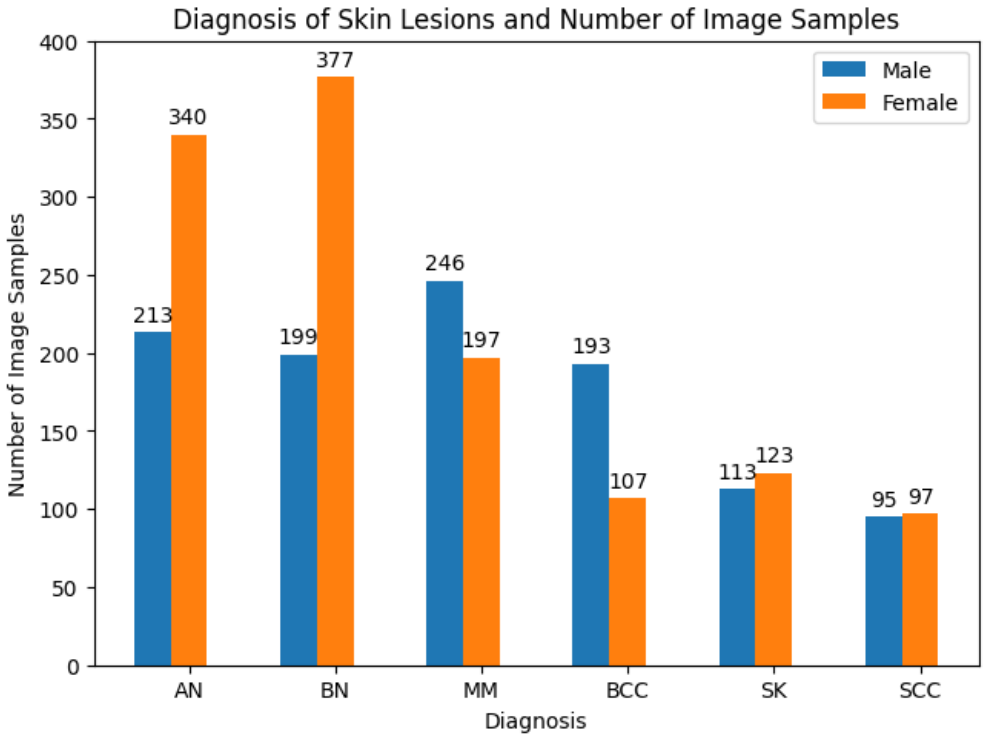

Figure 4.21: Number of image samples relating to the diagnosis and sex of the patients. There are more female than male patients.

As shown in figure [**nhs-location**] describing the location of the skin lesions and several image samples. There are more samples on the posterior torso (back) compared with other skin lesions with any of the others. There are only a couple of samples for lateral, palms/soles, and oral/genital. This data was originally modified because it had a total of 28 descriptors, so they were grouped into 8 similar to the ISIC dataset. This can be seen in more detail in figure 4.3.3.

Figure 4.3.6 describes the number of image samples relating to the diagnosis and sex of the patients. Interestingly there are almost double the number of female patients being diagnosed for AN and BN compared with MM where there are slightly more male patients and BCC where there is almost double male.

Image samples described in figure 4.3.6 demonstrate the age of patients compared with their diagnosis. Understandably, AN and BN which are neavus are from younger patients, while SK, SCC, and BCC appear in older adults. MM is primarily from patients at the age of 50 to 70.

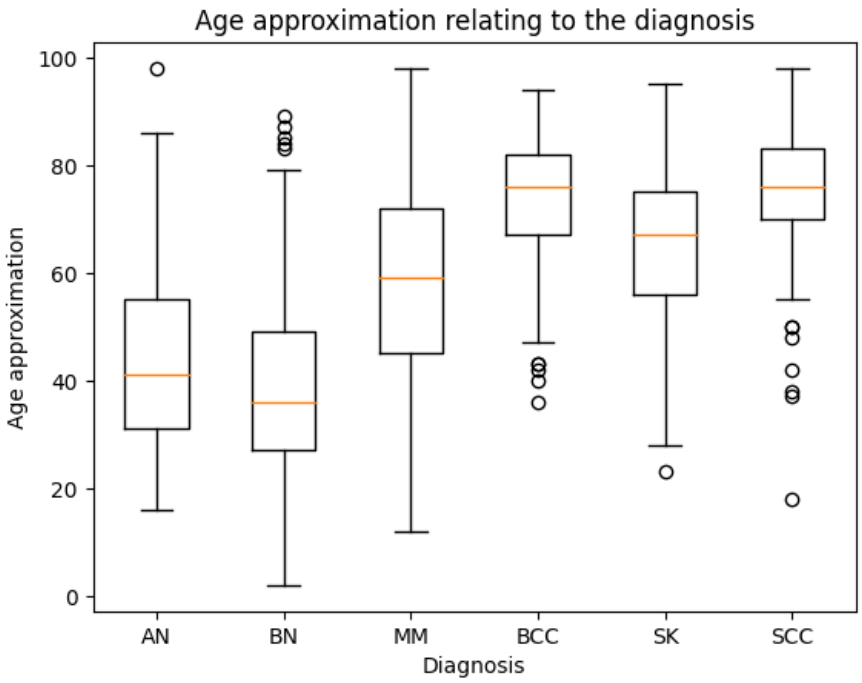

Figure 4.22: Boxplot describing the age of patients and the diagnosis.

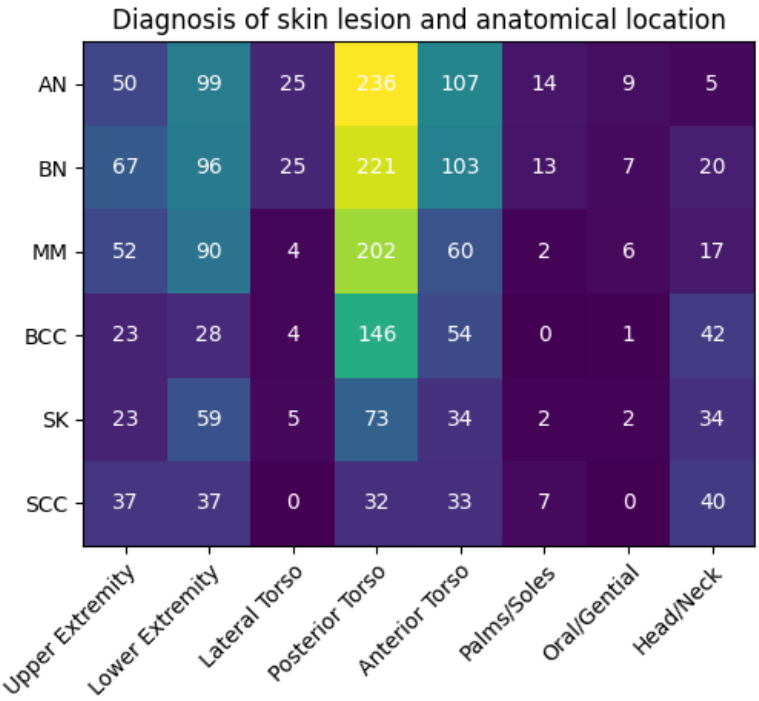

Figure 4.23: Number of image samples relating to the diagnosis of the image.

4.3.7 Issues

One fundamental problem is the overutilisation of private or privately annotated datasets, making a direct comparison of algorithms (especially relating to ABCD rules) difficult. Some are between benign and malignant[23, 19, 4, 3] while others utilise private or never mention any datasets[19, 35, 41, 29, 48]. None compare their rules, likely because of subjectivity depending on the dermatologists that labelled them. Ideally, more datasets and labels should be public to assess individual rules and reach objective measurements. Until then, testing algorithms conform with malignant, suspicious, or benign.

The data presented is from David Wen, et al. [**Wen2022**], but is updated to include meta-data relating to diagnostic procedures for testing. Although there are thousands of images for dermoscopic and macroscopic. Data is lacking relating to diagnostic procedures.

Chapter 5

A Comparison of Melanoma Segmentation Algorithms using Neural Networks and Statistical Models

5.1 Introduction

This section includes a discussion on segmentation techniques for the detection of melanoma. The goal is to produce 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, 20] 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. This chapter includes a simple border analysis technique called fractal box counting to assess the benefits of using segmentation algorithms with accurate border cut-offs.

5.2 Related Works

An approach by Albanhli[2] uses deep learning-based segmentation algorithm using YOLOv4-arkNet and active contouring for melanoma and skin lesion detection and segmentation. This technique provides classification of the skin lesion and a segmentation, demonstrating a high level of practicality for clinical decision support systems.

Seeja R D[**seeja2019**] proposed a technique that utilizes a convolutional neural network (CNN) based on a U-net model architecture for the segmentation based on colour, texture and shapes. The U-net model architecture is a popular choice for image segmentation tasks due to its ability to capture both local and global features effectively.

Hyunju Lee[**Lee2020**] proposed a technique that utilizes an edge fill method called u-otsu for segmentation, using the U cahnnel from the YUV colour space to calculate the histogram. Otsu calculates the optimal threshold value to separate foreground and background pixels based on the histogram of the image.

Another technique by Pedro[28] uses a newly developed technique called Local Binary Patterns Clustering (LBPC). Using a Local Binary Pattern (LBP) filter by subtracting the grayscale image from the LBP filter after a Gaussian filter, resulting in the creation of a mask. This has been successfully used for the detection of melanoma.

5.3 Skin Lesion Segmentation

Segmentation plays a crucial role in melanoma detection because it separates melanoma from healthy skin. Accurate segmentation is essential for various aspects of melanoma diagnosis and treatment, and classification[2] including improving the detection of of ABCD rules[**Lee2020**]. One of the important features is border, which to improve accruacy is

A range of traditional segmentation techniques including SegNet, U-net 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.

5.3.1 Semantic Pixel Wise Segmentation (SegNet)

SegNet is a deep learning architecture that is used for semantic image segmentation for melanoma detection. It was originally developed by[11] 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.

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 5.1 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 5.1 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].

5.4 Border extraction

To prove the usefulness of segmentation techniques with an accurate border cut-off a technqiue developed by Ali[4] is implemented that utilises machine learning with extracted data including zernike moments, fractal box-coutning, and convexity measurements. Fractal box-counting is used to measure the irregulrity of the border.

The fractal box counting technique is a commonly employed technqiue for analysing the fractal properties. It involves dividing a fractal object or pattern intoa grid of equally sized boxes and counting the number of boxes that contain a portion of the fractals. The process is repeated with different box sizes until the boxes until the relationship between the box sizes and number of boxes is analysed determining the fractal dimension[**Hamburger1996**]. Essentially a more complicated border with corners and convexes will have more boxes and therefore a higher fractal score, than for example a border wth smooth corners and edges which has a lower score. This should provide some evidence on the usefulness of an accurate border.

5.4.1 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 5.1 (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

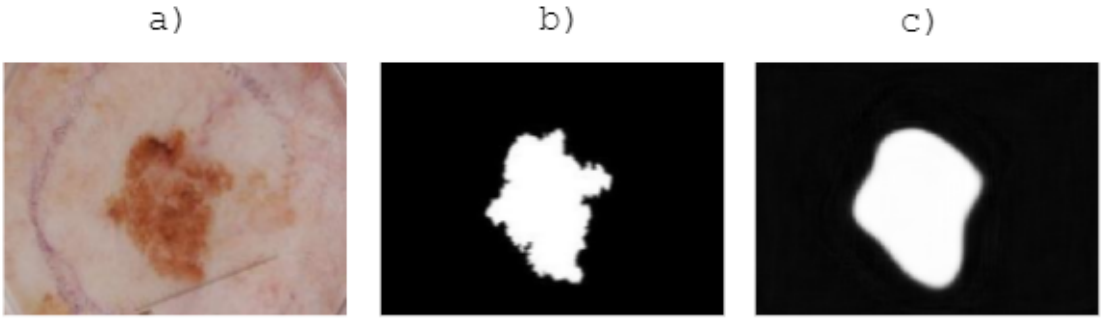

Figure 5.1: Demonstrating the Semantic Pixel-Wise Segmentation (SegNet) results showing the a) original image, b) expert ground-truth and c) SegNet results.

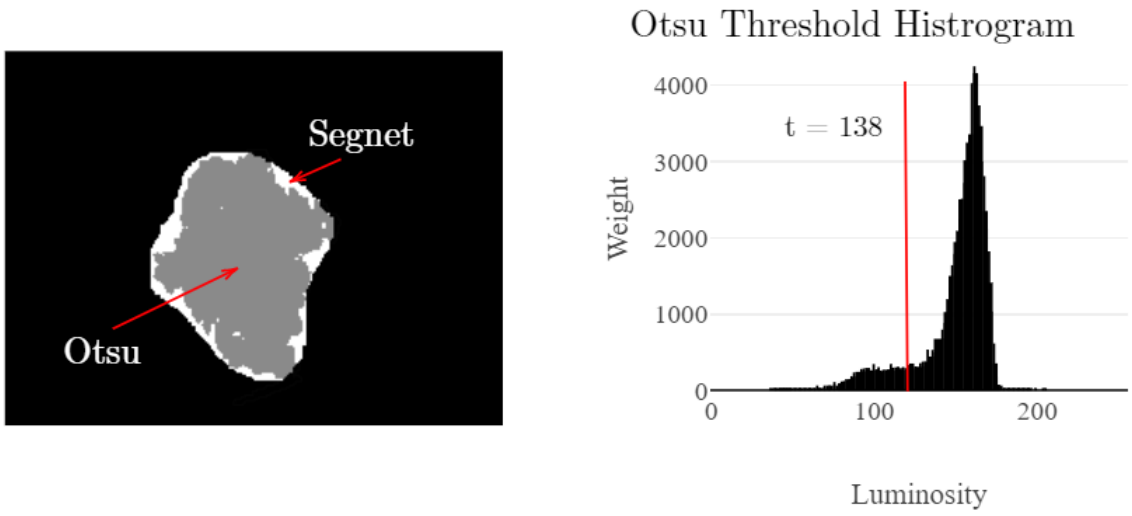

Figure 5.2: Otsu thresholding alongside ground-truth mask, where grey Otsu and white is SegNet. The bar chart shows the histogram with an otsu threshold of 138.

probability w is computed from the histogram in figure 5.2, 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.

$$\sigma_w^2(t) = w_0(t)\sigma_1^2(t) + w_1(t)\sigma_2^2(t) \tag{5.1}$$

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 5.2 describes the difference between otsu and SegNet.

5.4.2 LBPC segmentation

Local Binary Patterns (LBP) is a texture descriptor commonly used for augmenting the image improving classification accuracy[28, 20]. First, equation 5.2 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 5.3 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,

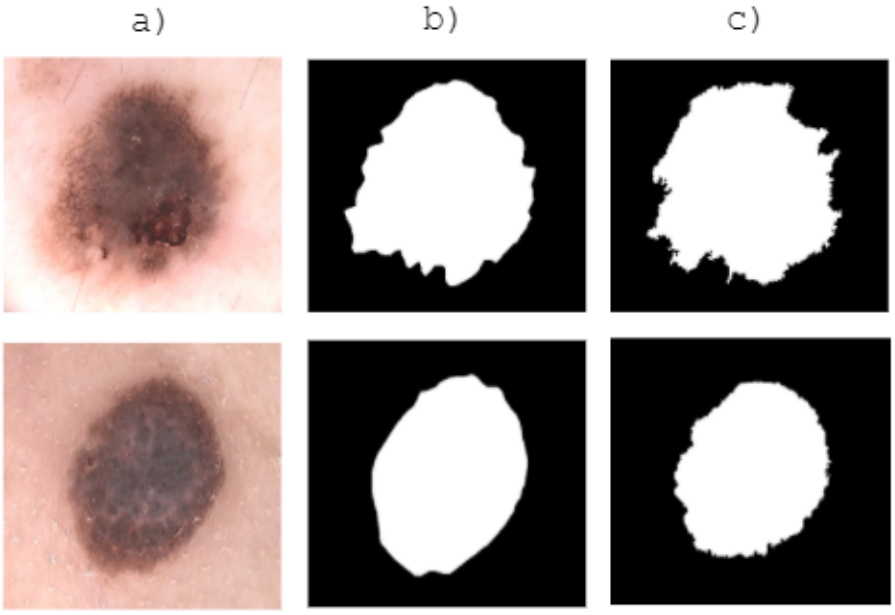

Figure 5.3: Local Binary Pattern Clustering (LBPC) showing the a) original image, b) ground-truth, and c) LBPC. LBPC successfully exaggerates the border cut-off on the skin lesions with regular and irregular borders

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.

$$LBP(gp_x, gp_y) = \sum_{p=0}^{P-1} s\big(gp - gc\big)2^p \tag{5.2}$$

$$s\big(x\big) = \begin{cases} 1, & x \geq 0; \\ 0, \text{otherwise.} \end{cases} \tag{5.3}$$

Figure 5.4.2 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 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, 20]. 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.

5.5 Comparison of techniques

Overall the accuracy of the techniques demonstrates that SegNet is the most reliable technique. However, comparing the techniques in **??** we can demonstrate that it produces a smudge effect and fails to capture the border cut-off from the skin lesion, but it is however successful at finding the location of the skin lesion.

Both statistical models of LBPC and Otsu threshold generated an accurate border cut-off between the skin and the skin lesion As previously mentioned, measuring the border cut-off and exaggerating irregular borders are helpful when calculating the ABCD rules.

It might be beneficial to combine SegNet and LBPC using SegNet to find the skin lesions' location, followed by adjusting the border cut-off using LBPC. A similar technique using otsu threshold and Segnet is described by[30].

5.6 Joint Neural network and statistcal model approach

Chapter 6

Analysing and Updating ABCD Rules Detection Algorithms

6.1 Introduction

This chapter holds a discussion on some of the most popular ABCD (Asymmetry, Border, Colour, and Dermoscopic Features) algorithms including their implementation, and updating the algorithms. They are compared using the $PH/^2$ dataset and updated relating to their accuracy. Supprisingly, many of these technqiues are tested on whether they effectively find melanoma and not individual features. So, this will be the first time some of these techniques are tested and documented regarding the ABCD rules.

6.2 Discussion

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[15]. One concern is the production of realistic, but incorrect results[17]. 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, and dermoscopic features).

6.2.1 Preferred Diagnostic Procedures

Diagnostic procedures are procedures that are preformed on patients in order to diagnose conditions. Regarding the diagnosis of melanoma there are many types that have been utilised for the detection of melanoma and the most favourable is CASH, ABCD rules, and Total Dermoscopy Score (TDS). The ABCD rules and TDS were commonly used because of its simplicity and effectiveness.

Diagnostic procedures are usually based on the doctors medical experience. For example the use-case of this project is for general practioners (GPs), many of which have likely never seen or attempted to diagnose melanoma, and many of which will not have access to dermoscopes for the analysis of dermoscopic structures. So, in this use-case diagnostic procedures including ABCD rules, and CASH are suitable because of their simplicity. The method used by the NHS is also ABCD rules, which is the reasoning for using this method.

Interestingly dermatologists will utilise dermoscopic features and textures, which are more accurate but requires sufficient training for the detection.

Asymmetry

Asymmetry analysis is a fundamental component in the early detection of melanoma because it often exhibits asymmetric shapes[3]. Meaning that the shape, colour and, texture matches asymmetrically more often in benign lesions. For example, as melanoma grows the central area begins to waste away leaving a hollow area covered by thin skin, showing dermoscopic features. As it grows the edges become more irregular producing an uneven shape often relating to irregular borders and asymmetrical shape. Diagnostic procedures have been developed to detect these uniuqe characteristics.

Bi-fold is a diagnostic procedure designed to support the recognition of 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.

6.2.2 Related Works

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[19] 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[19] 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 a SVM model.

6.3 A Novel Asymmetry detection technique using Bi-Fold, 3D Euclidean distance, and Superpixels

This section describes a novel machine learning technique for the automatic detection of melanoma

6.3.1 Bi-fold

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).

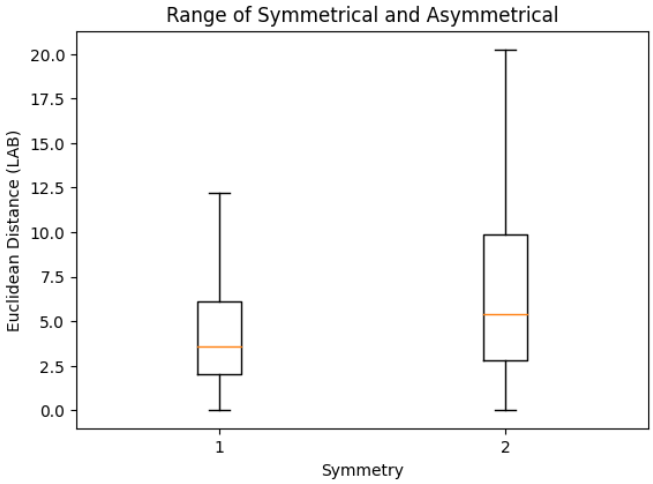

Figure 6.1: This diagram is a summary of the PH2 dataset after using bi-fold, euclidean distance of colour. The value on the right would be a threshold.

6.3.2 3D Euclidean Distance

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 6, then the lesion is asymmetrical.

The diagram shown below in 6.3.2 is a compilation of all the images within the PH$\hat{2}$ dataset showing the threshold range after applying bi-fold, euclidean distance of colour, but before applying the threshold. As can be seen a threshold of 6 covers all of the symmetrical values, but still roughtly covers half of the asymmetrical values. This demonstrates that the technique produces many false positives when regarding asymmetrical values. 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 holds poor results. Furthermore, there are a lot of fliers 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.

6.3.3 Superpixels using Simple Linear Iterative Clustering (SLIC)

Superpixels is an algorithm for grouping pixels into a grid format, but with flexible borders that can adjust to regions with similar features. Unlike the original technique averaging specific squares in a grid[19], they are segmented related to colour, texture, and other properties. The reason for using this technique is to increase boundry adherence and to group features that might otherwise be split into seperate groups. This overall improves the accuracy of the algorithm.

This technique uses simple linear iterative clustering (SLIC) algorithm and was first introduced by Achanta et al.[1]. This technique combines both k-means and graph-based segmentation. Firstly you define the desired number of superpixels as k and the approximate size of each superpixel as S, which is usually $S = /square(N/k)$ and N is the number of pixels in the image. Secondly for the centre of each cluster a search space is assigned to the cluster. For each group you measure the spatial distance which is the Euclidean distance between each pixel and the cluster center. Each pixel is assigned to the cluster with the nearest centroid. The cluster centres are then recalculated by taking the mean colour and position of the pixels assigned to each cluster. Followed by new pixels being assigned to the centroid relating to euclidean distance. This process is repeated depending on the number of iterations as i are assigned. From this point each pixel is assigned to a cluster.

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 ...

6.4 Experimental 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 10 by 10 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. The aim is to make a 10 by 10 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 6.4 demonstrates different borders when changing the C for compactness, where 100 generates a square grid similar to the original technique. The border becomes

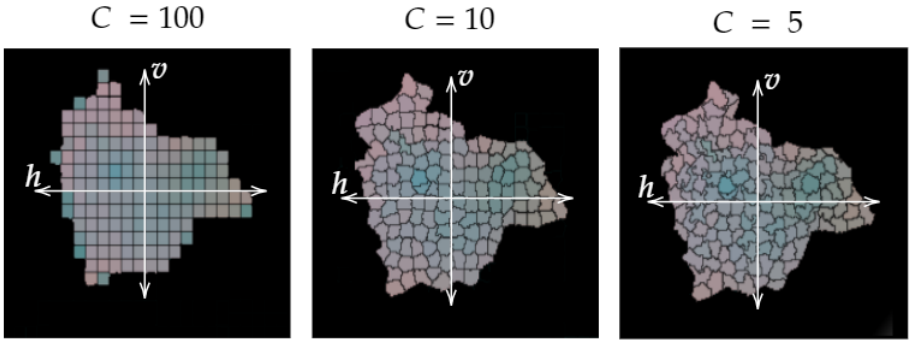

Figure 6.2: This diagram shows the skin lesion split relating to superpixels instead of averaging squares.

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 (relating to the line through the centre defined by the bi-fold) 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.

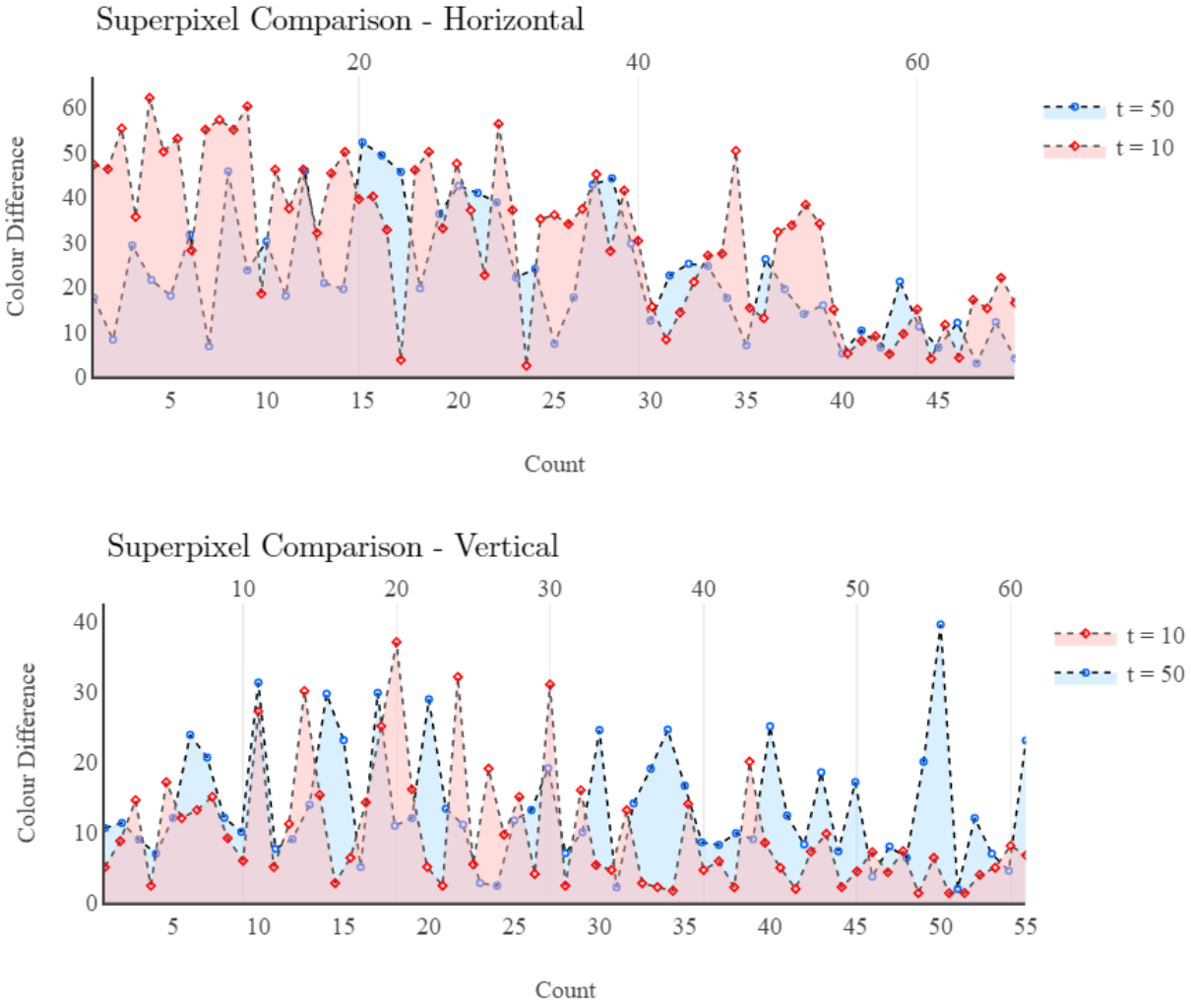

Figure 6.3: This diagram shows the difference between averaging squares and using superpixels, with the threshold of 10 implying curves and 50 being square. The horizontal colour difference is improved, making it more likely to be seen asymmetrical. The vertical comparison is roughly the same, except for removing a false positive of 40.

6.5 Border Detection Using Zernike Moments, Fractal Box-Counting, and Convexity

6.6 A Novel Colour Analysis Approach using Colour Ranges, and SVM

6.7 Results

6.8 Conclusion

Chapter 7

Dermoscopic features extraction

7.1

Chapter 8

Combined ABCD Rules and Dermoscopic Structures using Bayesian Network

8.1 Introduction

8.1.1 Challenges

Chapter 9

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 10

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] Alejandro Barredo Arrieta et al. "Explainable Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI". In: *Information Fusion* 58 (2020), pp. 82–115. ISSN: 15662535. DOI: 10.1016/j.inffus.2019.12.012.

[9] 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.

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

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

[17] 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.

[18] 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.

[19] 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.

[20] 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.

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

[22] 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.

[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 11

Tables

Chapter 12

Appendix