Melanoma Classification in a Clinical Environment

Transfer Report

University of South Wales

Elliot Naylor

2020

Contents

List of Figures

Chapter 1

Introduction

1.1 Introduction

Skin Cancer is considered one of the most invasive and serious public health problems with a mortality rate of 1,115 for non-melanoma and 2,347 for Melanoma skin cancer from 2015 to 2017 in the United Kingdom and are frequently increasing [1]. There are limited signs in diagnosing melanoma and 90% of patients can remain unaffected for the first 5 to 10 years before it is fatal. Melanoma has many similarities to a mole, making it easy to go unnoticed. Although, if found at an early stage it is the easiest treatable form of cancer through surgery, improving survivability.

To improve on diagnosing melanoma at early stages dermoscopes were developed that improve the visualisation of features by blocking out light sources. These features include streaks, dots and pigment networks that are otherwise invisible to the naked eye. The National Health Services (NHS) of Wales has recently been utilising these tools providing the perfect opportunity to streamline them for frequent use.

Alongside a dermoscopy a Computer-Aided diagnoses system (CAD) will be developed, which utilises algorithms to better visualise the important key features within an image. A CAD system is followed by a classification process providing a secondary opinion to a doctors diagnoses. The aim is to closely follow the same diagnoses procedures that might be used on a patient. This means the doctor can be guided along a proven diagnoses procedure without having prior knowledge. Therefore the doctor can interpret and learn from the procedure for a correct diagnoses without contacting one of the few dermatologists.

1.2 Skin Lesions

Skin lesions refers to a section of skin that has an abnormal appearance or growth compared to the skin around it. These are separated into two categories including primary which are abnormal skin conditions at birth consisting of birthmarks and moles. Secondary skin conditions are abnormalities obtained after birth which include a range of diseases, cancers and benign lesions. Many of these skin conditions are caused by a range of factors including age, UV light and genetics.
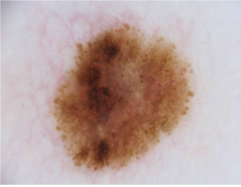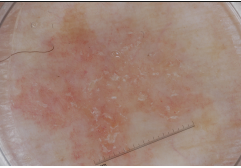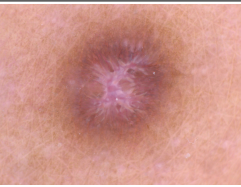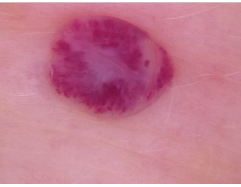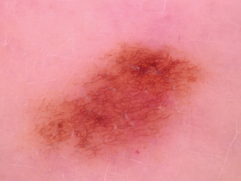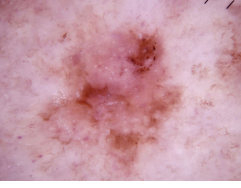
There are over 3000 known skin disorders in the area of dermatology some of which are so rare that it is unlikely for there to be any relevant images samples for analysis. This section holds a discussion on some of the most commonly found benign lesions and skin cancers that this project is concentrated around, the most important of which being melanoma.

1.2.1 Benign Lesions

Benign lesions are a non-cancerous areas of skin that are safe and unlikely to require any medical treatment. These are common and often share features with dangerous forms of skin cancers including Asymmetry, border and colour. As such, they are frequently given priority over other more serious lesions, slowing down the medical diagnosis procedure. One of the prime examples of this Serborrheic Keratosis (SK) which has similarities to melanoma. The following benign lesions discussed are Melancytic nevus (mole), benign keratosis, dermatofibroma and vascular lesions.

1.2.2 Skin Cancers

Skin cancer is one of the most common cancers that forms in the upper most layers of skin. Malignant cells divide without control spreading from the point of origin. There are many different types of skin cancers under categories of non-melanoma and melanoma. Non-melanoma refers to Basal Cell Carcinoma (BCC) and Squamous Cell Carcinoma, which are safer because there is considerably less risk of metastasis; spreading to other regions of the body. Regardless skin cancers are easily curable [2] if found at the earliest stages of development. Melanoma however has a higher likelihood of going metastatic, which can only take 6 weeks, making it deadlier than other forms of skin cancer.

Skin Lesion	Type	Description	Image Sample
Melanocytic Nevus	Benign	Melanocytic nevus or a mole is a harmless growth that often appears during childhood, birth or from prolonged exposure to UV light. They appear as pigmented clusters of cells representing a light to dark brown patch of skin that can protrude from the surface.	
Serborrheic Keratosis	Benign	The direct causes of these lesions are unknown, but they become more prevalent as the patient ages so genetics might play a role. These lesions are completely harmless, but the formation of them might be a symptom to a range of other problems.	
Dermatofibroma	Benign	This is a harmless skin lesion that forms within the dermis layer of skin that often related to insect bites or an immune system imbalance. However, being a red tumour like lesion these can be often misplaced with different skin cancers such as desmoplastic melanoma. [3].	
Vascular Lesion	Benign	Vascular lesions are harmless abnormalities otherwise known as birthmarks. These consist of three main categories of haemangiomas, vascular malformations and pyogenic granulomas. Vascular tumours can form which can be benign or malignant.	
Basal Cell Carcinoma	Malignant	Basal cell carcinoma (BCC) is one of the most commonly occurring skin cancer that develops on areas that have suffered from UV damage and develops. These lesions could take years to develop, making it the least dangerous type of skin cancer.	
Squamous Cell Carcinoma	Malignant	Squamous cell carcinoma (SCC) is caused by prolonged exposure to UV radiation and is prominent in people who sunburn easily and are over 70 years old. This type of skin cancer is not the most common but is one of the most devastating form of non-melanoma skin cancers.	
Melanoma	Malignant	Melanoma is the most common and dangerous forms of skin cancers relating to 4% of the population. As melanoma evolves it burrows down through the skin eventually spreading to other regions of the body including lungs, liver, bones, brain and intestines [4]. This is called metastases, which frequently results in death [5]	

Rules	Description	Example
Asymmetry	Melanoma is typically asymmetrical, meaning that if a line was drawn down the centre and the sides do not match there is a high likelihood that it is melanoma.	Image
Border	Irregular borders are normally a sign of melanoma. Normal moles have a smooth edge all the way around, while this has jagged or notched edges.	Image
Colour	Variations in colour are common in melanoma. Moles typically are a single shade of brown. As it evolves the colour red, white or blue can appear.	Image
Dermoscopic Structures	The diameter is for the analysis of structureless areas, pigment networks, atypical networks, dots and globules. These structures are required for the diagnoses of rarer melanoma and allow for a higher accuracy overall.	Image
Evolution	Any change in size, shape and colour could prove that an area is melanoma. Other factors include bleeding and itching.	Image

Table 1.1: Example of the ABCDE rules with a description and an image. These images were obtained from the ISIC 2019 dataset[8, 9, 10].

1.3 Diagnosis Procedures

Diagnoses procedures were developed to simplify the process and to pursue higher accuracy [6] results from doctors that are not trained specifically in the analysis of skin lesions. Some of these include menzies, 7-point checklist, CASH, ABCD, CHAOS, BLINCK, TADA and pattern analysis. The most popular is ABCD rules because of its combined simplicity and accuracy [?] when recognising between benign and malignant skin lesions.

The ABCD rules include Asymmetry, Borders, Colours, Diameter/dermoscopic structures and has been expanded to include Evolution[7]. Each feature is assigned a scoring system that can be used to diagnose skin lesions.

The 7-point checklist (7-PCL) is another effective[11] diagnosis method that considers the change in size, irregular pigmentation, irregular border, inflammation, itch/sensation, size (larger than 7mm) and oozing/crusting. This has similarities to the ABCDE rules, but, this method not only uses sight to assess a lesion, but considers the patients sensation and the feel of the lesion. This would be difficult to implement within a framework because the

relevant data describing the sensation might not be available.

Another methods called the 3-point Checklist (3-PCL) is a simple diagnoses method that aims to analyse the asymmetry, atypical Network and blue-white structures within a lesion. These features are very specific compared to the other methods and naturally with the lack of some common features within melanoma it seems likely for this method to be less effective. This method is described for use in a home environment, but atypical networks would be difficult for a non-doctor to recognise.

The Menzies method is a complex technique that has the highest accuracy[7] for diagnosing melanoma using negative and positive features. Negative feature include checking whether the lesion is symmetrical or of a singular colours, either of which define it is non-melanoma. Positive features include blue-white veil, multiple brown dots, pseudopods, radial steaming, scar-like depigmentation, peripheral black globules, five to six colours, mutliple blue-gray dots, broadened network. This method is naturally difficult to utilise within a clinical environment because of the expertise required recognise the features. Another problem is that a diagnoses needs to be fast because of the quantity of patients being processed, but, the number of rules does not support this.

Pattern analysis is a simultaneous analysis of all components within a lesion. By far the most difficult method, requiring time and great expertise to utilise in a clinical environment. This is likely to only ever be used by expert dermatologists that have possibly analysed hundreds of lesions. The method consists of two categories including global features and local features. Global features describe the structure of the lesion which could include multicomponent patterns, unspecific pattern (Structureless or irregular), parallel patterns (Ridges, palms and soles). Local features analyses the components including atypical pigment networks, dots/globules, asymmetrical, irregular streaks, five or six colours, blue-white veil, etc. If the doctor can recognise each component and the correlation between them, this has the potential to be the most effective method for diagnoses.

Another complementary sign is the "ugly duckling" approach, which urges for a comparison between other moles on the patient's body. The one that is unlike any of the others is often suspect for malignancy [12]. This is an interesting method as it requires less prior experience to correctly diagnose the patient. However, there are no direct examples proving the effectiveness in a clinical environment.

The goal of these methods is to simplify the process to improve the performance and accuracy of a diagnosis. However, the rules are not always effective against different types of melanoma. Examples of these are Amelanotic melanoma which is colourless and lacks the

Criteria	Description	Score	Weight
Asymmetry	In 0, 1, or 2 axes; assess not only contour, but also colors and structures	0 - 2	x 1.3
Border	abrupt ending of the pigment pattern at the edge in 0 - 8 segments	0 - 8	x 1.3
Colour	Presence of 6 colours (white, red, light brown, dark brown,. blue-gray, black)	1 - 6	x 0.5
Dermoscopic Structures	Presence of structureless areas, pigment networks, atypical networks, dots and globules	1 - 5	x 0.5

Table 1.2: Describes the scoring system for Total Dermoscopy score (TDS) for diagnosing melanoma[15], which is added up as: [(Asymmetry x 1.3) + (Border x 0.1) + (Colour x 0.5) + (Dermoscopic structure x 0.5)]

pigments required for analysing border and colour [13]. Another is called nodular melanoma which appears darker and lacks in colour variation (black) or dotted features that are often required diagnoses. This is likely why the ABCD rules was modified from diameter to dermoscopic structures, providing more than enough information for an accurate diagnoses.

1.3.1 Total Dermoscopy Score (TDS)

The Total Dermoscopy Score (TDS) is a defined method for assessing the score of the ABCD criteria. Scores less than 4.75 indicates that the lesion is benign, 4.8 to 5.45 indicate a suspicious lesion and anything higher has a high likelihood of being melanoma. Each rule is assigned a score and a weight which are used to assess the structures. Such scoring systems have been implemented into CAD systems successfully[14].

1.3.2 Computer-aided Diagnosis (CAD)

Most cases are first assigned to a general practitioner (GP), which do not have training in dermatology and skills are more globally oriented. This means they might struggle to recognise some of the skin features needed for a correct diagnoses. In these scenarios a secondary opinion is required, but, there are often limited dermatologists on hand, which can often lead to a misdiagnoses. To combat this CAD systems were developed to assist these doctors with an automatic secondary opinion; improving the likelihood of more frequently correct diagnoses. The main types of CAD systems are broken down into a smaller criteria

including dictionary based features, clinically relevant features and deep learning[16].

Bag-of-Features approaches involve further annotating of training data and assigning a word to each section of an image, which could relate to the ABCD rules or dermoscopic features. This can be achieved using a saliency map to separate an image into super pixels and assigning relevant words to each region. This is followed by classifying each area relating to the colour and texture descriptors [17]. However, these answers are often found to be ambiguous and difficult to interpret. This is because each area is assigned a word or multiple words, but do not directly show or enhance feature. Therefore the person using it would need the relevant experience to recognise the features effectively.

Clinically relevant features describe systems that follow a segmentation, features extraction and classification[18] approach. These steps are designed to segment the lesion to quantify the ABCD rules[19, 20] and other dermoscopic features. These hand-crafted features using edge filters can provide a basis for a diagnoses, but, this is sometimes met with scepticism because it is difficult to quantify the features for medical meaning. However, a recent paper has improved interpretability by analysing each feature with an SVM before merging the results using Bayesian fusion. This means each result is visualised and assigned a score similar to TDS possibly improving the diagnoses accuracy. Furthermore, this is a risky area to research because it is beginning to converge with texture extraction and deep learning methods; where clinical features are separated during classification through means of explainability. This means that using hand-crafted features could become irrelevant in the next couple of years.

Texture extraction methods utilise segmentation and border to visualise the ABCD rules; D being Diameter instead of Dermoscopic structure. Smaller dermoscopic features are not made interpretable in these examples and the area is instead extracted as a texture as GLCM, LBP, etc. Therefore all features are analysed at a given time with exceptionally high accuracies. This holds the benefit of having a higher accuracy than extracting clinically relevant features, but holds only the bare minimum of visualising clinically relevant features to doctors. Biases will therefore be harder to recognise possibly leading to rarer lesions being misdiagnosed.

Deep learning models are the current highest accuracy techniques that consist of training models directly from image samples. These methods have been improved to produce a visualisation through heatmaps and loosely visualising some relevant features, but the results cannot be retraced [21], making them un-trustworthy. While there are methods being developed that allow for explainability, the results are still not interpretable[22] enough for use

within a clinical environment. Therefore many useful machine learning algorithms should not be utilised within a clinical environment. Furthermore, this style of technique requires a large amount of training data which is not always available, especially for rarer skin conditions. This means there is a likelihood of the results being less accurate in rarer conditions with no means of proving otherwise.

1.3.3 Explainability

The European General Data Protection Regulation (GDPR and ISO/IEC 27001) came into force in May 2018, making "black box" approaches difficult to utilise in a medical environment. Although this is not a ban, the outcome of the results needs to be re-traceable [23]. These techniques are consistently relied on within medical imaging and without explainability doctors cannot interpret, modify or learn from the provided diagnoses. As such, when using machine learning algorithms, a trusted form of XAI or CAD system must be utilised.

The reasoning behind retractable results is because society is becoming reliant on "black box" approaches that the results of which cannot be analysed. For example, in scenarios where a machine learning model is trained from mostly common lesion samples with a specific skin colour, it might have a lower accuracy against other rarer variations. As these algorithms are not always retractable and there is a lack of image samples within these groups it is difficult to assess whether these algorithms work effectively in rarer scenarios. Therefore, it is ethically and morally incorrect to utilise such methods within a clinical environment. To avoid malpractice the doctor must be provided with relevant and interpretable results to be considered, but not depended on for a diagnoses.

1.4 Datasets

A dataset is a collection of image samples that are often used to compare and validate performance and accuracy of algorithms described in journals, without implementing each technique needed for the comparisons. Furthermore, machine learning techniques require large amounts of image samples for training and without the same amount of relevant images it would be impossible to reproduce some algorithms. For example, one of which utilises a private dataset of 129,450 images consisting of 2,032 different diseases [24], achieving the highest accuracy in classification so far. As a comparison, the largest public dataset called ISIC 2019 has 25,331 images containing 8 different diseases. This is one of the biggest concerns in medical imaging as comparisons with public datasets are not always made,

Dataset	Image Samples	Image Type	Availability
ISIC 2019	13,000	Dermoscopic	Public
BCN20000	19,424	Dermoscopic	Public
HAM10000	11,788	Dermoscopic	Public
NHS	unsure	Dermoscopic	Private
Atlas	80	Dermoscopic	Public
MED-NODE	170	Macroscopic	Public

Table 1.3: List of datasets that are used throughout this project. Some of these datasets share images, making the total amount of unique images 25,581

bottlenecking active research within a topic [25]. For this reason, competing against deep learning models has become less prominent and there has been a greater focus on other topics including image synthesis, explainable AI (XAI) and some CAD systems.

Other datasets including MED-NODE [26] provide macroscopic image samples for comparing and creating techniques without a dermoscope. Macroscopic images could have large variations in noise, angles and lighting conditions where the algorithm might fail. For example, the training process of a machine learning algorithm is specific to the type of data it analyses. If the model is used on an image that has a different variation in lighting and angles the algorithm is likely to fail. As such, rules must be followed by the doctor taking the image to insure accurate results.

1.5   Challenges - Project Scenario

This project has some unique components being that the traditional diagnostic procedure used by the dermatologists are based around touching and feeling through pinch tests. This might include an analysis of roughness, firmness, bleeding, etc. Other information that is considered are age and a lesions location on a body. These components are rarely considered within a CAD framework. A possibility would be to analyse the lesion based around the ABCD rules to separate the component into a benign or malignant skin lesion. The framework would then request additional information through a pinch tests, age and location to narrow the results down to a type skin condition. This additional information would prove valuable to the dermatologists examining the diagnoses at a later date.

1.5.1 Clinical Environment

In a clinical environment, sessions are often short because of the number of patients being diagnosed at a given time on a range of topics including dermatology. For many general practitioners it might be their first attempt at diagnosing a patient with a skin related issue. As described previously many benign lesions have a range of similar features to dangerous cancers, the most common of which are moles and melanoma. It is unrealistic to assume that each doctor has the relevant experience to correctly diagnose a skin lesion. Diagnoses methods including the ABCD rule were created to find some of the most common and dangerous skin lesions such as melanoma. However, there are over 3000 different skin lesions making it unlikely for a doctor to diagnose each effectively. Therefore, a secondary opinion is required from a dermatologist to improve the accuracy of the diagnoses.

Assuming the dermatologist prioritises undiagnosed lesions they can still be left weeks before being diagnosed. However, the main complication is when a general practitioner incorrectly label a skin lesion, leaving it unchecked by a dermatologist indefinitely. This is because it is often impractical to use their limited time to be looking through a range of already diagnosed skin conditions. This is where a computer aided diagnoses system (CAD) can provide a secondary opinion to the doctor by providing tools that would standardize and improve the accuracy of the diagnostic procedure.

The final diagnoses should be defined by the doctor and not the algorithm. For example, the algorithm should be built to provide relevant data supporting the classification process relating ABCD rules or other smaller features. These should be built to persuade and even teach the doctor of better diagnoses methods, without a too convincing diagnosis.

1.5.2 Data Samples

There are two different environments in which images of skin lesions have been captured which includes hospitals with access dermatologists and surgeries with general practitioners (GPs). Available equipment varies between these locations and general practitioners will often use digital cameras to capture skin lesions while dermatologists have access to dermocopes and other specialised equipment. Therefore the data samples will have variations between dermoscopic and macroscopic images depending on where they were captured.

One of the underlining problems with using macroscopic images for analysis is that they are less likely to contain smaller features which can improve the accuracy of a diagnosis[]. Furthermore, melanoma can contain 6 different colours and, in a room too bright or too dark

these colours and other details can be difficult to interpret, leading to an incorrect diagnosis. Most data samples available are

There was no mention of rules used to take these macroscopic images, meaning they might have large variations in lighting, angles and blurring. This makes it substantially difficult to utilise any illumination or noise removal techniques before using them to train a classification model. The training process is not controlled, and any intervening elements might be considered features, worsening the accuracy. It might be possible to train an algorithm only using the limited dermoscopic examples, which would have improved accuracy [27] compared to utilising macroscopic images.

1.6 Summary

Melanoma is one of the most dangerous skin cancers because it is difficult to differentiate between other skin conditions. For example, diagnosing with the ABCD rules is simple and allows for a fast and accurate diagnoses of the most common types of skin lesions, but, is inefficient at diagnosing some types benign lesions (SK) and rare types of melanoma[28]. Alongside the support of diagnostic procedures and Computer-Aided diagnostic (CAD) systems the accuracy continues to increase.

The goal of this project is to improve the diagnostic accuracy of skin lesions within clinical environments through the development of a CAD system. This system would ideally be capable of producing results that can be retraced and understood by the doctor. One of the central problems with this project is that surgeries lack access to specialised equipment including dermoscopes. This means the system has to be developed for and using macroscopic images. These images could potentially contain infinite variations in lighting directions, colour, shadows, blurring, etc. These will hide structures and alter colour intensities. Images are then less likely to contain reliable and consistent features for the development of machine learning algorithms. This means there will be a decreased accuracy compared to the analysis of dermoscopic images.

1.6.1 Aim

- Develop a Computer-aided System (CAD), capable of aiding and clinicians in the classification of skin lesions.

1.6.2 Objectives

- Detect and visualise ABCD rules.

- Detect smaller features including pigment networks, milia-like cysts (MLC), streaks, globules and blue-white veil.

- Classify defined features within a machine learning algorithm.

1.7 Contributions to knowledge - Ideas section for now

Many of the best border extraction techniques capture the surrounding area of the lesion perfectly, but don't enhance the border for smaller irregularities. This can improve visualisation to doctors of border irregularity rule and segmenting with these features can improve the accuracy of an SVM [29].

It would be possible to extract features relating to starry or cloudy milia-like cysts before classification. Seborrheic Keratosis (SK) was found to have 99.1% sensitivity against melanoma for cloudy (larger) milia cysts [30], meaning that locating these could improve the accuracy between these skin conditions. The papers showing this technique are very old and don't test this with an SVM or against a range of skin conditions. Another paper [31] specifies that a larger number (more than 3) and a smaller MLCs can indicate SK from melanoma. However, MLC are almost invisible under a non-polarised light of a dermoscope.

Improved visualisation of menzies features and colours would be useful.

Chapter 2

Literature Review

2.1 Introduction

Diagnosing skin lesions is a time-consuming task that requires high-level expertise and knowledge for consistent results. There are frequently a lack of dermatologists available and training new doctors is undesirable because of the cost. Computer-Aided Diagnostic (CAD) frameworks were developed to assist less experienced doctors and has been proven to effectively improve the accuracy and efficiency within clinical environments[32]. However, many frameworks are not implemented in clinical sessions because they provide a parallel diagnosis which prevents the clinicians ability to understand and learn from the results. An example of one of these techniques utilises a machine learning algorithm called Convolutional Neural Networks (CNN), which have reached the accuracy of professional-level dermatologists[24]. While highly accurate, these systems lack a clear description on the relevance of individual features that support the diagnostic procedure.

This review focuses on discussing techniques that automate the diagnostic procedure. An example of this is the ABCD rules[33] which includes asymmetry, border, colour, diameter. Another method utilises dermoscopic features to locate globules, reticular structures streaks, milia like cysts, etc. These methods are followed by feature combination methods including Bayesian Fusion to combine the results[34]. The most important correlation in these techniques is that they both algorithmically generate hand-crafted features. Followed by classification, allowing for the entire process to be visualised before the classification process providing proof of the data used for each technique.

2.2 Methodology

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

1. Image Acquisition – The method in which the image data is captured. This might include a dermoscope, macrosocpic images, stereoscopic dermoscope and High-Dynamic Rendering (HDR). Each method will effect the features differently and will effect the accuracy of classification. The most effective and popular method is the dermoscope.

2. Pre-Processing – Image samples are altered to enhance features, colours and remove obstructing components from the image. These components include hair, scab and other skin components that are not being analysed. When using macroscopic images specular removal is utilised.

3. Segmentation – Finding the region of interest allows for a higher accuracy classification technique preventing skin features from being considered. Capturing the area can be used to analyse clinical features including asymmetry and border.

4. Feature Extraction - Gathering features through filtering and morphology approaches. This includes capturing and analysing features related to the ABCD rules and other relevant features. This allows for results to be retraced and analysed

5. Classification – Features from the segmentation and feature extraction process are are classified and combined.

2.3 Image Acquisition

Unaided visual inspection of skin lesions has proven to be suboptimal because of obstructing reflections and variations in lesion size. For this reason, tools have been developed that have improved the sensitivity and specificity when analysing and recognising features. The most common of these tools is called an epiluminescence microscopy, otherwise known as a dermoscope. This removes obstructing elements improving the reliability of the images.

2.3.1 Dermoscopy

Dermoscopy is a non-invasive technique for examining skin structures that are not visible to the naked eye. It was found to be especially valuable at diagnosing patients with melanoma

at an early stage [26, 35] with confidence reaching 95% [36]. The structure includes streaks, dots, pigment networks and blue-white areas, which are more easily analysed with this tool [37, 38].

The technique requires the placement of an oil and direct contact with the surface before inspection with a specialised tool. Dermoscopes can vary in lighting conditions including non-polarized (NPD) and polarized (PD) lighting conditions. Both techniques differ in accuracy between different types of skin cancers. For example, when considering non-melanoma PD highlights the presence of blood vessels and has increased sensitivity when analysing basal cell carcinomas (BCC). However, NPD is vital for locating Milia-like cysts which are important to recognising seborrheic keratoses [39]. Hybrid dermoscopes are popular because they can toggle between both PD and NPD.

Many papers today utilise a dermoscope with outstanding results to accuracy[]. Without this, some defining features would be invisible, decreasing the likelihood of a correct clinical diagnoses. These are widely available today and are usually cheap to obtain. Including this many are designed as attachments onto mobile devices allowing for images to be captured and immediately analysed through CAD systems. This can cater towards the short clinical sessions and allows for data to be saved easily.

2.3.2   Macroscopy

Macroscopy describes analysis without using screening tools that only depending on the naked eye. Many skin analysis techniques utilise dermoscopy tools, but they are not always obtainable in a dermatology environment making image capturing with standard cameras of great interest because of their wide availability. However, utilising macroscopic images requires further pre-processing because of the presence of specular reflections [40, 41] and background data.

One technique that uses macroscopic imaging achieved an 81% accuracy making it comparable to using dermoscopic imaging [42]. However, the images need to be taken under good conditions for accurate results when considering the noise, angle and lighting conditions. Anything outside of this range is likely to decrease the accuracy, producing misleading and inaccurate results

This often means that in a clinical session the doctor must capture an image in good conditions otherwise the results could be misleading or inaccurate. Different times of the day can even affect the lighting conditions, meaning there is always a likelihood of the algorithm being used under bad conditions. Ideally macroscopy should only be focused on if there is

no other data available, otherwise it is more substantial to focus within an area that uses screening tools with an increased reliability and accuracy.

Macroscopy images has not achieved much attention in recent day because it has an increased complexity than other techniques [43]. This is because without a dermoscope certain features are invisible. It also involves the correction of lighting issues in the image which can produce noise which can decrease the quality of samples. There are many intervening elements that can completely alter the results. This means the accuracy is bound to be less reliable than techniques using a dermoscope.

2.3.3 Spectroscopy & Others

Other screening tools for analysing skin are available through spectroscopy [44]. Specroscopy is used for the dispersion of an objects light into component colours. Allowing for physical properties to be analysed including mass, luminosity and composition. Comparisons between dermoscopy and spectroscopy have been made, but it is difficult to tell whether the results provide any further relevant data to dermatologists [32]. However, there is protentional for the combined method to be used within deep learning with an improved accuracy [45], but relevant training data is lacking.

As mentioned previously direct contact with a dermoscope is required for the accurate analysis of lesions. This is possibly the reason stereoscopic analysis is less popular, as more details relating to the shape can be captured but the benefit of having direct contact is lost. However, there are methods that reconstruct and estimate the shape and depth of a lesion through a traditional dermoscope [46].

Using custom tools is often a way to increase the accuracy of diagnoses but developing and buying them is often not possible for a non-profit organisation like the NHS. It is more reliable to use frequently used and proven technologies such as a mobile device and a dermoscopic attachment.

2.3.4 High-Dynamic Range

High-dynamic range (HDR) uses tone-mapping which aims to balance illumination intensity throughout an imaged. Changing the shutter speed on a camera can change variations in light intensity making the image appear brighter or darker. By combining images captured with a range of shutter speeds using tone mapping an image can be produced with balanced lighting conditions throughout. This could prove valuable to macroscopic images because the

presence of specular reflections is largely where accuracy is lost. However, this technique is known to produce noise potentially removing feature components from an image that would prove valuable to classification.

Using HDR facilitates in identifying the structure of the lesion [47]. However this method increases contrast which destroys valuable clues to recognising the melanoma or non-melanoma [48]. Another article demonstrates that it can be utilised to recognise hypopigmented lesions [49] that could improve the accuracy of a diagnosis.

2.4   Segmentation

Medical imaging segmentation is the process of partitioning an image into multiple segments for easier analysis. These areas can be separated manually by a dermatologist to create a ground truth or by an algorithm capable of separating the area relating to colour and texture. A segmentation mask consists of every pixel being assigned a category, labelled for their characteristics. Skin cancer segmentation is usually separated into lesion and non-lesion. Furthermore, areas can be separated into more partitions and can share the same space.

Many neural network approaches to segmentation utilise Fully Convolutional Networks (FCN)[50] which describes a CNN using a transposed convolution or deconvolution. This is achieved by feeding the network images of lesions and masks. The network then gathers features relating to the mask location and can use these spatial features to generate further masks for lesions. There are dozens of examples of this, one of which is SegNet [51], which was not originally designed for skin lesions but is proven to be effective with skin lesions. While these methods are useful they do not suit the CAD criteria as the results need to be obtained using non-machine learning techniques.

Saliency-based segmentation techniques [52, 53] make use of saliency mapping to simplify or change the data into something easier to analyse. This can be achieved by generating superpixels through a method called Simple Linear Iterative Clustering (SLIC). These have a low computational overhead and allow for the analyses and segmentation of an area relating to colour and texture. Generating a saliency map then locates unique areas within an image relating to texture and colour. Many other techniques improve the accuracy within lesion segmentation by applying a threshold before locating areas of interest [54, 55].

Another method uses laplacian filters, binarization and morphological operations to segment lesions. This heavily concentrates on pre-processing to remove [56] blurring and other

anomalous data. Stickers used to cover anomalous lesions and the lens shape are detected and removed to ensure high accuracy results when classifying.

Otsu binarization technique is a thresholding technique that is effective at locating the border of a melanoma after segmentation [?, 57]. One of the techniques uses the saliency-based segmentation approach to capture the area, followed by an Otsu threshold to capture the border more accurately.

One papers aims to capture an irregular lesion border and was found to produce more accurate results when classifying, compared to a perfectly accurate border. This is achieved by using a Local Binary Pattern Clustering (LBPC) [29] followed by subtracting it against a gray scale to find the border. The method goes on to describe classification methods using an SVM or FNN presenting the extracted border in a readable form improving the accuracy and visualisation of the results.

Interestingly, many of the machine learning approaches lack the ability to create an accurate border around a skin lesion. This proves the importance of segmenting features using thresholding and clustering techniques such as LBPC because of the higher quality results and better explainability. The border irregularity and asymmetry can be classified using an SVM alongside other extracted features, for a final result.

2.5   Handcrafted Features

Hand-crafted features describes the process of using pre-defined features created using filtering and morphology algorithms. Alternatively non-hand-crafted features are automatically defined during the training process through a CNN or deep learning. This technique can be described as a "black box" approach because the process is encapsulated and the features used for classification are difficult to interpret. Alternatively, the benefit of using hand-crafted features is the assurance that the training models only utilise appropriate data of an image for classification, instantiating further trust in the use of the technique. This allows for the machine learning process to be developed around specific diagnostic procedures such as the ABCD rules. This is instead of providing a direct diagnosis of a skin lesion without any explanation, which is currently not acceptable for use in a clinical environment.

2.5.1   Asymmetry

Asymmetry refers to drawing a line down the middle of a skin lesion and comparing the two halves to see if they match. This is done on both the horizontal and vertical axis and can be

a warning sign of melanoma. Asymmetry can be measured using the shape[58], colour[59] and texture[60].

Measuring the shape involves using the mask generated through the segmentation process that captures the area of the skin lesion[58]. The centroid and rotation of the skin lesion is found using moments. A line is then drawn across the centroid into two equal halves of both vertically and horizontally. Each vertical and horizontal side is compared is subtracted against the opposite opposite half finding which areas of the skin lesion are asymmetrical in shape.

To measure the colour asymmetry [59] of the skin lesion it is split into 20 x 20 pixels and converted into the LAB colour space where the average colour is computed for each block. The euclidean distance of the colour is measured against the principal block on the opposite half of the skin lesion. A threshold is set that represents the colour difference visible to the human eye in the LAB colour space[61]. If over half of these colour differences are over the threshold the lesion is considered colour asymmetrical. If there is no opposite side of the lesion the result is ignored as this method is only meant for measuring colour. To prevent colours being separated regarding brightness, luminance is separated and calculated using the same euclidean distance technique.

Measuring similarities in texture can be achieved by using a SIFT based similarity and projection profiles[60]. SIFT is scale invariant and is therefore useful for texture components with varying texture quality. The skin lesion is split into vertical and horizontal across the center into four halves. Texture components on the proportional halves are compared with a higher value representing similarities. This is followed by a projection profile in the x and y directions then comparing the histograms. These results are then used to train a decision tree and was found to have an 80% accuracy at finding asymmetry.

2.5.2 Border

Border refers to the area around the skin lesion that is measured for irregularities by splitting it into 8 equal sections through the centroid. Each border within a section is checked for a tight corner which is then considered an irregular border. When conforming to the TDS a score is added for each section of the border ranging from 0 to 8.

To measure border irregularity contours were found by splitting the skin lesion into 8 areas around the center. Followed by calculating a fitting error for each area. If the error is larger than the (0.05 x contour) that area is considered irregular[59].

Another paper calculates the compactness of each border by first calculating the contour

around the area of the lesion containing x and y positions. Compactness is then calculated by measuring the space between each position. The tighter the curves and corners the increased amount of contour positions revealing irregular borders within a segment. Border irregularity is then calculated by combining them into a irregularity index from all the segments[62].

Fractal dimensions (FDs) is a statistical index measuring the detail in a pattern changing the scale in which it is measured. One of the techniques is called box-counting and will increase an index value if there is more detail around the edges of the skin lesion. The higher the value means there is an increased chance of it being malignant whereas the lower the value means it is likely benign. These have recently been used within machine learning alongside zernike moments and convexity measurements for a high accuracy border irregularity classification[63]. Although results appear to be ambiguous because the output would be either "irregular" or "regular" border (not relating to TDS), and would not be retractable because of the use of a CNN.

2.5.3 Colour

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

Finding colour variations has been achieved by calculating the normalized standard deviation of the red, green, and blue components[?]. The normalization process improves the recognition of higher normal skin pigmentation which would show pigmentation levels making comparisons easier between different skin lesions.

Mimicking the method in which dermatologists locate the areas with colour and the size of those zones a method was proposed[?]. This method utilises a joint learning using kohonen map and k-means clustering. For each skin lesion in the dataset 5 pixels are chosen at random to create a 5x5 kohonen map represented by 25 neurons in a neural network. The colour variations are then determined on a 25-dimensional vector obtained by finding the proportions of pixels projected onto each of the 25 neurons of the map. K-means is then used to classify the skin lesions set by the amount of colours found by the dermatologists. In this scenario only 4 colours were present in the dataset while there could be 7 colours. Eventually the colour components are represented as a 42-dimensional vector and is passed into a KL-PLS based classifier for the detection of variations in colour at 66%. This is an

interesting approach, but it can be argued that the use of a neural network to first represent the colour distribution of the dataset would prevent results from being retraced easily. Which is the reason handcrafted features are being used for this project.

Another method locates the amount of colour variations by first converting the image into the LAB colour space matching the colour ranges that can be perceived by human eyes. Followed by finding the average colour distribution of the dataset and assigning each colour as a threshold range. The euclidean distance between each colour threshold is compared with the each pixel colour [59], finding the closest matching colour of the six variations. To prevent the classification of small abnormalities the area of the colours found are measured to find if they are more than 5% of the area of the skin lesion. This approach uses a colour range for white, light brown and dark brown. However, for the other colours there is a single threshold value, which would be unlikely to cover the ranges of the colours including red, blue-gray and black.

2.5.4 Dermoscopic Structure

2.6 Classification

Classification is the process of locating key features within an object of an image allowing it to be detected frequently without human support. These methods have shown the potential of diagnosing different types of skin cancer with a high accuracy alongside improving the performance of a dermatologist [64] and are based on the ABCD criteria [65, 66, 67]. These methods in detecting skin cancer are increasing in potential reaching a higher accuracy than professional dermatologists [24].

Many techniques within medical imaging analysis makes use of a technique called Convolutional Neural networks (CNN). This change started when deep neutral networks began surfacing and outperforming other established models [68]. They are designed to automatically learn spatial hierarchies and features (non-handcrafted) through a set of building blocks called convolution, pooling and fully connected layers. These allow for flexibility [69] when creating a model by rearranging layers into what is most suitable for detecting a particular object. The disadvantage is that such models need thousands of relevant training data for a high accuracy, which is sometimes out of reach.

Fully Convolutional Residual Networks (FCRN) is a technique that combines both segmentation and classification into a single network improving performance in both techniques [70]. This is achieved by creating a very deep network that consists of 50 layers. How-

ever, these networks usually have more parameters and require greater amounts of training samples to prevent overfitting.

Using data in conjunction with image samples can improve results. One paper describes the combination of a CNN and the results of 112 professional dermatologists using gradient boosting [71]. This was achieved by sending questionnaires to dermatologists which were organised into different skin conditions alongside certainty ranging from 0 to 10. These human decisions supported in outlining common mistakes and improved the accuracy of the results by 1.36% reaching 82.95%. Moreover, using metadata alongside images to train a CNN frequently increases the accuracy [72, 73, 74].

SVM

Recent classification is often achieved using a wide range of machine learning techniques, but when using extracted features mainly Support Vector Machine (SVM) are used. There are some papers that achieve similar results by passing Gabor filtered images into CNNs [75, 76], ANNs [77] or even FCNs.

Reproducing many of these algorithms is usually impossible because the results are directly related to the dataset used. Although there are new variations in techniques it is largely considered finished because the accuracy cannot be obtained or improved without access to extensive data, such as Googles transfer model. Moreover, the most accurate model created so far [24] is available within a mobile application. Meaning that further relevant image data is being gathered making it difficult to compete with the accuracy of skin cancer algorithms.

2.7 Discussion

When attempting to automate the ABCD rules the segmentation performance is depended on to find the area of the skin lesion. Both asymmetry and border irregularity depend on the border of the lesion to be measured effectively.

Segmentation performance must be highly accurate. For example, by visualizing the results calculations can be made on the border irregularity or asymmetry described by the ABCD rules, proving the algorithms are working as expected. This is essential to the projects objectives; which are to visualise clinically relevant features within an image.

Measuring asymmetry is a substantial method for classification because of the large selection of techniques utilising shape, colour and texture used to measure the asymmetry of a skin lesion. These techniques utilise unique components and each technique can be conformed

to the TDS with minor changes. It would be interesting to utilise these components together for a possibly higher accuracy technique. Another approach to texture would be to utilise bag of features instead of SIFT to see if improvements can be made. However bag of features has been scrutinised for being ambiguous within a clinical environment[].

Border analysis is one of the highest accuracy techniques using fractal dimensions, zernike moments and convexity for analysis. However the output result is ambiguous because it returns "irregular" or "regular". It would be beneficial to split the data into 8 border components and pass these into a CNN for classification separately. This would conform to the TDS and could prove more valuable to the clinicians using it.

Finding colour variations appear to be lower in accuracy. This is because many datasets use records from a number of different doctors making the results subjective and inconsistent. This makes testing and proving the accuracy of the technique an especially difficult task because of the lack of accurate data. However, the quality of datasets proves the necessity for adopting a standard for more consistent results between different clinicians.

2.8   Summary

Chapter 3

Development

3.1 Introduction

This section includes examples of a number of algorithms and the method and how they work.

[y=.08cm, x=.18cm,font=] (0,0) − coordinate (x axis mid) (35,0); (0,0) − coordina""""te (y axis mid) (0,60); in 0,5,...,35 (,1pt) − (,-3pt) node[anchor=north] ; in 0,10,...,60 (1pt,) − (-3pt,) node[anchor=east] ;

plot[mark=o, mark options=fill=red] file $\text{div}_h orizontal.data$;

[y=.08cm, x=.18cm,font=] (0,0) − coordinate (x axis mid) (35,0); (0,0) − coordinate (y axis mid) (0,60); in 0,5,...,35 (,1pt) − (,-3pt) node[anchor=north] ; in 0,10,...,60 (1pt,) − (-3pt,) node[anchor=east] ;

plot[mark=o, mark options=fill=blue] file $\text{div}_v ertical.data$;

Chapter 4

NOT USED

To validate the results further the process was split up into border segmentation, colour extraction, texture extraction[20] and feature extraction. Border irregularity captured through segmentation (described previously) allows for border irregularity to be visualised, the colour extraction alongside border can analyse the asymmetry and colour. Followed by the feature extraction visualising features described in dermoscopic structure. This means that the ABCD rules can be fully visualised before the diagnoses through a means of classification, making the results retractable. This only leaves texture extraction which enhances the textural data of the lesion to allowing the lesion to be classified by an SVM, which has proven effective with dermoscopic structures. This has been tested extensively to prove which texture extraction methods reveal the most useful data when classifying the area of the lesion.

4.0.1 Texture Extraction

Texture extraction techniques enhance the texture components of an image improving the classification accuracy. They can be extracted from an image using a range of techniques including LBP, HOG[78] or heralic features [20, 78], which describes 14 texture extraction methods; mentioned more in depth below. There are a great deal of papers that aimed to find which one of these techniques has the most useful textural information when training an SVM. This textural data of entire lesions has been used to classify a lesions with 90?% accuracy and has proven valuable at extracting specific dermoscopic structures; some of which are pigment networks[79], globules and streaks.

Local binary pattern (LBP) is a dated but efficient texture operator that gathers a binary of the neighbouring pixels. The binary number is then converted to decimal and assigned to the center pixel. This in turn creates an enhanced version of the image improving on the

classification results.

Histogram of Oriented Gradients (HOG) is a feature descriptor that has found uses in object classification. The goal of this technique is to enhance the image to better describe the intensity of gradients and edge directions. This is usually used to detecting people within an image but has recently found a usefulness in skin lesion analyses.

Heralick features are a newer technique that describes a range of feature extraction methods, designed to enhance some of the most important features in an image. These techniques include Gray-level Co-occurrence Matrix (GLCM), Haar Wavelet Decomposition, Gray Level Histogram (GLH) and others. This was designed to categorise and evaluate texture extraction methods within a classification process to find which method provides the highest accuracy.

Out of the techniques described in Heralick features, GLCM was found to be the most effective with an SVM classifier achieving the highest accuracy of 92% without separating individual dermoscopic features.

4.0.2 Colour Extraction

Colour should be calculated representing the area and shade which can be first achieved by averaging the colour to reduce noise within the image. This is followed by masking unrelated features relating to ulcer and crust and followed by k-means to separate areas by colour intensity.

Hue of each pixel - (Skin cancer diagnostics with an all-inclusive smartphone application)

Normalized standard deviation of colour components - (Combination of features from skin pattern and ABCD analysis for lesion classification)

Kohonen map - Euclidean distance between each pixel and suspicious colour to separate them into groups - (Classification of malignant melanoma and benign skin lesions: Implementation of automatic ABCD rule)

4.0.3 Dermoscopic Structure Extraction

Dermoscopic structures are visual features that can distinguish melanoma or non-melanoma from benign lesions. Without considering these common structures it would be harder to find the differences in skin conditions including SK and melanoma, which can share many features. These techniques often use Gabor or Sobel filtering techniques to locate and enhance edges before using a texture extraction method to classify the individual structures. This is then

followed by a fusion method that combines the results from the individual Support Vector Machine.

Pigment Network

Pigment networks can be described as a brown net-like structure commonly surrounding a lesion. This network is useful when diagnosing melanoma because a typical networks are uniform in colour and width of the lines and gaps. Therefore accelerated cellular growth can be distinguished because of the erratic changes in shape and colour of this structure, proving valuable when recognising skin cancers.

One of the methods used to locate typical networks uses machine learning to locate areas of interest, followed by Gabor filtering and morphology to find uniform areas[79]. This method is robust because the lesion mask is not required for locating these structures, which means that in a scenario the segmentation algorithm fails, this retains its accuracy.

Pigment network detection using directional filters[80].

One of the more recently released methods can distinguish between typical and atypical pigment networks using directional Gabor filtering and machine learning[81] with an accuracy of 84.6% in the PH2 dataset. This method first locates the area of a lesion using a Gabor filter bank of 19, before extracting the texture and colour to classify the lesion into either a typical or atypical network.

The pigmented network is extracted before passing these results into a machine learning model to classify whether the lesion is melanoma and what if the network is typical and atypical. While this technique is substantially high in accuracy it lacks the ability to show regions that are atypical.

Dots or Globules

Inverse non-linear diffusion to enhance darker circular areas within an image [82].

Streaks

[83]

Other

The structures mentioned in this section are only the most commonly exploited, which means that the accuracy is likely to reach a peak that cannot be passed without considering other

smaller features including Milia-like cysts, angulated lines, white structures, granularity, etc. However the predicament is that providing further features might further confuse the diagnoses procedure for a GP. The features above while providing an accuracy of around 80% will still provide enough valuable data for an accurate diagnoses at lower levels.

4.1 Data Augmentation

Data augmentation is the method in which data can be modified or organised to increase the efficiency when training a model. This is important because relevant data samples (especially of rarer skin conditions) are often not available. Therefore ensuring that the training process is efficient can allow for higher accuracies to be reached with less training data.

4.1.1 Image Synthesis

Lesion Synthesis is the generation of realistic image samples and has found to be useful in medical imaging by allowing for further variation in data augmentation [84]. For example, when training a classification model to improve the accuracy and prevent overfitting the model is augmented by transforming, rotating and scaling. However, these augmented samples can deform the skin cancer data into unrealistic samples, increasing false positives. Generating synthetic images allows for variations in image data without unrealistic image transformations, improving the accuracy of classification algorithms [85].

Generative Adversarial Networks (GAN) are a style of neural network that uses game logic to generate realistic images. The training process consists of two networks training in conjunction called the discriminator and the generator. The generator attempts to create realistic images while the discriminator identifies synthetic ones from the real [84]. After a period of training (around 3000 iterations) the synthetic images become indistinguishable from real ones. Using these generated samples to train a classifier can improve the performance on rarer skin conditions [86], segmentation accuracy [86] and classification accuracy [85].

There are many different variations of GANs with different advantages, some of which are PoGO Alerts Network (PGAN), Laplacian Pyramid GAN (LAPGAN), Decision Diagram GAN (DDGAN) and Deep Convolutional GAN (DCGAN). Both DDGAN and LAPGAN proved valuable because they can produce diverse image samples with smaller datasets with an improved performance. However, LAPGAN sometimes generates high frequency artifacts and requires adjustments during the training process unlike DDGAN which is easier to train

[87].

GANs require extensive image data for training which can be scarce in medical imaging [87]. However to overcome this issue ISIC in 2017 created a challenge where only 2000 provided dermoscopic images can be used [9], giving academics a chance to engage and implement journals. The ISIC 2019 dataset has a total of 25,331 training images and 8,238 validation images, supporting the implementation and comparisons of GANs.

Since the new implementations of GAN synthesising images is far more reliable. However, the training process is still elusive and more difficult to train than discriminative models. Another issue is when generating images, the more samples produced creates a larger variety of samples with a lower quality. Therefore, to produce realistic samples less images need to be generated, which have minor variations in shape and size to the original dataset.

The only use for this within medical imaging is data augmentation which can improve the generalization when training a classifier. However, some papers outside of medical imaging use it to repair image samples through colour [88] and lighting correction [89]. Lighting correction is achieved by using an existing technique to extract noise which is corrected using a Label Denoising Adversarial Network (LDAN). This generates a ground truth of the natural skin which is used for training a deep CNN to extract lighting variations. By using a similar technique, it might be possible to repair macroscopic images to estimate and remove lighting conditions.

4.1.2 Transfer Learning

There has been recent development within an area called transfer learning which allows for networks to be partially trained on unrelated image samples before the training is finalised within the area of interest. Recent example of these are GoogleNet [90], AlexNet [91] and vgg19 [92], allowing for models to be trained with less data and for companies to keep their datasets private. However, appropriate transfer models need to be chosen carefully as they are likely to provide different results [?]. Therefore, having access to the training data is still more reliable.

AlexNet was successfully used to train a DCNN model with less training data [93] and gathered a higher quality model.

Transfer learning has found to be useful when extracting spatial features from a CNN model [94]. The main problem is the sheer amount of features returned from a trained model, making organising through them a difficult process. Transfer learning shortens the training process, optimising the network and results in returning less features. This is commonly

followed by another optimisation process to allow for features to be organised and interpreted. This is one of the techniques which allows for machine learning models to be interpreted otherwise k-means, filtering and threshold methods are relied on to gather these features.

4.2   Explainability

Since 2017 machine learning approaches have been made difficult to use in medical environments because explainability is lacking [23]. For example, machine learning is often referred to as a "black box" approach because the method and the features used to generate a classification model is often left undescribed. This is one of the biggest problems as the model is often incomprehensible to the developers and unmodifiable, reducing trust in machine learning [95].

Another concern is how explainable results can be interpreted by people [96]. Interpretability is a method in which the results are presented to the user in a way that can be understood and considered. The problem has remained unsolved because it is subjective concept and hard to formulate a single answer. Furthermore, it is domain-specific meaning there is no general-purpose answer to this problem [97]. Within medical imaging an ideal method might be to point to areas of the image describing the features associated with the ABCD rule. However, explanations generated from machine learning algorithms are still frequently incorrect [97].

In a medical environment, explainable algorithms need to be easily interpretable for a fast and effective diagnosis. Currently these algorithms can catalogue the training process which can be read and analysed to further understand the classification process. However, without the expertise and training it remains uninterpretable to a doctor. There has been a rise in techniques that interpret the produced data into a form that people can understand. However, this stage can manipulate the explanation to produce a realistic, but incorrect results [98].

One of the major pitfalls within machine learning is the uncertainty whether the data produced from explainable algorithms can be interpreted by people [99]. For example, doctors are taught of the ABCD rules which define a set of the most dominant features which over time evolve into more dominant versions of those features. Machine learning algorithms find the correlation in shapes, size, rotation and the differences in obtained feature locations. Meaning that the locations found could be different from the ABCD rules and unrelated to classifying the lesion, except with a high accuracy.

This holds the potential in allowing for doctors to better themselves by locating skin cancer using frequently unconsidered features. However, at what level are these captured features completely uninterpretable by another human and if they cannot be understood can the explanation be trusted [96]. One method for improving the reliability of the features was proposed by segmenting and labelling the features of interest to guide the training process, like using metadata and ground truths. Another example of this is a pedagogical model called TED [100], which is a framework designed to increase explainability by controlling the learning strategy of the network. Although this is largely built to improve reliability, it has the potential benefits to improve the accuracy and training times of models. However, this approach does not remove the "black box" but can improve the results.

Regardless of there being speculation on the kind of explainability can be achieved there has still been a fast progression in a range of different techniques including linear proxy models, decomposition, occlusion procedure and decision trees.

Linear proxy models (LPM) are best described through LIME [101] which is a method that probes behaviour of a model based on the perturbations of the input. A local linear model is then constructed which serves as a simplified proxy which interprets the decision making. One of the desirable benefits of this technique is that the results are quantifiable, and models can be compared more in depth. Moreover, trained models are analysed, while other techniques utilise the training process.

Another LPM method is known as Shapley Additive exPlanations (SHAP) and is a unified technique connecting optimal credit allocations with local explanations using classic Shapley values from game theory [102]. Interpretability has been considered with the development of TreeExplainer, allowing for models to be visualised in graphs and the generation of a heatmap relating to the areas used for classification in an image [103]. The variety of visualisation tools allows for interpretation of the model on higher and a lower level, whether its for doctors or computer scientists.

Visualization after recognition is progressively receiving attention [104] because it can be interpreted fast and easily, which is paramount to a medical environment. Deep Taylor decomposition was developed for this reason and can produce a visualisation of the key features using a layer called relevance propagation. This returns a heatmap visualising the most relevant features located in image, showing false positives which are otherwise hard to find. This has similarities to the decomposition techniques mentioned in the segmentation section, allowing for images to be processed by the model.

Grad-Cam is a technique that presents visual explanations through decomposition [22],

highlighting the most important regions in the image when predicting the region. This functions with CNN models with a range of different layers all without re-training or changes to architecture. Studies were conducted proving that people could successfully discern a strong model from a weak one. Proving there is a level of interpretability.

Another method uses a codebook which is when common image features are associated with a word. This is achieved by using Scale-Invariant Feature Transform (SIFT) to capture feature descriptors and k-means to optimally locate areas of interest. The classification process then returns associated features with a description, possibly providing adequate explanations to dermatologists [105]. However, accuracy is still lacking compared to CNN related methods.

Decision trees can be generated from CNN models at the semantic level, mining all the potential decision modes. This is considered an explainable model because once the model is converted all nodes and decision making is visible and can be analysed [106]. While this is explainable and even modifiable, the decision trees are often overly complicated and difficult to interpret. There are techniques that optimise the decision trees [103], however, it has the possibility of lowering accuracy.

While there is a massive range of techniques that can be used to improve on interpretability of machine learning, the acceptable techniques and the method in which they can be compared is still being debated. Many papers create extravagant complicated techniques claiming that it is explainable only to specify that they are unsure whether the technique is acceptable. This is proof that the area is new and further progression is needed before a standard is widely accepted.

The advantage of explainable models is to make use of past techniques that were not developed with interpretability in mind. These past models have incredibly high accuracies, but cannot be used in a medical environment because the answers are not re-traceable. They have been developed for this reason, however, the techniques are still lacking to accurately return features relating to the ABCD rules. Produced heatmaps are useful for debugging a model, but do not provide any validation to a doctor. It seems that the stage between explainability and interpretability in these techniques are still lacking too much to provide adequate validation. In a scenario where an algorithm was being used on a mobile device instead of in a clinical environment this would be a possibility.

4.3 Summary

There is simply too much speculation on the explainability of techniques today, making it difficult to warrant implementing them for a medical environment. Recent papers have taken a step back and left "black box" approaches to develop filtering, clustering or thresholding techniques for segmentation instead. Frameworks have been defined that separate the classification process into sections to allow for visualisation and the back tracking of results.This allows for features relating to the ABCD rules to be defined, supporting the classification process. Undoubtedly these are not the highest accuracy methods for lesion classification, however, all the results have to be retractable for a use in a medical environment.

Settled on using a dermoscope with polarised and non-polarised light.

Similar to TED, features could be split into groups so they are re-traceable and possibly a slight increase in accuracy.

Bibliography

[1] Cancer Research UK. Melanoma skin cancer statistics — Cancer Research UK, 2019.

[2] C Newlands, R Currie, A Memon, S Whitaker, and T Woolford. Non-melanoma skin cancer: United Kingdom National Multidisciplinary Guidelines. *The Journal of Laryngology & Otology*, 130(S2):S125–S132, 2016.

[3] Lucy L. Chen, Natalia Jaimes, Christopher A. Barker, Klaus J. Busam, and Ashfaq A. Marghoob. Desmoplastic melanoma: A review. *Journal of the American Academy of Dermatology*, 68(5):825–833, 2013.

[4] William E. Damsky, Lara E. Rosenbaum, and Marcus Bosenberg. Decoding melanoma metastasis. *Cancers*, 3(1):126–163, 2011.

[5] A. Sandru, S. Voinea, E. Panaitescu, and A. Blidaru. Survival rates of patients with metastatic malignant melanoma. *Journal of medicine and life*, 7(4):572–576, 2014.

[6] D. S. Rigel, J. Russak, and R. Friedman. The Evolution of Melanoma Diagnosis: 25 Years Beyond the ABCDs. *CA: A Cancer Journal for Clinicians*, 60(5):301–316, 2010.

[7] Cristina Carrera, Michael A. Marchetti, Stephen W. Dusza, Giuseppe Argenziano, Ralph P. Braun, Allan C. Halpern, Natalia Jaimes, Harald J. Kittler, Josep Malvehy, Scott W. Menzies, Giovanni Pellacani, Susana Puig, Harold S. Rabinovitz, Alon Scope, H. Peter Soyer, Wilhelm Stolz, Rainer Hofmann-Wellenhof, Iris Zalaudek, and Ashfaq A. Marghoob. Validity and Reliability of Dermoscopic Criteria Used to Differentiate Nevi From Melanoma. *JAMA Dermatology*, 152(7):798, 2016.

[8] Philipp Tschandl, Cliff Rosendahl, and Kittler Hararld. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Sci Data 5*, 2018.

[9] Noel C.F. Codella, David Gutman, M. Emre Celebi, Brian Helba, Michael A. Marchetti, Stephen W. Dusza, Aadi Kalloo, Konstantinos Liopyris, Nabin Mishra, Harald Kittler, and Allan Halpern. Skin lesion analysis toward melanoma detection: A challenge at the 2017 International symposium on biomedical imaging (ISBI), hosted by the international skin imaging collaboration (ISIC). *Proceedings - International Symposium on Biomedical Imaging*, 2018-April:168–172, 2018.

[10] Marc Combalia, Noel C. F. Codella, Veronica Rotemberg, Brian Helba, Veronica Vilaplana, Ofer Reiter, Cristina Carrera, Alicia Barreiro, Allan C. Halpern, Susana Puig, and Josep Malvehy. BCN20000: Dermoscopic Lesions in the Wild. 2019.

[11] Fiona M. Walter, A. Toby Prevost, Joana Vasconcelos, Per N. Hall, Nigel P. Burrows, Helen C. Morris, Ann Louise Kinmonth, and Jon D. Emery. Using the 7-point checklist as a diagnostic aid for pigmented skin lesions in general practice: A diagnostic validation study. *British Journal of General Practice*, 63(610), 2013.

[12] J. Daniel Jensen and Boni E. Elewski. The ABCDEF rule: Combining the 'ABCDE rule' and the "ugly duckling sign" in an effort to improve patient self-screening examinations. *Journal of Clinical and Aesthetic Dermatology*, 8(2):15, 2015.

[13] M. A. Pizzichetta, R. Talamini, I. Stanganelli, P. Puddu, R. Bono, G. Argenziano, A. Veronesi, G. Trevisan, H. Rabinovitz, and H. P. Soyer. Amelanotic/hypomelanotic melanoma: Clinical and dermoscopic features. *British Journal of Dermatology*, 150(6):1117–1124, 2004.

[14] Joanna Jaworek-Korjakowska. Automatic detection of melanomas: An application based on the ABCD criteria. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 7339 LNBI:67–76, 2012.

[15] Armand B. Cognetta, Thomas Vogt, Michael Landthaler, Otto Braun-Falco, and Gerd Plewig. The ABCD rule of dermatoscopy: High prospective value in the diagnosis of doubtful melanocytic skin lesions. *Journal of the American Academy of Dermatology*, 30(4):551–559, 1994.

[16] Catarina Barata, M. Emre Celebi, and Jorge S. Marques. A Survey of Feature Extraction in Dermoscopy Image Analysis of Skin Cancer. *IEEE Journal of Biomedical and Health Informatics*, 23(3):1096–1109, 2019.

[17] Catarina Barata, Jorge S. Marques, and Teresa Mendonça. Bag-of-features classification model for the diagnose of melanoma in dermoscopy images using color and texture descriptors. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 7950 LNCS:547–555, 2013.

[18] Eugenio Vocaturo, Diego Perna, and Ester Zumpano. Machine Learning Techniques for Automated Melanoma Detection. *Proceedings - 2019 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2019*, pages 2310–2317, 2019.

[19] Samy Bakheet. An SVM framework for malignant melanoma detection based on optimized HOG features. *Computation*, 5(1), 2017.

[20] Youssef Filali, My Abdelouahed Sabri, and Abdellah Aarab. An improved segmentation approach for skin lesion classification. *Statistics, Optimization and Information Computing*, 7(2):456–467, 2019.

[21] Christopher J. Kelly, Alan Karthikesalingam, Mustafa Suleyman, Greg Corrado, and Dominic King. Key challenges for delivering clinical impact with artificial intelligence. *BMC Medicine*, 17(1), 2019.

[22] Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-cam: Why did you say that? visual explanations from deep networks via gradient-based localization. *Revista do Hospital das Cl??nicas*, 17:331–336, 2016.

[23] Andreas Holzinger, Chris Biemann, Constantinos S. Pattichis, and Douglas B. Kell. What do we need to build explainable AI systems for the medical domain? *ArXiv*, 2017.

[24] Esteva Andre, Kuprel Brett, Novoa Roberto A., Ko Justin, Swetter Susan M., Blau Helen M., and Thrun Sebastian. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639):115–118, 2017.

[25] Yuji Roh, Geon Heo, and Steven Euijong Whang. A Survey on Data Collection for Machine Learning: A Big Data - AI Integration Perspective. *IEEE Transactions on Knowledge and Data Engineering*, pages 1–1, 2019.

[26] Giuseppe Argenziano and H. Peter Soyer. Dermoscopy of pigmented skin lesions - a valuable tool for early diagnosis of melanoma. *Lancet Oncology*, 2(7):443–449, 2001.

[27] G Alden Holmes, Janna M Vassantachart, Brittanya A Limone, Michael Zumwalt, Jane Hirokane, and Sharon E Jacob. Using Dermoscopy to Identify Melanoma and Improve Diagnostic Discrimination. *Federal practitioner : for the health care professionals of the VA, DoD, and PHS*, 35(Suppl 4):S39–S45, 2018.

[28] Cristina Carrera, Sonia Segura, Paula Aguilera, Massimiliano Scalvenzi, Caterina Longo, Alicia Barreiro, Paolo Broganelli, Stefano Cavicchini, Alex Llambrich, Pedro Zaballos, Luc Thomas, Josep Malvehy, Susana Puig, and Iris Zalaudek. Dermoscopic clues for diagnosing melanomas that resemble seborrheic keratosis. *JAMA Dermatology*, 153(6):544–551, 2017.

[29] Pedro M.M. Pereira, Rui Fonseca-Pinto, Rui Pedro Paiva, Pedro A.A. Assuncao, Luis M.N. Tavora, Lucas A. Thomaz, and Sergio M.M. Faria. Skin lesion classification enhancement using border-line features – The melanoma vs nevus problem. *Biomedical Signal Processing and Control*, 57, 2020.

[30] S. M. Stricklin, W. V. Stoecker, M. C. Oliviero, H. S. Rabinovitz, and S. K. Mahajan. Cloudy and starry milia-like cysts: How well do they distinguish seborrheic keratoses from malignant melanomas? *Journal of the European Academy of Dermatology and Venereology*, 25(10):1222–1224, 2011.

[31] Akane Minagawa. Dermoscopy–pathology relationship in seborrheic keratosis. *Journal of Dermatology*, 44(5):518–524, 2017.

[32] Lavinia Ferrante di Ruffano, Yemisi Takwoingi, Jacqueline Dinnes, Naomi Chuchu, Susan E. Bayliss, Clare Davenport, Rubeta N. Matin, Kathie Godfrey, Colette O'sullivan, Abha Gulati, Sue Ann Chan, Alana Durack, Susan O'connell, Matthew D. Gardiner, Jeffrey Bamber, Jonathan J. Deeks, and Hywel C. Williams. Computer-assisted diagnosis techniques (dermoscopy and spectroscopy-based) for diagnosing skin cancer in adults. *Cochrane Database of Systematic Reviews*, (12), 2018.

[33] Abder Rahman H. Ali, Jingpeng Li, and Guang Yang. Automating the ABCD Rule for Melanoma Detection: A Survey. *IEEE Access*, 8:83333–83346, 2020.

[34] Javier López-Labraca, Miguel Ángel Fernández-Torres, Iván González-Díaz, Fernando Díaz-De-María, and Ángel Pizarro. Enriched dermoscopic-structure-based cad system for melanoma diagnosis. *Multimedia Tools and Applications*, 77(10):12171–12202, 2018.

[35] Andrea Pennisi, Domenico D. Bloisi, Daniele Nardi, Anna Rita Giampetruzzi, Chiara Mondino, and Antonio Facchiano. Skin lesion image segmentation using Delaunay Triangulation for melanoma detection. *Computerized Medical Imaging and Graphics*, 52:89–103, 2016.

[36] 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. *Yearbook of Dermatology and Dermatologic Surgery*, 2009:378–379, 2009.

[37] Howard Zhou, Mei Chen, Le Zou, Richard Gass, Laura Ferris, Laura Drogowski, and James M. Rehg. Spatially constrained segmentation of dermoscopy images. *2008 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, Proceedings, ISBI*, pages 800–803, 2008.

[38] Sait Suer, Sinan Kockara, and Mutlu Mete. An improved border detection in dermoscopy images for density based clustering. *BMC Bioinformatics*, 12(SUPPL. 10), 2011.

[39] Steven Q. Wang, Stephen W. Dusza, Alon Scope, Ralph P. Braun, Alfred W. Kopf, and Ashfaq A. Marghoob. Differences in dermoscopic images from nonpolarized dermoscope and polarized dermoscope influence the diagnostic accuracy and confidence level: A pilot study. *Dermatologic Surgery*, 34(10):1389–1395, 2008.

[40] Maryam Ramezani, Alireza Karimian, and Payman Moallem. Automatic Detection of Malignant Melanoma using Macroscopic Images. *Journal of Medical Signals and Sensors*, 4(4):281–290, 2014.

[41] Verosha Pillay and Serestina Viriri. Skin cancer detection from macroscopic images. *2019 Conference on Information Communications Technology and Society, ICTAS 2019*, 2019.

[42] Ioannis Giotis, Nynke Molders, Sander Land, Michael Biehl, Marcel F. Jonkman, and Nicolai Petkov. MED-NODE: A computer-assisted melanoma diagnosis system using non-dermoscopic images. *Expert Systems with Applications*, 42(19):6578–6585, 2015.

[43] Robert Amelard, Jeffrey Glaister, Alexander Wong, and David A. Clausi. Melanoma Decision Support Using Lighting-Corrected Intuitive Feature Models. *Computer Vision Techniques for the Diagnosis of Skin Cancer*, pages 193–219, 2014.

[44] David J. Carpenter, Mirabelle B. Sajisevi, Nikita Chapurin, Clifford Scott Brown, Tracy Cheng, Gregory M. Palmer, Daniel S. Stevenson, Caroline L. Rao, Russell P. Hall, and Charles R. Woodard. Noninvasive optical spectroscopy for identification of non-melanoma skin cancer: Pilot study. *Lasers in Surgery and Medicine*, 50(3):246–252, 2018.

[45] Nils Gessert, Marcel Bengs, and Alexander Schlaefer. Melanoma detection with electrical impedance spectroscopy and dermoscopy using joint deep learning models. *ArXiv*, page 37, 2020.

[46] T. Y. Satheesha, D. Satyanarayana, M. N.Giri Prasad, and Kashyap D. Dhruve. Melanoma Is Skin Deep: A 3D Reconstruction Technique for Computerized Dermoscopic Skin Lesion Classification. *IEEE Journal of Translational Engineering in Health and Medicine*, 5, 2017.

[47] Toshitsugu Sato and Masaru Tanaka. Improved detection of dermoscopic structures by high dynamic range image conversion. *The Japanese Journal of Dermotology*, 123(2):121–131, 2013.

[48] Toshitsugu Sato and Masaru Tanaka. A case of a superficial spreading melanoma in situ diagnosed via digital dermoscopic monitoring with high dynamic range conversion. *Dermatology Practical & Conceptual*, 4(5), 2014.

[49] Ralph P. Braun and Ashfaq Marghoob. High-dynamic-range dermoscopy imaging and diagnosis of hypopigmented skin cancers. *JAMA Dermatology*, 151(4):456–457, 2015.

[50] Yading Yuan, Ming Chao, and Yeh Chi Lo. Automatic Skin Lesion Segmentation Using Deep Fully Convolutional Networks with Jaccard Distance. *IEEE Transactions on Medical Imaging*, 36(9):1876–1886, 2017.

[51] Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(12):2481–2495, 2017.

[52] Euijoon Ahn, Jinman Kim, Lei Bi, Ashnil Kumar, Changyang Li, Michael Fulham, and David Dagan Feng. Saliency-Based Lesion Segmentation Via Background Detection in Dermoscopic Images. *IEEE Journal of Biomedical and Health Informatics*, 21(6):1685–1693, 2017.

[53] Oludayo O. Olugbara, Tunmike B. Taiwo, and Delene Heukelman. Segmentation of Melanoma Skin Lesion Using Perceptual Color Difference Saliency with Morphological Analysis. *Mathematical Problems in Engineering*, 2018:1–19, 2018.

[54] Kai Hu, Si Liu, Yuan Zhang, Chunhong Cao, Fen Xiao, Wei Huang, and Xieping Gao. Automatic segmentation of dermoscopy images using saliency combined with adaptive thresholding based on wavelet transform. *Multimedia Tools and Applications*, 2019.

[55] Haidi Fan, Fengying Xie, Yang Li, Zhiguo Jiang, and Jie Liu. Automatic segmentation of dermoscopy images using saliency combined with Otsu threshold. *Computers in Biology and Medicine*, 85:75–85, 2017.

[56] Maryam Bibi, Anmol Hamid, and Samabia Tehseen. Automated Skin Lesion Detection towards Melanoma. *ICST Transactions on Scalable Information Systems*, 0(0):159800, 2018.

[57] E. Meskini, M. S. Helfroush, K. Kazemi, and M. Sepaskhah. A new algorithm for skin lesion border detection in dermoscopy images. *Journal of Biomedical Physics and Engineering*, 8(1):109–118, 2018.

[58] Ihab S. Zaqout. Diagnosis of Skin Lesions Based on Dermoscopic Images Using Image Processing Techniques. *International Journal of Signal Processing, Image Processing and Pattern Recognition*, 9(9):189–204, 2016.

[59] Reda Kasmi and Karim Mokrani. Classification of malignant melanoma and benign skin lesions: Implementation of automatic ABCD rule. *IET Image Processing*, 10(6):448–455, 2016.

[60] 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. *PLoS ONE*, 15(6), 2020.

[61] Nikolaos E. Myridis. Ultra-realistic Imaging: Advanced Techniques in Analogue and Digital Colour Holography, by Hans Bjelkhagen and David Brotherton-Ratcliffe. *Contemporary Physics*, 55(3):247–248, 2014.

[62] Ihab Zaqout. Diagnosis of Skin Lesions Based on Dermoscopic Images Using Image Processing Techniques. *Pattern Recognition - Selected Methods and Applications*, 2019.

[63] Abder-Rahman Ali, Jingpeng Li, Guang Yang, and Sally Jane O'Shea. A machine learning approach to automatic detection of irregularity in skin lesion border using dermoscopic images. *PeerJ Computer Science*, 6:e268, 2020.

[64] Marchetti M.A., Liopyris K., Dusza S.W., Codella N.C.F., Gutman D.A., Helba B., Kalloo A., Halpern A.C., Soyer H.P., Curiel-Lewandrowski C., Caffery L., and Malvehy J. Computer algorithms show potential for improving dermatologists' accuracy to diagnose cutaneous melanoma: Results of the International Skin Imaging Collaboration 2017. *Journal of the American Academy of Dermatology*, 2020.

[65] K. M. Clawson, P. J. Morrow, B. W. Scotney, D. J. McKenna, and O. M. Dolan. Computerised skin lesion surface analysis for pigment asymmetry quantification. *International Machine Vision and Image Processing Conference, IMVIP 2007*, pages 75–82, 2007.

[66] M. Emre Celebi and Azaria Zornberg. Automated quantification of clinically significant colors in dermoscopy images and its application to skin lesion classification. *IEEE Systems Journal*, 8(3):980–984, 2014.

[67] A. Rajesh. Classification of malignant melanoma and Benign Skin Lesion by using back propagation neural network and ABCD rule. *Proceedings - 2017 IEEE International Conference on Electrical, Instrumentation and Communication Engineering, ICEICE 2017*, 2017-Decem:1–8, 2017.

[68] Alexander Selvikvåg Lundervold and Arvid Lundervold. An overview of deep learning in medical imaging focusing on MRI. *Zeitschrift fur Medizinische Physik*, 29(2):102–127, 2019.

[69] Agnieszka Mikołajczyk and Michał Grochowski. Data augmentation for improving deep learning in image classification problem. *2018 International Interdisciplinary PhD Workshop, IIPhDW 2018*, pages 117–122, 2018.

[70] M. Prathiba, Deepa Jose, R. Saranya, and Nandhinidevi. Automated Melanoma Recognition in Dermoscopy Images via Very Deep Residual Networks. *IOP Conference Series: Materials Science and Engineering*, 561(1), 2019.

[71] Achim Hekler, Jochen S. Utikal, Alexander H. Enk, Axel Hauschild, Michael Weichenthal, Roman C. Maron, Carola Berking, Sebastian Haferkamp, Joachim Klode, Dirk

Schadendorf, Bastian Schilling, Tim Holland-Letz, Benjamin Izar, Christof von Kalle, Stefan Fröhling, Titus J. Brinker, Laurenz Schmitt, Wiebke K. Peitsch, Friederike Hoffmann, Jürgen C. Becker, Christina Drusio, Philipp Jansen, Georg Lodde, Stefanie Sammet, Wiebke Sondermann, Selma Ugurel, Jeannine Zader, Alexander Enk, Martin Salzmann, Sarah Schäfer, Knut Schäkel, Julia Winkler, Priscilla Wölbing, Hiba Asper, Ann Sophie Bohne, Victoria Brown, Bianca Burba, Sophia Deffaa, Cecilia Dietrich, Matthias Dietrich, Katharina Antonia Drerup, Friederike Egberts, Anna Sophie Erkens, Salim Greven, Viola Harde, Marion Jost, Merit Kaeding, Katharina Kosova, Stephan Lischner, Maria Maagk, Anna Laetitia Messinger, Malte Metzner, Rogina Motamedi, Ann Christine Rosenthal, Ulrich Seidl, Jana Stemmermann, Kaspar Torz, Juliana Giraldo Velez, Jennifer Haiduk, Mareike Alter, Claudia Bär, Paul Bergenthal, Anne Gerlach, Christian Holtorf, Ante Karoglan, Sophie Kindermann, Luise Kraas, Moritz Felcht, Maria R. Gaiser, Claus Detlev Klemke, Hjalmar Kurzen, Thomas Leibing, Verena Müller, Raphael R. Reinhard, Jochen Utikal, Franziska Winter, Laurie Eicher, Daniela Hartmann, Markus Heppt, Katharina Kilian, Sebastian Krammer, Diana Lill, Anne Charlotte Niesert, Eva Oppel, Elke Sattler, Sonja Senner, Jens Wallmichrath, Hans Wolff, Anja Gesierich, Tina Giner, Valerie Glutsch, Andreas Kerstan, Dagmar Presser, Philipp Schrüfer, Patrick Schummer, Ina Stolze, Judith Weber, Konstantin Drexler, Marion Mickler, Camila Toledo Stauner, and Alexander Thiem. Superior skin cancer classification by the combination of human and artificial intelligence. *European Journal of Cancer*, 120:114–121, 2019.

[72] Z. Liu, J. Sun, M. Smith, L. Smith, and R. Warr. Incorporating clinical metadata with digital image features for automated identification of cutaneous melanoma. *British Journal of Dermatology*, 169(5):1034–1040, 2013.

[73] Jason R. Hagerty, R. Joe Stanley, Haidar A. Almubarak, Norsang Lama, Reda Kasmi, Peng Guo, Rhett J. Drugge, Harold S. Rabinovitz, Margaret Oliviero, and William V. Stoecker. Deep Learning and Handcrafted Method Fusion: Higher Diagnostic Accuracy for Melanoma Dermoscopy Images. *IEEE Journal of Biomedical and Health Informatics*, 23(4):1385–1391, 2019.

[74] José Sergio Ruiz-Castilla, Juan José Rangel-Cortes, Farid García-Lamont, and Adrián Trueba-Espinosa. CNN and Metadata for Classification of Benign and Malignant Melanomas. *CNN and Metadata for Classification of Benign and Malignant Melanomas*, 11644 LNCS:569–579, 2019.

[75] Sertan Serte and Hasan Demirel. Gabor wavelet-based deep learning for skin lesion classification. *Computers in Biology and Medicine*, 113, 2019.

[76] M. Monisha, A. Suresh, and M. R. Rashmi. Artificial Intelligence Based Skin Classification Using GMM. *Journal of Medical Systems*, 43(1), 2018.

[77] Sharmin Majumder and Muhammad Ahsan Ullah. Feature extraction from dermoscopy images for melanoma diagnosis. *SN Applied Sciences*, 1(7), 2019.

[78] M. Attique Khan, Tallha Akram, Muhammad Sharif, Aamir Shahzad, Khursheed Aurangzeb, Musaed Alhussein, Syed Irtaza Haider, and Abdualziz Altamrah. An implementation of normal distribution based segmentation and entropy controlled features selection for skin lesion detection and classification. *BMC Cancer*, 18(1), 2018.

[79] Jose Luis García Arroyo and Begoña García Zapirain. Detection of pigment network in dermoscopy images using supervised machine learning and structural analysis. *Computers in Biology and Medicine*, 44(C):144–157, 2014.

[80] Catarina Barata, Jorge S. Marques, and Jorge Rozeira. A system for the detection of pigment network in dermoscopy images using directional filters. *IEEE Transactions on Biomedical Engineering*, 59(10):2744–2754, 2012.

[81] Sameena Pathan, K. Gopalakrishna Prabhu, and P. C. Siddalingaswamy. A methodological approach to classify typical and atypical pigment network patterns for melanoma diagnosis. *Biomedical Signal Processing and Control*, 44:25–37, 2018.

[82] Maglogiannis I. and Delibasis K.K. Enhancing classification accuracy utilizing globules and dots features in digital dermoscopy. In *Computer Methods and Programs in Biomedicine*, volume 118, pages 124–133, 2015.

[83] K. Kropidlowski, M. Kociolek, M. Strzelecki, and D. Czubinski. Nevus atypical pigment network distinction and irregular streaks detection in skin lesions images. *Signal Processing - Algorithms, Architectures, Arrangements, and Applications Conference Proceedings, SPA*, 2015-Decem:66–70, 2015.

[84] Xin Yi, Ekta Walia, and Paul Babyn. Generative adversarial network in medical imaging: A review. *Medical Image Analysis*, 58, 2019.

[85] Alceu Bissoto, Fábio Perez, Eduardo Valle, and Sandra Avila. Skin lesion synthesis with generative adversarial networks. *ArXiv*, 11041 LNCS:294–302, 2019.

[86] Amirata Ghorbani, Vivek Natarajan, David Coz, and Yuan Liu. DermGAN: Synthetic Generation of Clinical Skin Images with Pathology. *ArXiv*, 2019.

[87] Christoph Baur, Shadi Albarqouni, and Nassir Navab. MelanoGANs: High Resolution Skin Lesion Synthesis with GANs. *ArXiv*, 2018.

[88] Jie Li, Katherine A. Skinner, Ryan M. Eustice, and Matthew Johnson-Roberson. WaterGAN: Unsupervised generative network to enable real-time color correction of monocular underwater images. *IEEE Robotics and Automation Letters*, 3(1):387–394, 2018.

[89] Hao Zhou, Jin Sun, Yaser Yacoob, and David W. Jacobs. Label Denoising Adversarial Network (LDAN) for Inverse Lighting of Faces. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 6238–6247, 2018.

[90] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 07-12-June:1–9, 2015.

[91] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6):84–90, 2017.

[92] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. In *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, 2015.

[93] Khalid M. Hosny, Mohamed A. Kassem, and Mohamed M. Foaud. Classification of skin lesions using transfer learning and augmentation with Alex-net. *PLoS ONE*, 14(5), 2019.

[94] Tallha Akram, Hafiz M.Junaid Lodhi, Syed Rameez Naqvi, Sidra Naeem, Majed Al-haisoni, Muhammad Ali, Sajjad Ali Haider, and Nadia N. Qadri. A multilevel features selection framework for skin lesion classification. *Human-centric Computing and Information Sciences*, 10(1), 2020.

[95] Domenico Monaco. What is the Explainable-AI and Why is Important. *ArXiv*, pages 1–5, 2019.

[96] L H Gilpin, D Bau, B Z Yuan, A Bajwa, and ... Explaining explanations: An overview of interpretability of machine learning. *2018 IEEE 5th . . .*, 2018.

[97] Diogo V. Carvalho, Eduardo M. Pereira, and Jaime S. Cardoso. Machine learning interpretability: A survey on methods and metrics. *Electronics (Switzerland)*, 8(8), 2019.

[98] Amirata Ghorbani, Abubakar Abid, and James Zou. Interpretation of Neural Networks Is Fragile. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33:3681–3688, 2019.

[99] Anh Nguyen, Jason Yosinski, and Jeff Clune. Deep Neural Networks are Easily Fooled. *Computer Vision and Pattern Recognition, 2015 IEEE Conference on*, pages 427–436, 2015.

[100] Michael Hind, Dennis Wei, Murray Campbell, Noel C.F. Codella, Amit Dhurandhar, Aleksandra Mojsilović, Karthikeyan Natesan Ramamurthy, and Kush R. Varshney. TED: Teaching AI to explain its decisions. *AIES 2019 - Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, pages 123–129, 2019.

[101] Marco Ribeiro, Sameer Singh, and Carlos Guestrin. "Why Should I Trust You?": Explaining the Predictions of Any Classifier. *ArXiv*, pages 97–101, 2016.

[102] Scott M. Lundberg and Su In Lee. A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 2017-Decem:4766–4775, 2017.

[103] Scott M. Lundberg, Gabriel Erion, Hugh Chen, Alex DeGrave, Jordan M. Prutkin, Bala Nair, Ronit Katz, Jonathan Himmelfarb, Nisha Bansal, and Su-In Lee. From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence*, 2(1):56–67, 2020.

[104] Yiqi Yan, Jeremy Kawahara, and Ghassan Hamarneh. Melanoma Recognition via Visual Attention. *International Conference on Information Processing in Medical Imaging*, 11492 LNCS:793–804, 2019.

[105] Kai Hu, Xiaorui Niu, Si Liu, Yuan Zhang, Chunhong Cao, Fen Xiao, Wanchun Yang, and Xieping Gao. Classification of melanoma based on feature similarity measurement

for codebook learning in the bag-of-features model. *Biomedical Signal Processing and Control*, 51:200–209, 2019.

[106] Quanshi Zhang, Yu Yang, Haotian Ma, and Ying Nian Wu. Interpreting cnns via decision trees. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2019-June:6254–6263, 2019.