Automatic Detection of Melanoma in Dermoscopic Images of Skin Lesions

by

Amith Jooravan
Student Number: 20150183

A dissertation submitted in fulfilment of

the requirements for the degree of

Master of Engineering (MEng)

at the

Durban University of Technology
Faculty of Engineering and the Built Environment
Department of Electronic and Computer Engineering

Supervisor: Dr S. Reddy

November 2022

DECLARATION

1. I know and understand that plagiarism is using another person's work and pretending it is one's own, which is wrong.

2. This essay/report/project is my own work.

3. I have appropriately referenced the work of other people I have used.

4. I have not allowed and **will not allow** anyone to copy my work with the intention of passing it off as his/her own work.

_____ _____

Amith Jooravan Date

Student Number: 20150183

This research was duly supervised by Dr S. Reddy at the Durban University of Technology. This dissertation is the student's own work, every cited work or text have been properly referenced. It has not been partially or fully submitted at any other University.

Approved for Final Submission by:

_____ _____

Supervisor: Dr S. Reddy Date

ABSTRACT

Melanoma, which is an aggressive form of skin cancer, has the highest mortality rate of all skin cancers, especially if there is a late diagnosis. The diagnosis of melanoma is usually conducted in two stages; initially an invasive biopsy of a skin lesion under concern is conducted and subsequently the respective removed tissue undergoes laboratory analysis. A crucial component of the first stage is the highly subjective process of determining which skin lesions require a biopsy. The biopsy process may be painful, costly and time consuming, however in healthcare, a case of false positive may result in a patient being unnecessarily alarmed, while a case of false negative, a type-2 error that does not detect a positive case, may have dire consequences. This is owing, in the main, to the survival rate of melanoma being correlated to the stage of the cancer when first diagnosed. This research proposes a non-invasive method to assist with the reduction of false negative classifications associated with skin lesions that are considered as benign candidates instead of melanoma. In this research, 218 dermoscopic images from the Interactive Atlas of Dermoscopy dataset were used. The selected lesions were in early stages of melanoma, being in situ or less than 0,76mm thick. Of these, 178 were used for training and 40 for testing. The training and test dataset were balanced, comprising a 50/50 split, for melanoma and non-melanoma cases. Three classification algorithms were considered in this research; these include *k*-nearest neighbours (KNN), Naïve Bayes and linear support vector machine (LSVM). To identify potential candidates of skin lesions for biopsy, the algorithms consider the asymmetry, border, colour, and diameter of the skin lesions; this is referred to as the ABCD rule. This research proposes the use of a LSVM machine learning algorithm to classify a skin lesion as being either melanoma or non-melanoma with a view of minimising false negative rate of the investigated classification algorithms. Classification accuracy of 87.5% and a false negative rate of 5% is achieved.

PUBLICATIONS

A component of this research entitled “Comparative Study of Binary Classifiers for Reducing False Negative Detections of Melanoma in Skin Lesions” was presented at the IEEE 8th International Conference on Engineering and Emerging Technologies (ICEET 2022), 27-28 October 2022, Kuala Lumpur, Malaysia.

ACKNOWLEDGEMENTS

I would like to thank my supervisor, Dr Seren Reddy, for providing guidance, direction, and support with my research dissertation. I am truly grateful for the support I have received from my family and colleagues.

This study is dedicated to the light in my life, my inspirational daughter, Vidhya, who has provided me with the impetus to succeed.

TABLE OF CONTENTS

Declaration ..... **ii**

Abstract ..... **iii**

PUBLICATIONS ..... **iv**

Acknowledgements ..... **v**

Table of contents ..... **vi**

List of figures ..... **ix**

List of tables ..... **xii**

I. Introduction **1**

A. The purpose and relevance of the research 1

B. Background 2

1. Composition and Structure of Human Skin 2

2. Skin Lesions 3

3. Skin Cancer 3

4. Detection/Diagnosis of Melanoma 8

5. Biopsy procedure of Melanoma 12

C. Problem definition and challenges 12

D. Structure of dissertation 13

E. Scope Limitation 14

F. Contributions of the Dissertation 14

II. Literature Review **15**

A. Dermoscopic Image processing 15

1. Stage 1: Image Acquisition 16

2. Stage 2: Pre-processing and Filtering 17

3. Stage 3: Image Segmentation....20
4. Stage 4: Feature extraction....21
5. Stage 5: Classification....25
B. Related Work....27
III. Proposed methodology....31
A. Image Acquisition....31
B. Image Pre-processing....33
1. RGB to grayscale....33
2. Brightness adjustment....34
3. Contrast adjustment....34
C. Image Segmentation....36
D. Feature Extraction and Analysis....40
1. Asymmetry....40
2. Border Irregularity....42
3. Colour....43
4. Diameter....46
E. Classification of Analysed Images....47
IV. Results and DISCUSSION....54
A. Results of classification....54
B. Analysis of SVM results....57
C. Comparison with similar studies....60
V. Conclusion....62
A. Summary....62
B. Recommendations....63
C. Future work....63

VI. Reference list 64

VII. Appendix.....74

A. Training dataset.....74

B. Test dataset 82

LIST OF FIGURES

Fig. 1: Cross-section of skin [8].....2

Fig. 2: Cross section of skin showing the location of the 3 main types of skin cancer [19].....4

Fig. 3: Images showing the varying appearance of Basel cell carcinoma lesions [20].....5

Fig. 4(a)–(d): Images showing the varying appearance of Squamous cell carcinoma lesions [21].....6

Fig. 5: The 5 stages of melanoma [22] 7

Fig. 6(a)–(d) : Images of melanoma at various stages of progression[23] 7

Fig. 7: Diascopic test shows blanching of lesion [25] 8

Fig. 8: (a) Handheld dermoscope and (b) digital dermoscope [30] 9

Fig. 9: a) Normal light reflecting of the skin and (b) the resultant digital image captured [31].....9

Fig. 10: a) Polarised light refracting below the epidermis and (b) the resultant image captured on a dermoscope [31].....10

Fig. 11: Biopsy of lesion a) excisional biopsy, b)punch biopsy[26].....12

Fig. 12: The five stages towards Computer Aided Diagnosis.....16

Fig. 13: Examples of noise and unwanted objects: (a) hair, (b) blood vessels and (c) trapped air bubbles [48] 17

Fig. 14: Qualitative Image Enhancement, a) original image, b) enhanced image [50].....18

Fig. 15: Quantitative Image Enhancement, a) original image b) enhanced image for c) Edge detection[54] 18

Fig. 16: Area without overlap: a) folded along y axis and b) folded along x axis[73] 22

Fig. 17: Fractional dimension using the box counting method [79] 23

Fig. 18: Examples of colour scores for melanoma. a) Dark, light brown, red and white present, b) blue, dark brown, light brown and white present[82] 24

Fig. 19: The 5 stages towards melanoma detection 31

Fig. 20: Sample of images rejected due to being a) out of focus or b),c) incomplete [39].....32

Fig. 21: Original image from database, a) black frame around image, b) black frame removed from image. 32

Fig. 22: Pre-processing of image, a) Original RGB image, b) Grayscale conversion of dermoscopic image 33
Fig. 23: Segmentation of lesion with various brightness levels a) no change, b) 20% increase c) 40% increase 34
Fig. 24: Individual RGB colour channels of dermoscopic images, a) Original RGB image, b) Red channel only, c) Green channel only, d) Blue channel only. 34
Fig. 25: Histogram stretching: a) Original blue channel image, b) histogram of original image, c) Stretched image, d) histogram of stretched image, e) histogram equalised image with tile size = 8, and f) histogram of equalised image. 35
Fig. 26: a) Histogram equalised image b) Image binarization using Otsu's method and c) Image binarization using locally adaptive thresholding. 36
Fig. 27: Hair and noise removal using morphological processes, b) dilation (SE = 4), ...37
Fig. 28: Image segmentation, a) Boundary mask around the lesion and b) final segmented image 38
Fig. 29a-c: Examples showing the segmentation of the largest region while excluding the other smaller regions 38
Fig. 30: Various edge detection methods for edge detection of a skin lesion, sample1 39
Fig. 31: Various edge detection methods for edge detection of a skin lesion, sample 2 39
Fig. 32: Example of polygon for calculating the centroid [115] 40
Fig. 33: Plotting of centroid for various segmented lesions 41
Fig. 34: Circularity Index calculation of various shapes [117] 42
Fig. 36: Subtractive colour models of dermoscopic lesions: a) Original RGB image, b) Cyan, c) Magenta, and d) Yellow 44
Fig. 37: Six colours of interest count for melanoma lesions: (a) Lesion sample 1 with colour count 2, (b) Lesion sample 2 with colour count 2, (c) Lesion sample 3 with colour count 3, and (d) Lesion sample 4 with colour count 3. 45
Fig. 38: KNN with two classes A and B with k=1 and k=3 48
Fig. 39: Various hyperplanes that may separate the data into the 2 classes 50
Fig. 40: Linear SVM classification of linearly separable data with maximised margins 51

Fig. 41: Parallel coordinate plot for linear SVM 53

Fig. 42: Confusion Matrix for *k*-Nearest Neighbours for 40 test samples 55

Fig. 43: Confusion Matrix for Naïve Bayes for 40 test samples.....55

Fig. 44: Confusion Matrix for Linear SVM for 40 test samples.....56

Fig. 45: Incorrect classification of a lesion (False Negative) 58

Fig. 46: Incorrectly classified lesions (false positive) 58

LIST OF TABLES

Table I:.....11

Table II: 25

TABLE III: 30

Table III: 43

Table IV:.....44

Table V: 45

Table VI:.....46

Table VII:.....56

Table VIII: 59

Table IX:.....60

I. INTRODUCTION

A. The purpose and relevance of the research

Owing to the geographic location of South Africa the country is exposed to relatively high amounts of sunlight and subsequent ultraviolet (UV) radiation. Irrespective of skin colour and ethnicity, excessive exposure to UV radiation has shown to have adverse effects on humans including being a significant contributing factor to skin cancer [1].

According to the South African Melanoma Advisory Board there is a two-step process for the diagnosis of skin cancer, such as melanoma. The first step involves the visual inspection of a skin lesion by a dermatologist and if the prognosis cannot rule out skin cancer, then the second step involves a biopsy and diagnosis by a pathologist [2]. To perform a biopsy, a portion or the entire region of a suspicious mole is initially removed by a physician and the specimen is subsequently sent to a pathologist for analysis. The result is relayed back to the physician for onward communication to the patient. This process is invasive and, at present, time consuming and costly.

A direct examination of a suspected skin lesion is approximately 60% accurate in detecting melanoma and the naked eye analysis of a dermoscopic image of a skin lesion is shown to be slightly higher at 68% [3-5]. With the dramatic improvements in digital image capture technology together with the exponential increase in computer processing power, it has become viable for digital images of skin lesions to be analysed and classified using computer vision algorithms.

This research provides a comparative analysis of three classification algorithms that may be used to estimate the probability of melanoma being detected in dermoscopic digital images of skin lesions. The three algorithms considered are linear support vector machine, k-nearest neighbours, and Naïve Bayes; the optimal classification is based on accuracy, spec1ificity and sensitivity. The proposed algorithm will have the hyper parameters tuned to produce a balance between accuracy and minimising false negative classifications. The objective of this study focuses on the use of machine learning to aid physicians with the non-invasive classification of skin lesions.

B. Background

1. Composition and Structure of Human Skin

Skin, which is the largest organ of the human body, forms a protective barrier to the environment that covers the surface of the entire body. It forms the first line of defence against pathogens [6], regulates the temperature and the hydration of the body and provides sensation. Skin is made up of three basic layers; this is illustrated in Fig. 1. The outermost layer is the epidermis, followed by the dermis and finally the hypodermis or subcutaneous layer [7].

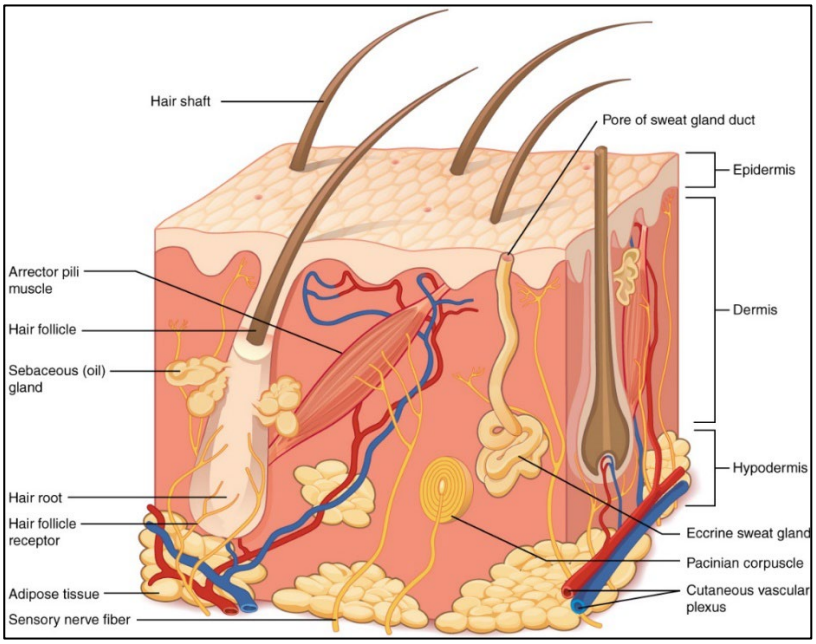


Fig. 1: Cross-section of skin [8]

a) Epidermis

The uppermost layer of skin is the epidermis. It varies in thickness throughout the body; for example, it is thinner on the eyelids and thicker on both the palms of the hands and soles of the feet. The epidermis, which produces new skin cells to replace dead skin cells, has five specialised cells [9]; these include (i) Melanocyte cells, which provide skin colour by producing the pigment melanin, (ii) Keratinocyte cells that produce the protein keratin to strengthen skin, (iii) Basal cells, which create new skin cells as older cells are pushed to the surface of the skin, (iv) Squamous cells that provide a filtration and diffusion of molecules, and (v) Merkel cells, which function as touch or sensory receptors.

b) Dermis

The dermis is the layer below the epidermis and above the hypodermis. The dermis is made up of connective tissue, muscle, neurons, blood and lymph vessels, hair follicles and sweat glands. The epidermis is avascular and, consequently, the cells of this layer must get their oxygen and nutrients from capillaries in the dermis [7]. The dermis protects against mechanical injury, provides nutrients to the epidermis, and helps to heal wounds [10].

c) Hypodermis/Subcutaneous layer

The hypodermis is the layer below the dermis. It is primarily made up of fat and contains nerves and blood vessels. It connects the skin to the fibrous tissue around muscles[8]. The fat provides thermoregulation and assists with shock absorption [10].

2. Skin Lesions

A skin lesion is an area of the skin that appears different to the surrounding skin, such as a lump, bump, ulcer, sore or pigmentation on the skin [11, 12]. Lesions can appear on a specific area of the body or could be across the entire body and could be caused by a simple bruise, cut, scrape or even an allergic reaction. Some lesions are caused by cancerous cells; these are commonly found in areas exposed to the sun [13].

3. Skin Cancer

Skin cancer, which affects humans of all skin tones, is the most common cancer globally; it accounted for over 1 million deaths worldwide in 2018 [14, 15]. Skin cancer occurs when regular skin cells abnormally mutate and multiply and consequently form a tumour. The spreading of a tumour to other tissues may be termed as cancerous or malignant. When tumours spread to other organs in the body through the blood or lymph system it is referred to as metastasis [16, 17].

The most common cause of skin cancer is overexposure to ultraviolet (UV) light [18, 19]. Areas of the body exposed to sun are the most vulnerable. There are 3 main types of skin cancer, namely basal cell

carcinoma, squamous cell carcinoma and melanoma. Fig. 2 below shows the location of these three main types of skin cancer.

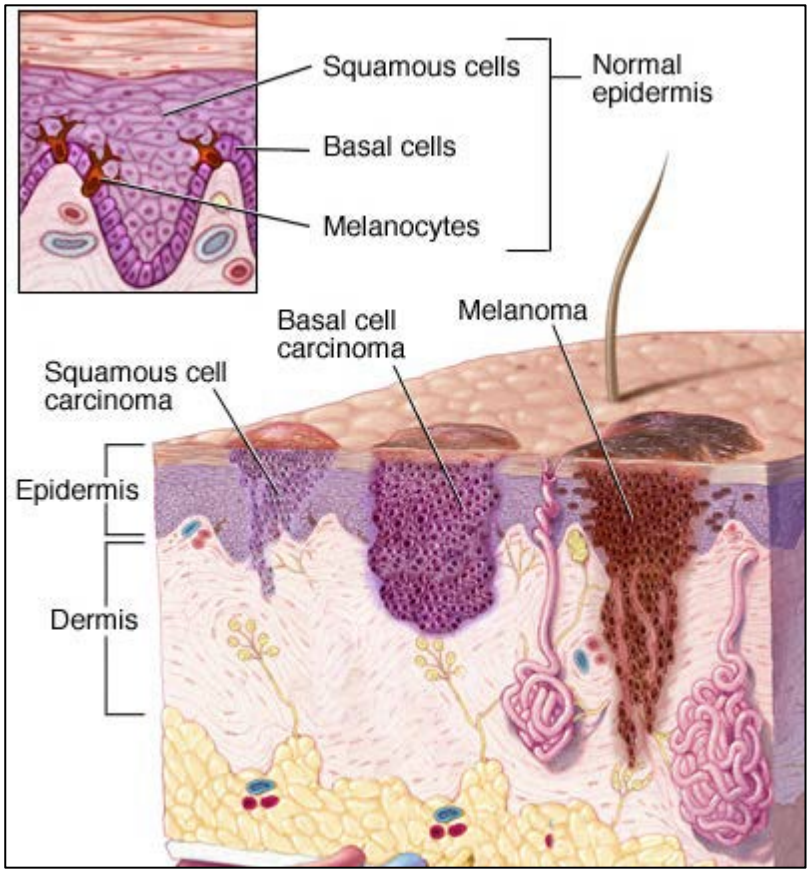


Fig. 2: Cross section of skin showing the location of the 3 main types of skin cancer [19]

a) Basel cell carcinoma (BCC)

Basel cell carcinoma (BCC), which is the most common type of skin cancer, forms from the mutation of the Basel cells in the lower part of the epidermis. It is linked to DNA damage through ultraviolet (UV) light and consequently is mainly found on areas of the body that are exposed to the sun. BCC, which has a slow growth rate, is very rarely metastatic, and does not lead to many deaths. However, if left untreated, the tumours may grow and lead to disfigurement. BCC can vary in its appearance and is characterised by reddish pink growths, open sores, or shiny translucent bumps. Figure 3 shows the different appearances of BCC lesions. Figure 3a shows an open sore that does not heal. Figure 3b, shows the BCC as a shiny nodule located on the face. In figure 3c, the BCC appears as a reddish patch near the eyebrow and in figure 3d the BCC lesion is a pink growth with a raised and rolled edge.

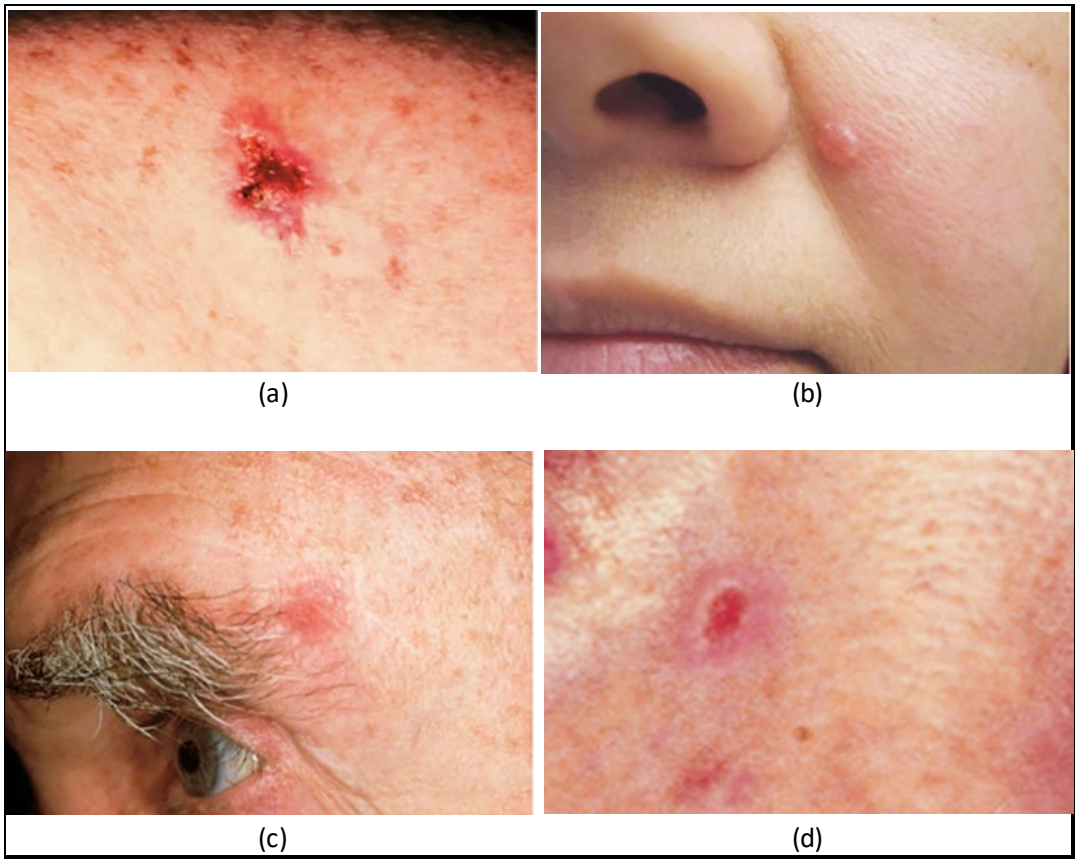


Fig. 3: Images showing the varying appearance of Basel cell carcinoma lesions [20]

b) Squamous cell carcinoma (SCC)

Squamous cell carcinoma (SCCl), which is the second most common skin cancer, mainly occurs on sun-exposed areas of the body, such as the face, ears, and hands. The keratinocyte in the epidermis undergoes a DNA change which may lead to a change in the cells. Consequently, these mutated cells may grow out of control and possibly form SCC. This type of carcinoma is characterised by a red nodule, or a flat lesion with a firm red nodule, or a scaly lesion with a crusty surface. If left untreated, SCC may become metastatic.

Fig. 4 shows the various appearances of SCC. Figure 4a shows a lesion on the back of a hand that presented as a scaly red patch with irregular borders; it may be crusty or bleed. Fig. 4b shows an elevated lesion on a lip with a depressed centre; it may also bleed. Figure 4c and d shows crusty lesions on the ears; these may appear as open sores or warts and may bleed.

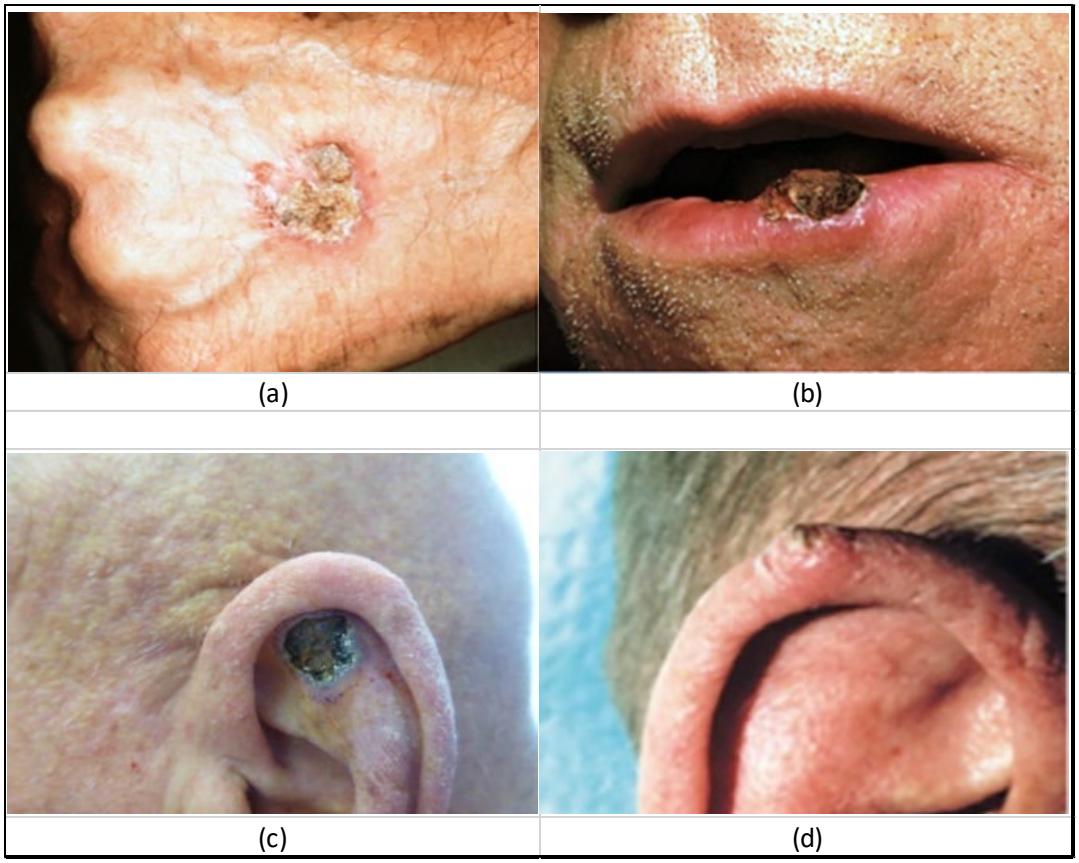


Fig. 4(a)–(d): Images showing the varying appearance of Squamous cell carcinoma lesions [21]

c) Melanoma

Melanoma, which is often characterised by a brownish black mole on the skin, develops when the melanocyte cells undergo uncontrolled cell division. When the melanocyte cells are exposed to high levels of sunlight it produces more melanin. Melanin is a brown pigment that gives skin its tone. When melanoma occurs in people with dark skin tones it is more likely to occur in areas not normally exposed to the sun, such as the palms of the hands and soles of the feet. It is the least common of the skin cancers, however, it has the highest mortality rate.

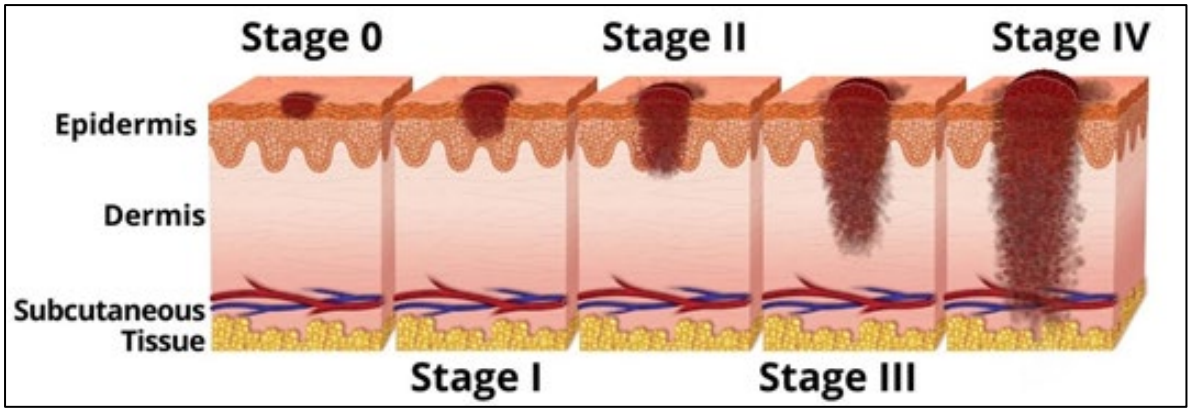


Fig. 5: The 5 stages of melanoma [22]

The lesion thickness, which is measured from the skin surface to the deepest point of the lesion, is an indicator of the progression stage of the melanoma. Melanoma has five stages; these are shown in Fig. 5. Fig. 6a illustrates a melanoma at either stage 0 or stage 1. Stage 0 is when the cancer cells are only in the epidermis and has not grown into deeper tissue; this is referred to as situ melanoma. Stage 1 is when the melanoma is not more than 2 millimetres thick and has not spread to the lymph nodes.

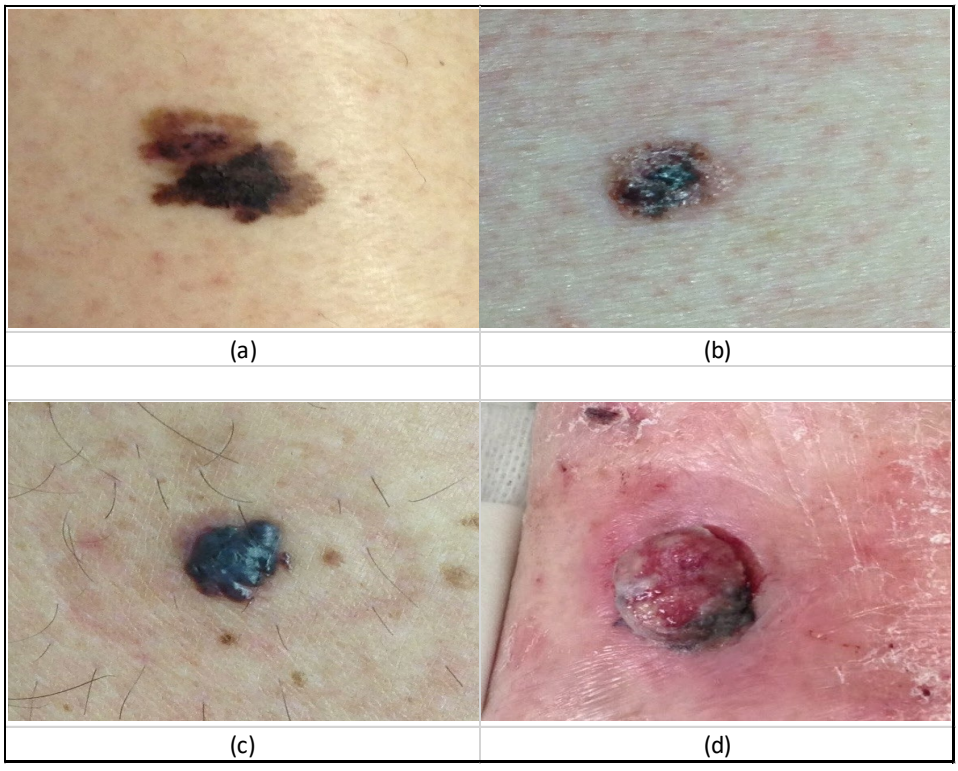


Fig. 6(a)–(d) : Images of melanoma at various stages of progression[23]

Fig. 6b illustrates melanoma at stage 2; it is 1-4 millimetres thick and is only in the skin, and there is no sign that it has spread to lymph nodes or other parts of the body. Fig. 6c shows a melanoma during stage 3; in this case it has spread to the nearby lymph nodes. Fig. 6d illustrates a melanoma at stage 4 whereby the cancer has spread to the lymph nodes and/or other organs, such as the brain, bones, and/or liver. Early diagnosis of melanoma may be critical to the survival rate of a patient. The longer the melanoma is left untreated there may be a greater risk of the cancer progressing from stage 0 towards stage 4.

4. Detection/Diagnosis of Melanoma

Melanoma is the most lethal of the skin cancers; this is owing, in the main, to its metastatic nature. The number of cases of melanoma has continued to increase over the last 50 years [24] and the ability to detect this type of cancer early is vital for recovery. Detection requires both a physical examination and a biopsy; these are discussed below.

a) Physical Examination

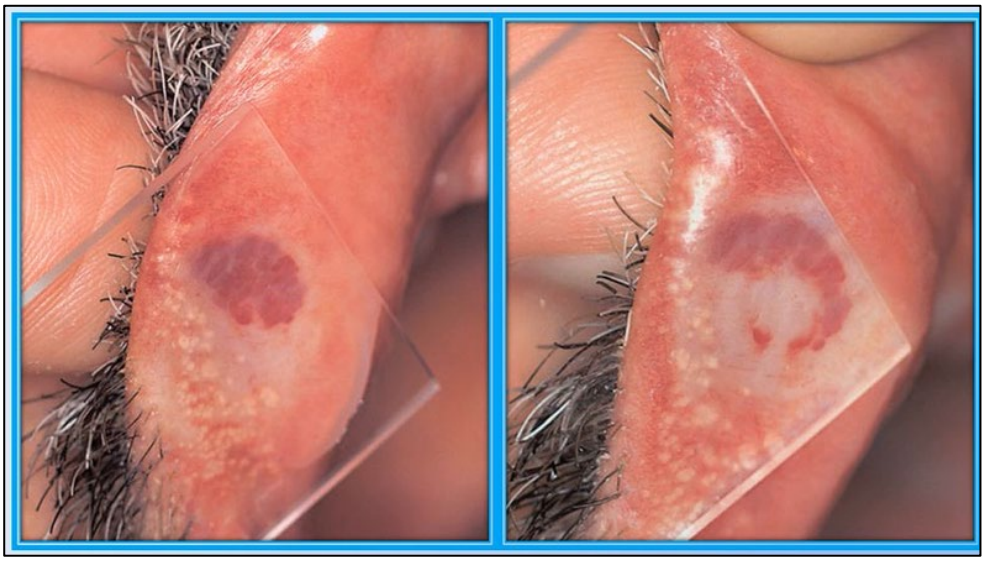


Fig. 7: Diascopic test shows blanching of lesion [25]

The first step in diagnosing melanoma is a visual examination. This may be conducted monthly through self-examination or annually by a doctor [26]. A doctor may use a diascopy to assess the blood

dilation and pigmentation of the lesion. This is done by placing a glass slide firmly against the lesion and seeing if the redness is either caused by blood in the dilated superficial vessels or due to haemorrhaging [27-29]. In the case of the former, blanching may occur; this is illustrated in Fig. 7.

Fig. 8: (a) Handheld dermoscope and (b) digital dermoscope [30]

A dermoscope, which is also referred to as a dermatoscope or an epiluminescence microscope, is a handheld magnifying instrument that may be used by a doctor for a closer visual examination of a lesion; this is shown in Fig. 8a. A digital dermoscope is an enhanced device that can digitally take a high-resolution photograph of a lesion. This type of device, which is shown in Fig. 8b, allows for lesions to be inspected more closely as well as compared to past and future images of a lesion site.

Fig. 9: a) Normal light reflecting of the skin and (b) the resultant digital image captured [31]

Fig. 9 shows an image of a lesion captured using normal light reflecting off the surface of the epidermis. A dermoscope uses polarised light to view underlying tissues below the epidermis. Fig. 10 shows how the polarised light filters are used to capture images with greater detail of underlying features compared to using normal light.

Fig. 10: a) Polarised light refracting below the epidermis and (b) the resultant image captured on a dermoscope [31]

Non-invasive detection and classification of melanoma is usually performed using a combination of three methods [32, 33]; these are the (i) ABCD/E rule, (ii) Menzies method, and (iii) 7-point checklist. The ABCD/E rule is an acronym for asymmetry, border, colour, diameter, and evolution, of the lesion.

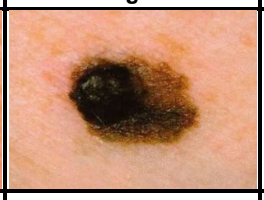
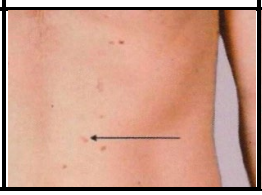
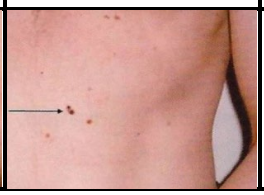
The Menzies method investigates eleven features. Two are negative features; these include observing (i) a symmetrical pattern and (ii) the presence of a single colour. Nine are positive features; these include (i) a hazy blue-white veil, (ii) focal collections of multiple brown dots, (iii) pseudopods, (iv) radial streaming, (v) scar-like depigmentation, (vi) peripheral black (not brown) globules, (vii) multiple colours (5 or 6 colours excluding white), (viii) multiple irregular sized blue/grey dots, and (ix) broadened pigment network. If the lesion under investigation has at least one of the positive features and neither of the two negative features, the result may be diagnosed as positive.

The 7-point checklist considers three major and four minor criteria; the former has a weighted score of 2 each and the latter has a weighted score of 1 each. The major criteria are (i) atypical pigment network, (ii) atypical vascular pattern and (iii) a blue-whitish veil. The minor criteria are (i) irregular streaks,

(ii) irregular pigmentation, (iii) irregular globules and (iv) regression/pepper-like areas. A negative diagnosis is obtained if the cumulative score is less than 3.

b) Application of the ABCD/E rule

TABLE I:
Images showing the ABCDE rule **[34]**

	Benign	Malignant	
Asymmetry The benign lesion is symmetrical			The malignant lesion is asymmetrical as the two halves do not match
Border The border is smooth and even			The borders are irregular and uneven
Colour The lesion is a single colour and is usually a shade of brown			The malignant lesion has several colours with different shades of brown, black, red, white and blue-grey.
Diameter Benign lesions are usually less than 6mm			The malignant lesion is generally larger and has a diameter greater than 6mm
Evolving The lesion does not change over time			The malignant lesion evolves by changing colour, shape and thickness.

In 1985, Dr Robert Friedman developed the method for visually differentiating between benign nevi, which are raised growths on the skin that are often dark, and malignant melanoma [35]. The clinical characteristics of early melanoma was referred to as the ABCD criteria. This refers to asymmetry, border, colour, and diameter. Later, E was added to the acronym to represent evolving. This ABCDE method is described in Table I.

5. Biopsy procedure of Melanoma

A biopsy is performed by a physician where a suspected lesion is partially or fully removed and sent to a pathology laboratory for analysis. The analysis of the lesion under a microscope could be used to determine if the lesion is benign or malignant. The suspected lesion is removed through scraping and/or cutting (excisional) or using a punch (circular blade); the former is illustrated in Fig. 11a and the latter in Fig 11b. The excisional biopsy uses a traditional scalpel to completely remove the lesion and some healthy tissue around the lesion. The exposed area may be stitched closed. Local anaesthesia is commonly used for a skin biopsy.

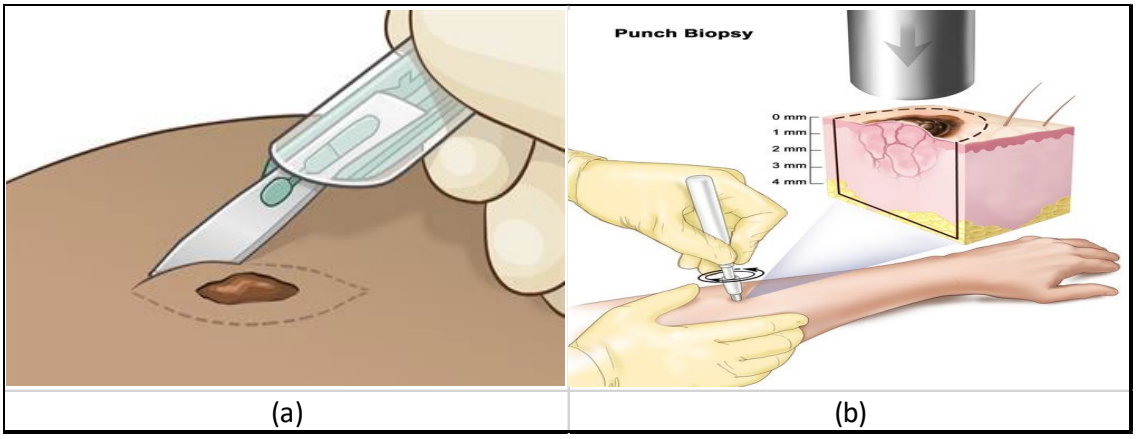


Fig. 11: Biopsy of lesion a) excisional biopsy, b)punch biopsy[26]

C. Problem definition and challenges

There are several stages required in estimating the presence of melanoma in dermoscopic digital images of skin lesions; these include image acquisition, filtering, segmentation, feature extraction and classification.

Image acquisition requires sourcing a relevant and reliable database of dermoscopic images that would allow for the training and testing of the system. Image filtering involves the pre-processing of the images using one or more filters to remove noise and unnecessary features like hair as well as enhance features by filling in holes for further processing. The image will also undergo light correction to reduce the effects of glare. The removal of noise from the images may improve the contours of the image for

segmentation. Image segmentation involves the delineation of the region of interest (foreground) from the rest of the image (background). This is based on the characteristics of an individual pixel or a cluster of pixels. This assists with determining the background and foreground of the image. The lesion needs to be segmented from the background to analyse the features. Feature extraction involves the implementation of the ABCD method. The acronym ABCD denotes asymmetry, borders, colour, and diameter of the lesion and represents the method of visually identifying the features of melanoma. These features need to be analysed and quantified to provide data that will be used to determine the classification of the lesion. Image classification is used to estimate the probability of the reference image indicating the presence of melanoma. The classification must ensure that false negatives are minimised as this would result in potential misdiagnosis and possible delay in treatment. The latter has shown to reduce the survival rate of patients.

D. Structure of dissertation

This dissertation comprises five chapters with preliminaries and an appendix. Chapter 1 introduces the study and highlights the relevance and need for the study. The biology of human skin is introduced with the various skin lesions and its characteristics. It discusses the traditional procedures with respect to detection and diagnosis of melanoma. It also identifies challenges and faced during the study and presents the scope limitations. Chapter 2 provides the relevant literature review. This chapter introduces the technology that combines medicine, engineering, and computer science. This chapter also highlights the previous work done in the field of computer vision-based melanoma classification. Chapter 3 looks at the research methodology employed in the study. It provides a high-level overview of the proposed algorithm and its implementation. It also discusses the tools, hardware and software used to accomplish the research outcomes. Chapter 4 analyses the data gathered from the research. The data is extracted and subsequently classified using decision support tools. The proposed system is evaluated using accuracy, sensitivity (true positive) and specificity (true negative). The proposed system is compared to other similar studies. Chapter 5 provides the conclusion and possible further research. It also recognises the shortcomings and improvements of this research.

E. Scope Limitation

This research is based on the ABCD method of melanoma detection. It does not include the E for evolution as this research provides a snapshot of the diagnostic process and does not include long-term changes to the lesion.

The images used in this research are from the Interactive Atlas of Dermoscopy (EDRA) database. This was chosen as it was the only database to provide thickness of lesion and was used to train resident dermatologists. For these reasons, other databases, such as ISIC, HAM 10000 and PH^2, were excluded from this study. The patient's metadata, such as age, gender, race, family history and location of lesion, is not included in the classification process.

Classification methods are limited to three popular image classifiers; these include k-nearest neighbours (KNN), Naïve Bayes and linear support vector machines (LSVM). The database available that contains lesions in situ or less than 0,76mm thick is relatively small and would not be sufficient to train and test a deep learning neural network. Therefore, the deep learning neural network was excluded from this study.

F. Contributions of the Dissertation

This study provides a classification algorithm to be used as a decision support tool by physicians in the early diagnosis of melanoma that is specifically in situ or less than 0,76mm in depth. Furthermore, the accuracy of the algorithm has been balanced to provide a high false negative rate.

II. LITERATURE REVIEW

A. Dermoscopic Image processing

Medical imaging refers to the electromagnetic (EM) or acoustic emission and capture methods used to obtain images or representations of the body, bones, or tissue for diagnostic purposes. Images captured using acoustics or the low energy non-ionizing range of EM frequencies are non-invasive and relatively harmless to humans. Improvements in these technologies over the last few decades has led to major advances in medical imaging accuracy [36].

Medical imaging came to the fore in 1895 when the first X-Ray was developed using high frequency EM radiation to form an image on a sheet coated with barium platinocyanide. This discovery started a flurry of research into medical imaging across the world. The X-Ray is still used in modern medicine and imaging is essential in assisting with diagnosis and treatment.

There are many types of medical imaging used currently and it is dependent on the nature of the tissue or organ that needs to be viewed. This includes X-Rays, computerised tomography (CT) scan, magnetic resonance imaging (MRI), positron emission tomography (PET) scanner, ultrasounds and medical(clinical) photography [37].

Medical images are acquired to assist with diagnosis and further studies of the medical condition. The use of medical imaging is increasing owing, in the main, to the ease, availability, and improvement of technology. For example, as previously discussed, a digital dermoscope is able to capture the image of skin lesions, including with the use of polarising filters; subsequently, these images may be uploaded to a computer, post-processed and, with the use of algorithms, analysed.

Fig. 12 illustrates a flow diagram of the steps required to perform computer aided diagnosis from the acquisition of the image to the classification of the image. The reliability of the final output is dependent on the outcome of the prior cascaded stages. Input and output errors at each of the stages may be compounded and dramatically affect the accuracy of the overall result.

Fig. 12: The five stages towards Computer Aided Diagnosis

1. Stage 1: Image Acquisition

The first step to creating an image processing system is to gather the data to form the image. A digital dermoscope uses a polarised light source that connects to a digital camera with a CCD array or a CMOS sensor to acquire the digital image. The digital image is uploaded and in read into the computer programme. The image may be resized to ensure uniformity of the dataset as the cameras that captured the images may be set to varying sizes.

There are several image datasets available for use in detecting melanoma from clinical and dermoscopic images.

The PH2 dataset contains 200 dermoscopic images of lesions. This includes 160 nevi and 40 melanoma images. The images have been manually segmented to provide the ground truth [38]. The images are sized to 768x560 pixels at x20 magnification. All images underwent clinical diagnosis, however only the highly suspicious lesions have been verified by a histological pathology test.

The Interactive Atlas of Dermoscopy (EDRA) was created to train resident dermatologists. The dataset contains over 2000 images of clinical and dermoscopic images from over 1000 patients. The dermoscopic images were captured at 768x512 pixels[39]. The lesions are graded in four thickness levels namely, in situ, less than 0,76mm, 0,76mm-1,5mm and greater than 1,5mm.

The Human Against Machine (HAM10000) dataset is a collection of 10015 multisource dermoscopic images. The images are obtained using various types of dermoscopes and is stored using various mediums[40]. The dataset has 7 different categories of skin lesions with over 50% being confirmed via a histological pathology test and the rest by expert consensus or microscopy. The images obtained from the Medical University of Vienna have been digitised from diapositives using a Nikon scanner. The images were enhanced and resized to 800x600 pixels at 72dpi to create a standardised format in the dataset.

The International Skin Imaging Collaboration (ISIC) contains, as of 2020, 33126 dermoscopic images for training and 10982 for testing from more than 2000 patients [41]. The ISIC dataset is a collection

of numerous smaller datasets to form a large multi-source and multi country dataset with all melanoma cases have been verified by histological pathology tests. Cassidy et al. [42] has found over 14000 duplicates within the training data and also between the training dataset and test dataset.

2. Stage 2: Pre-processing and Filtering

The image data from the digital camera or dermoscope in its raw form may not produce the desired results when processed for feature extraction [43, 44]. The skin lesion needs to be isolated from the normal skin [45]. The raw image may contain irrelevant information that needs to be removed to extract the useful information and features from the image. To extract useful information, the image must undergo pre-processing and filtering. It is a cleansing process that prepares and standardises the images before building a system or model [46]. The aim of pre-processing is to remove unwanted noise and to make the desired features of the image more visible [47]. Some examples of these unwanted objects or noise include hair (refer to Fig. 13(a)), blood vessels (refer to Fig. 13(b)) and texture variation, such as trapped air bubbles (refer to Fig. 13(c)).

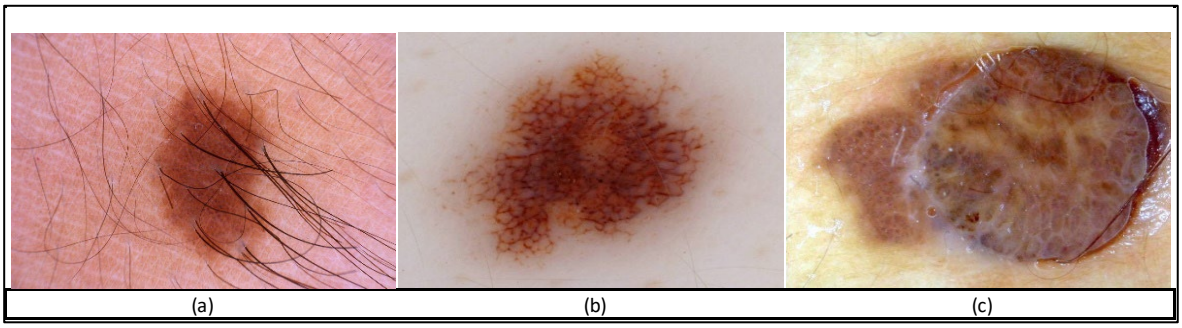


Fig. 13: Examples of noise and unwanted objects: (a) hair, (b) blood vessels and (c) trapped air bubbles [48]

Common methods for pre-processing of images of skin lesions include (a) image enhancement and (b) noise removal and image restoration [49].

a) Image Enhancement

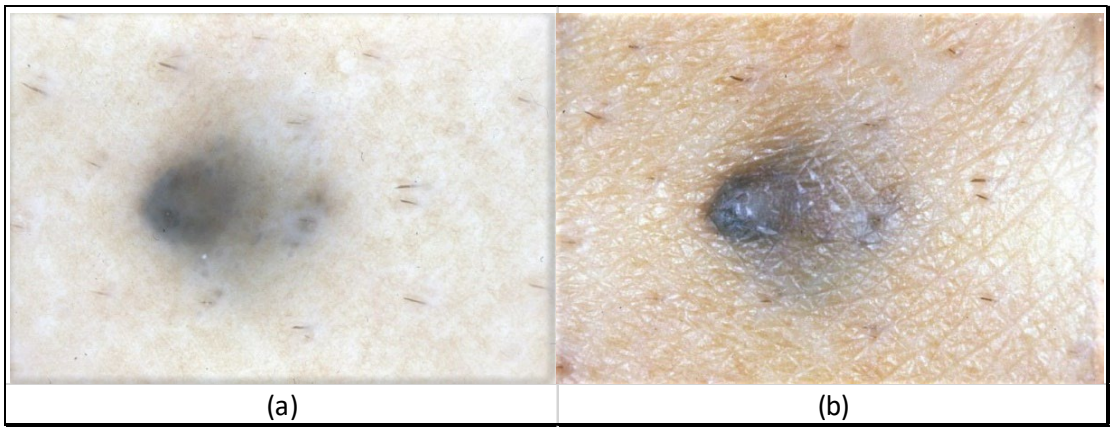


Fig. 14: Qualitative Image Enhancement, a) original image, b) enhanced image [50]

Image enhancement is a process of modifying or adjusting digital images. Image enhancement may be qualitative or quantitative. An example of qualitative enhancement is shown in Fig. 14, in this case, the image is adjusted to look better to a human observer (psychophysical) [51], An example of quantitative enhancement is shown in Fig. 15. In this case, the image is adapted so as to be more appropriate for a task like edge detection [52, 53].

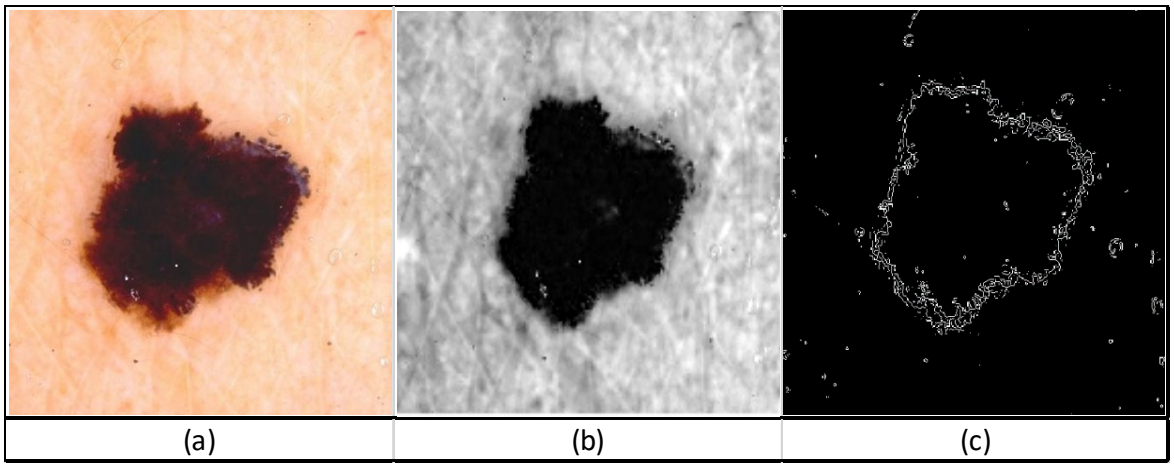


Fig. 15: Quantitative Image Enhancement, a) original image b) enhanced image for c) Edge detection[54]

Image enhancement is subjective and depends on what type of information the application requires. Image enhancement may be performed in both the frequency and spatial domains [55, 56]. Types of

enhancements may include, among others, brightness [56], saturation [57], contrast [55, 58], and histogram equalisation or stretching [59, 60] [61].

b) Noise Removal and Image Restoration

Noise refers to an unwanted, random signal that degrades the useful information in an image. Noise may be produced by internal and external factors in the process of creating a digital image, such as from lighting, the camera sensor temperature, lens distortion, sampling, dust and the transmission medium [62, 63]. Moreover, raw digital images are often quantised as part of the digital capture process, and this may consequently generate additional noise.

Camera sensors and transmission errors can create salt and pepper noise which refers to the degradation of the image by random maximum intensity(white) or dead (black) pixels in an image. It is also referred to as impulse noise.

Gaussian noise is used to model thermal noise and can be caused by the counting of photons especially under low light levels[64]. Gaussian noise is distributed evenly over the entire image and is additive.

Gaussian noise, $G(x)$, is mathematically represented by equation

$$G(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \quad (1)$$

Where x is the gray level image pixel, μ is the mean grey value and σ is the standard deviation.

Noise removal may be attained through the use of filtering; these may include, among others, mean [55, 65], median [65], and Gaussian [62] filters.

The mean filter replaces the value of the target pixel by the mean/average value of the neighbouring pixels. Therefore, the mean filter smooths an image. It uses a kernel, usually 3x3, larger kernels may be used for greater smoothing however this will increase blurring and reduce the edge preservation[55, 65].

The median filter may be used to remove salt and . Instead of replacing the target pixel value with the average value of the kernel, it is replaced by the median value. Therefore, a pixel that is noise or

unrepresentative of the image will not affect the mean value of the pixels within the kernel. It is a nonlinear operation and is more effective than mean filter as it reduces noise and preserves the edges[65]. The median filter does not suppress Gaussian noise very well. A Gaussian filter works very well to remove Gaussian/normal distribution noise. A Gaussian filter blurs or smooths the image to reduce intensity of the noise.

3. Stage 3: Image Segmentation

Segmentation refers to a process of splitting an image into different regions. The regions are dependent on what information is needed from the image. One type of segmentation involves the delineation of a region-of-interest (foreground) from the rest of the image (background). There are several ways to segment an image; these include, among others, colour, texture, boundary, and edge detection. Automatic digital image segmentation is an extremely active field of research and is one of the most crucial components in image processing and computer vision [62]. There are three broad categories for image segmentation [45]; these include (i) morphological, (ii) pixel, and (iii) region-based segmentation. A common morphological segmentation method is edge detection; some of the most popular techniques include Roberts, Sobel, Prewitt, Laplacian, and Canny [54, 66, 67]. Roberts, Sobel and Prewitt edge detection methods finds the edges where the gradient of the image is at a maximum using the respective approximation to the derivative[66]. The Laplacian edge detector finds edges by looking for the zero crossings and the Canny detector looks for the local maxima of the gradient of the input image. The canny edge detector uses two thresholds to find weak and strong edges and is therefore less sensitive to noise[54, 67].

Another of this segmentation method includes the use of erosion and dilation [62, 68, 69]. Dilation thickens the objects in an image whereas erosion thins the objects in a binary image. Erosion and dilation use a structuring element which is usually a small, odd dimension matrix. The structuring element moves to all pixels in the image and the pixel is adjusted according to the values of the neighbouring pixels. A common pixel-based segmentation algorithm is the k-means unsupervised clustering algorithm [70] [55]. K-means is performed in 2 stages. First, it calculates the k centroid and thereafter allocates each point to the nearest

centroid[70]. K-means is an iterative process that calculates the mean intensity for every class[55]. Each pixel is placed in a class with the closest mean, thereby segmenting the image.

Region-based segmentation may include the use of threshold values to separate objects in an image into different sub-regions of connected pixels, such as active contours [71][72]. The region of interest can be separated from the background by defining a single threshold value and is referred to as global threshold. Thresholding produces a binary image b from an intensity image I and T is the threshold value. The initial area of interest is specified and the curves evolve towards the boundaries on every iteration[72]. The closer the initial contour position is to the desired boundary, the faster the desired segmentation result can be obtained.

4. Stage 4: Feature extraction

On one hand, image segmentation isolates a region-of-interest (ROI), on the other hand, feature extraction attempts to identify distinguishing features or patterns from either an entire image or ROI that mat be used to enhance the efficacy of a classifier. For skin lesions, feature extraction is based on the dermoscopic ABCD (asymmetry, border, colour, and diameter) rule.

a) Asymmetry

There are various methods to calculate the asymmetry of a lesion. The centre of mass or centroid of the lesion is calculated. Then the image is folded or reflected along the x and y axis, as shown in Fig. 16, and finding the difference in the overlapping area will provide the degree of asymmetry as shown in Eq. (1):

$$Asymmetry\_Ref = \frac{\Delta A_{min}}{A_{total}} * 100\,, \tag{2}$$

where ΔA_{min} is the difference in area of the folded or reflected regions.

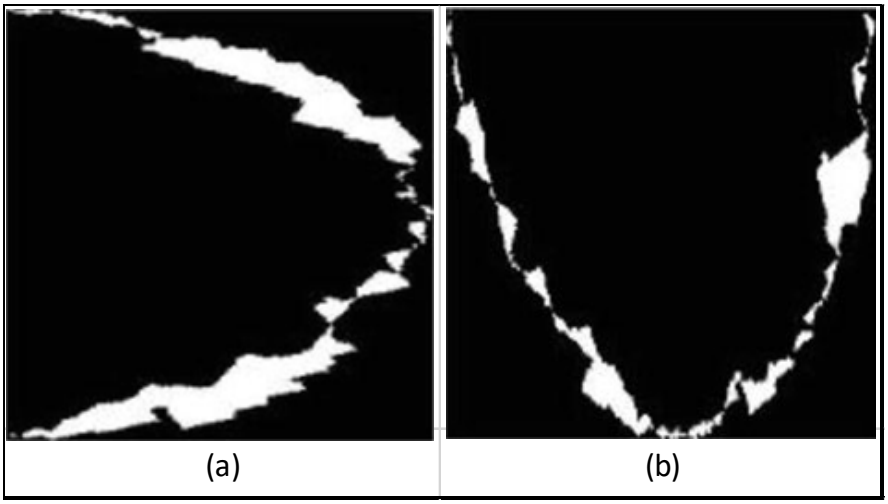


Fig. 16: Area without overlap: a) folded along y axis and b) folded along x axis[73]

Symmetry Distance (SD) may also be used to measure asymmetry [74]. The SD measure calculates the average displacement of several vertices as the original shape (P) is transformed into a symmetric shape (ST) as given by Eq. (2):

$$SD = \frac{1}{n}\sum_{1}^{n}\left\| P_i - \acute{P}_i \right\|^2 \quad , \tag{3}$$

where $P_i$and $\acute{P}_i$ are points in P and ST(P) respectively.

Skewness of an image can be used a measure of asymmetry[75]. Skewness is the amount of deviation of a projection from symmetry. The skewness of the horizontal and vertical skewness is given by Eqs. (3) and (4), respectively:

$$Skewness_H = \frac{M_{30}}{\sqrt{M_{20}^3}} \tag{4}$$

and

$$Skewness_V = \frac{M_{03}}{\sqrt{M_{02}^3}} \quad . \tag{5}$$

The degree of skewness is by using 2 third order moments, M_{30} and M_{03}. The image is biased to the left if the skewness if negative and conversely biased to the right if the skewness is positive.

b) Border Irregularity

Melanoma lesion border is characterised by an irregular and poorly defined border. The edge pixels of the segmented lesion make up the border. The isoperimetric quotient (IQ) uses the perimeter (p) and area (A) of the lesion to determine the border irregularity given by Eq. (5):

$$IQ = \frac{4A\pi}{p^2} \tag{6}$$

A circle provides an isoperimetric quotient of 1, and since melanoma is characterised with uneven or irregular borders, the isoperimetric quotient of melanoma will approach 0.

Another method of determine border irregularity is to analyse the turning points of the border [76]. The points of a function where the derivative changes sign may be referred to as the turning point. Turning points along the border of the lesion indicates the border irregularities.

The fractal dimension, D, of the lesion border is also an indicator of the smoothness of the border[77]. A smooth border produces a value closer to 1 and an irregular border provides a fractional dimension closer to 2. Using the box counting method [78], as shown in Fig. 17, the relationship between the fractional dimension and the and the number of boxes with borders, N(r), is given by Eq. 6:

$$N(r) = \lambda r^{-D}, \tag{7}$$

where, r, is the grid size of boxes and λ is a constant.

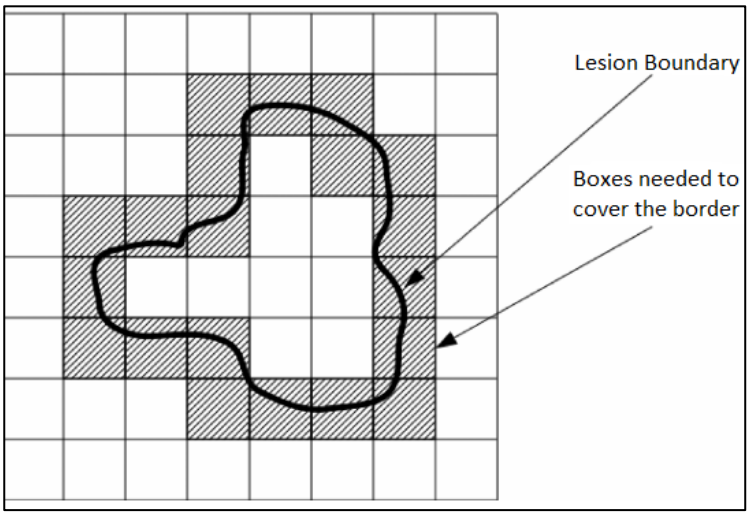


Fig. 17: Fractional dimension using the box counting method [79]

c) Colour

According to the Skin Cancer Foundation benign skin lesions are generally a single shade of brown [80, 81]. Melanoma is generally made up of up to six colours, namely, light brown, dark brown, black, red, white, and blue, as shown in Fig. 18. A blue-whitish veil in the lesion in a strong indicator of a positive melanoma diagnosis.

The light and dark brown is the melanin in the epidermis. Black is the melanin in the granular layer. Red and white is caused by inflammation and regression respectively. The blue melanin in the papillary layer. Each of these six colours, that is present in the lesion, will be awarded a score of 1 and then added to form a total colour score from 1-6.

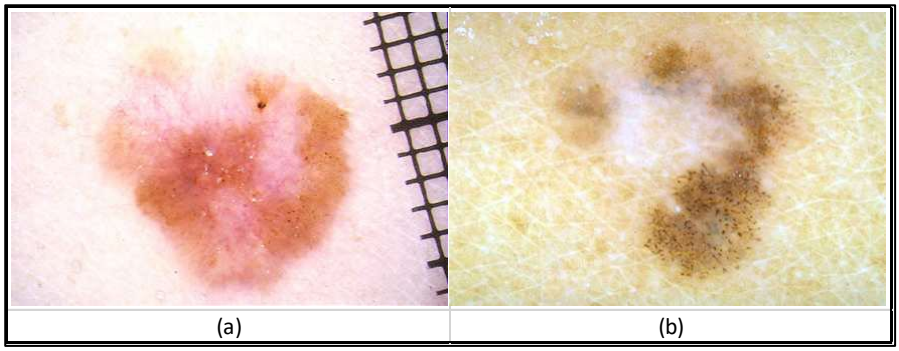


Fig. 18: Examples of colour scores for melanoma. a) Dark, light brown, red and white present, b) blue, dark brown, light brown and white present[82]

d) Diameter

Melanoma is typically greater than 6mm in diameter. There have been several methods used to calculate diameter of a lesion.

One approach is to draw a line from all edge pixels through the centre and then determine the average diameter [83]. A second approach is to find the coordinates of all the edge pixels and subsequently calculate the largest distance between pixel pairs that pass through the centroid; this is referred to as the major axis [84]. A third approach is to consider the Ferets diameter, which is the maximum distance between any two boundary points on the antipodal vertices of a convex hull that encloses the lesion [75]. A fourth option is to use Eq. (7):

$$D = \sqrt{\frac{4A}{\pi}} \quad , \tag{8}$$

where A is the area of the lesion[85].

5. Stage 5: Classification

Classification in this study refers to the technique of determining (or classifying) whether a lesion is melanoma or non-melanoma. The task of classifying images is relatively simple for humans, however, the interpretation and classification of these tasks by a machine is very complex. Over the past 20 year, with the exponential increase in computer processing power, especially GPUs, there has been significant improvements to both image capture devices and the efficiency of image processing.

TABLE II:
TOTAL DERMOSCOPIC SCORE AND WEIGHT FACTOR FOR THE ABCD CRITERIA [86]

ABCD Rule		
Diagnostic Criteria	**Score**	**Weight Factor**
Asymmetry	0–2	x 1.3
Border pigment pattern	0–8	x 0.1
Colour variation	1–6	x 0.5
Diameter and Different structures	1–5	x 0.5

The total dermoscopic score (TDS) is a summative scoring system that assigns weighted values to each ABCD feature [87]; refer to Table II. The TDS score ranges from 1-8, 9 and a score less than or equal to 4.75 indicates non-melanoma, a score of 4.8 – 5.45 indicates the lesion is suspicious and a score greater than 5.45 indicates the lesion has a high probability of being melanoma.

Artificial neural networks have also been used to successfully classify skin lesions. These techniques are very data hungry and computationally expensive and consequently were not considered as an area of investigation within this research.

This study focusses on three binary classifiers; these include k-nearest neighbours (KNN) [88], naïve Bayes [89], and linear support vector machines (SVMs) [90].

A KNN algorithm is a simple memory-based classification model that calculates the distance from a known test data point to predetermined number, K, of closest training data points. The distance between the test datapoint and the training data point may be calculated using several distance metrics. Examples of

distance metrics include Minkowski, Euclidean, city block, cosine, Chebychev, and chi square. The Minkowski distance metric, d_{st}, is given by:

$$d_{st} = \sqrt[p]{\sum_{j=1}^{n} \left| x_{sj} - y_{tj} \right|^p} \tag{9}$$

where n is the number of dimensions, p represents the order of the norm, and x and y are the data points. The Euclidean distance metric is a special case of Minkowski distance with $p=2$; the city block distance metric is a special case of Minkowski distance with $p=1$; and the Chebychev distance metric is a special case of Minkowski distance with $p=\infty$.

Naïve Bayes is a probabilistic classifier that works well with a high number of input features. Bayes theorem determines the probability of a class, given the probability of a prior classification [91, 92]. The extracted features independently and equally contribute to the classification of a lesion. There are several predictor distributions such as multinomial, Bernoulli, and Gaussian (normal). Multinomial distribution analyses the frequency of the extracted feature and is commonly used to classify text data. Bernoulli distribution is used for classification when the extracted features are binary, either 1 or 0. Gaussian distribution is used for classification by assuming the extracted features follow a normal distribution by calculating the mean, μ, and standard deviation, σ. This is mathematically expressed by:

$$g(x, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2(\sigma^2)}} . \tag{10}$$

An SVM, which was originally developed as a binary classifier, separates data into classes using linear or non-linear classification depending on the type of data provided. The line or boundary that is used to separate the data into classes is referred to as the hyperplane [93, 94]. The optimal hyperplane maximises a margin between the support vectors. There are several different kernels in SVM [95], such as linear, polynomial and Gaussian radial basis function (RBF); these are expressed by Eqs. (13-15). The linear kernel, which is suitable for linearly separable data, is given by

$$F\left(x, x_j\right) = sum\left(x . x_j\right) \ ; \tag{11}$$

the polynomial kernel is expressed by:

$$F(x, x_j) = (x.\, x_j + 1)^d \ ; \tag{12}$$

and the Gaussian RBF kernel, which is valuable when there is no prior knowledge of the dataset, is given by:

$$F(x, x_j) = exp\,(-\gamma \left\| x - x_j \right\|^2) \ , \tag{13}$$

where, $F(x, x_j)$ represents the decision boundary, d is the degree of the polynomial and γ is a value from 0 to 1 that influences overfitting.

A Linear SVM is a common choice for image classification in the biomedical field owing to its ease of use and flexibility. In this research, the optimal model is based on the probability of achieving a high degree of success from generalized unseen or test data and not necessarily on being the best classifier for the known data (overfitting).

B. Related Work

The diagnosis of melanoma from dermoscopic lesions has been an area of interest over the past 15 years owing, in the main, to its non-invasive nature and being considered as an objective decision support tool to physicians. This area of research has escalated over the recent years with the advent of machine learning techniques, providing direction for future studies and adding value to the research area. This section presents a short review of proposed methodology by other researchers.

Jayatilake and Ganegoda [96] segmented the lesion using contour edge detection, extracted the ABCD features and performed the classification using SVM, logistic regression and artificial neural networks. The HAM10000 dataset was used in the study.

Chaiyakhan et al. [97] used Otsu thresholding and clustering to segment the image. Texture features were extracted in addition to the ABCD features and subsequently classified via SVM and deep neural networks (DNN). The ISIC dataset was used with 240 melanoma and 260 benign lesions.

Thanh et al. [98] segmented the image using adaptive curvature and performed ABCD feature extraction. Classification was performed using the total dermoscopic score (TDS) on the ISIC dataset.

Gautam and Ahmed [99] used various online databases and tested the system with 146 images. Images were segmented using thresholding. The ABCD rule and texture analysis was used to extract the features. The images were classified using SVM.

Shiji et al. [100] analysed colour and texture features. The colour features were investigated in the RGB, hue, intensity, and saturation (HIS) domains. The PH2 database, which includes 200 images, was used to train and test the SVM classifier.

Majumder and Ullah [101] used 200 images from the PH2 dataset and segmented the lesions using Otsu and Chan-Vese models. The ABCD features were extracted and classified using artificial neural networks. The ABD shape-based features were calculated by finding the area, centroid, perimeter and the major axis. The 6 colours of interest were detected and scored from 1 to 6.

Amaliah et al. [83] used region growing to segment the images. The ABCD features were extracted using the asymmetry index, compactness index, colour homogeneity and average diameter drawn from edge pixels. The TDS was calculated to classify the images.

Senan and Jadhav [102] used 200 images from the PH2 dataset the PH2 dataset and segmented the images using active contours and morphological operators. The ABCD features were extracted. Asymmetry was calculated by folding the image along the major axis; border was determined by turning points; colour was scored from 1 to 6 based on the colours of interest; and the diameter was calculated using Eq. (7). TDS was used to classify the lesions based on the cumulative score.

Oukil et al. [103] used k-means to segment 200 dermoscopic images from the PH2 dataset. The 6 colours of interest and texture features were extracted. The classification was performed using KNN, SVM and ANN classifiers.

Amoabedini et al. [104] segmented the dermoscopic images using the Radon transform. The ABCD features were extracted with a focus on the colour symmetry and shape of the lesion. The extracted features are classified using TDS of weighted coefficients.

Ali et al. [75] segmented images using Otsu and morphological operations. Three features, viz. asymmetry, colour, and diameter, were extracted. The asymmetry feature was calculated using shape descriptors; colour was analysed by the 6 colours of interest; and the diameter was calculated using Feret's diameter. A decision tree algorithm was used for classification.

Chakkaravarthy [105] used watershed segmentation, together with the Sobel operator and morphological image processing to isolate the region of interest. Three features, viz. border, colour, and diameter, were extracted. The border is calculated using the circularity index; colour score is given by the 6 colours of interest; and the diameter is calculated using the area of the lesion. Classification was achieved using deep learning methods.

Al-Zyoud [106] used 200 dermoscopic images from the PH2 dataset. The dataset provides pre-segmented binary masks isolating the region of interest. A version of ABCD features was employed whereby D represented dermoscopic structures rather than the diameter, and only 5 colours of interest were analysed. The images were classified using the TDS weighting.

Messadi et al. [107] compared 3 methods, including mean shift, grow cut and histogram thresholding, to perform segmentation of 320 dermoscopic images. The ABCD features were extracted. The asymmetry was calculated by folding the image along the major and minor axes parring through the centroid. The border irregularity was determined by the compactness index and radial variance. The colour variations were measured by evaluating the colour distribution of the lesion and the diameter was found by measuring the major axis. Classification is achieved using a multi-layer neural network.

Ma and Tavares [108] used manually segmented images from 3 datasets. They initially considered 14 shape and 36 colour features, however, only 15 features were used due to high computational resources needed. The classification was performed using a tuned SVM algorithm with an RBF kernel.

As evidenced above, the ABCD guideline for identifying the characteristics of melanoma is a common approach for feature extraction. Moreover, traditional classifiers, such as SVM and KNN, along with TDS are commonly used to classify melanoma.

Table 3 below provides a summary of the architectures considered in this research.

TABLE III:
OVERVIEW OF VARIOUS ARCHITECTURES USED, HIGHLIGHTING THE DATABASES, TRAINING AND TEST SET SIZE, AND CLASSIFIERS USED

Researchers	Year	Database	Training/Test size	Feature extraction	Classifiers used
Jayatilake and Ganegoda [96]	2022	HAM 10000		ABCD	SVM Regression ANN
Chaiyakhan et al. [97]	2019	ISIC	500	ABCD	SVM DNN
Thanh et al. [98]	2020	ISIC		ABCD	TDS
Gautam and Ahmed [99]	2015	Various Online	146	ABCD	SVM
Shiji et al. [100]	2020	PH2	200		SVM ANN
Majumder and Ullah [101]	2019	PH2	200	ABCD	ANN

III. PROPOSED METHODOLOGY

In this chapter the methods used to develop the algorithm for melanoma detection are discussed. The proposed algorithm, which is illustrated in Fig. 19, is composed of five stages; these include (a) image acquisition, (b) image pre-processing, (c) image segmentation, (d) feature extraction and analysis, and (e) classification.

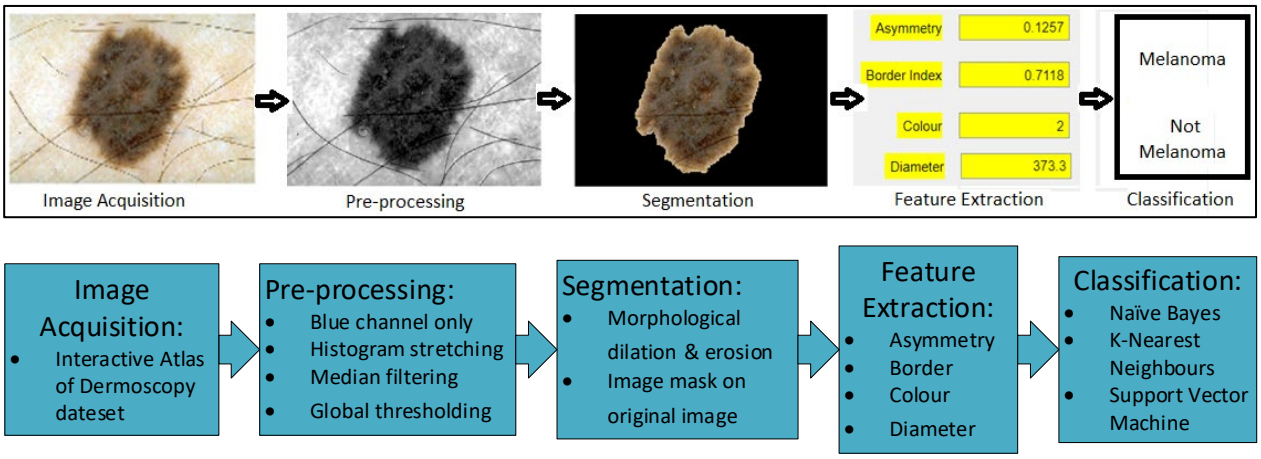


Fig. 19: The 5 stages towards melanoma detection

The research was conducted on a desktop computer with an Intel i7 (9th generation) processor, 32 GB RAM, an SSD storage device, and an Intel 630 graphics device. The operating system used was Windows 10 (64-bit) and MATLAB® version 2020b was used to write the code and create a graphical user interface.

A. Image Acquisition

The dermoscopic image dataset (Interactive Atlas of Dermoscopy) used in this research is provided by Simon Fraser University [59]. This dataset, which is one of the most diverse datasets available [40], is derived from two university hospitals and is used to train dermatology residents [109]. The database represents over 1000 clinical cases; moreover, all lesions were excised, and the clinical diagnosis was confirmed. The dataset includes normal clinical images captured by a Nikon F3 camera and dermoscopic images that were captured using a Dermaphot/Optotechnik dermoscope. The dermoscopic images were captured at a dimension of 768 x 512 pixels at 72dpi [110]. A minor shortcoming of the Interactive Atlas of

Dermoscopy dataset is that the segmentation masks (ground truths) for the lesions are not included; these had to be manually generated for this research.

This research investigates the early detection of melanoma; consequently, the images selected from the dataset needed to be in stage 0/1. The melanoma lesions chosen are in situ or less than 0,76 mm thick. This means the melanoma is still located on the epidermis and has not spread to deeper skin layers or to other tissue [111]. The chosen dataset was the only available dataset that indicated the thickness of the lesion.

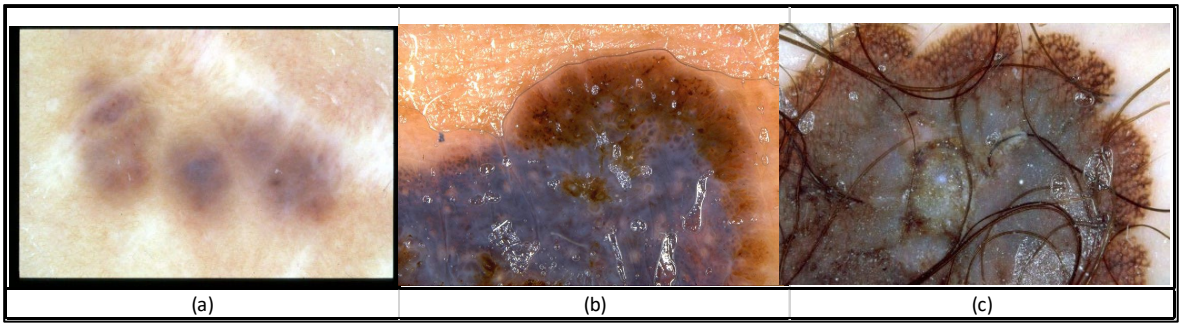


Fig. 20: Sample of images rejected due to being a) out of focus or b),c) incomplete [39]

For this research, 218 images (melanoma and non-melanoma cases) from the dataset were chosen; 178 for training and 40 for testing. In addition, the test dataset was split (20/20) to include an even number of melanoma and non-melanoma cases. The images used from the dataset were carefully selected and needed to be in focus, a complete lesion and all at the same magnification level.

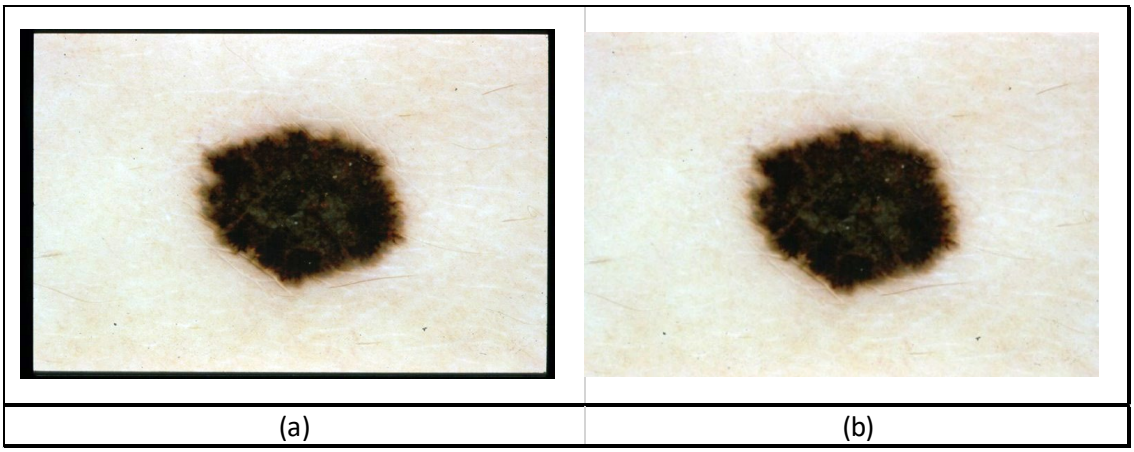


Fig. 21: Original image from database, a) black frame around image, b) black frame removed from image.

It is essential for the images to be complete as the entire border of the lesion is needed to allow for segmentation. An incomplete lesion would affect the asymmetry, border, and diameter values in the feature extraction stage. Images that are out of focus will negatively impact the edge detection for image segmentation. Unknown dermoscopic magnification would lead to unreliable diameter calculations. A few examples of the images that were rejected are shown in Fig. 20.

Some of the selected images contained a black frame/border; this created poor segmentation and border detection results as the area of the unwanted frame exceeded the area of the lesions. To rectify this such images needed to be cropped; a pre-sample and post-sample are shown in Fig. 21(a) and (b), respectively. The cropping of images was performed manually and not considered as part of the pre-processing stage.

B. Image Pre-processing

1. RGB to grayscale

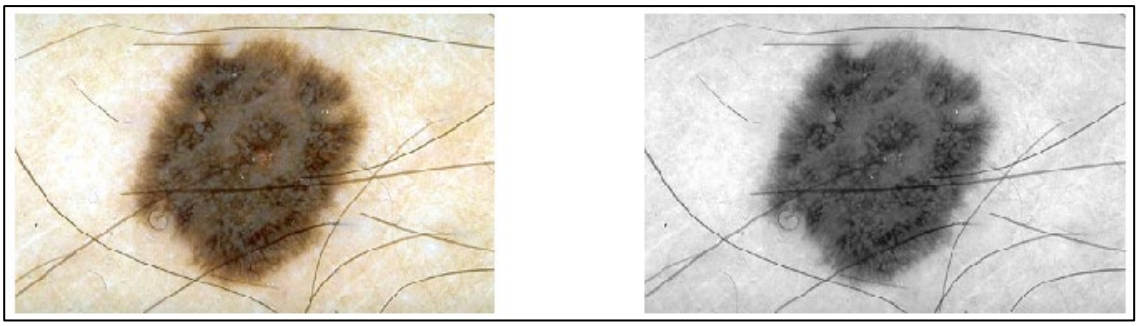


Fig. 22: Pre-processing of image, a) Original RGB image, b) Grayscale conversion of dermoscopic image

The first step of the pre-processing stage was the conversion of the source RGB images into grayscale. This is done to reduce the computational load and allows the algorithm to be applied more efficiently as there is now only one input channel [112]. The saturation and hue information is removed while the luminance (Y) is retained [113]. Fig. 22(a) shows an example of a source RGB image and Fig. 19(b) provides an illustration of the output of the conversion process.

2. Brightness adjustment

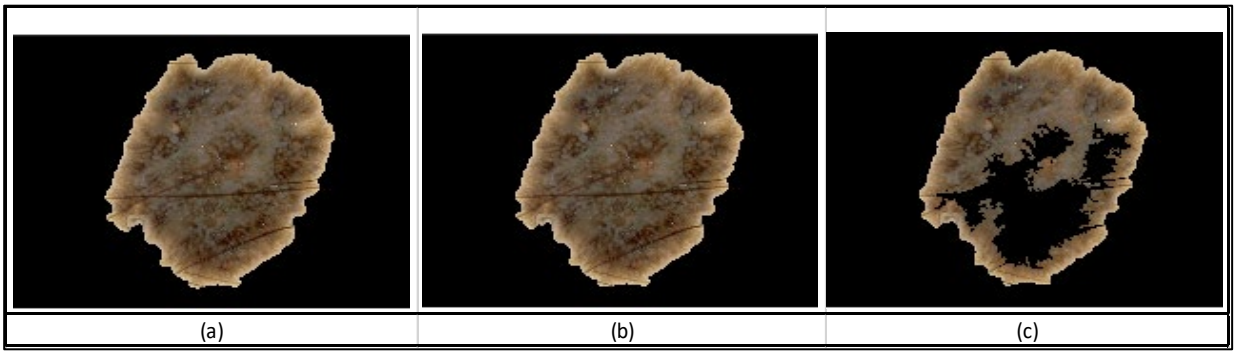


Fig. 23: Segmentation of lesion with various brightness levels a) no change, b) 20% increase c) 40% increase

The image brightness was left unchanged. Through experimental analysis, changing the brightness, saturated the image, which made edge detection and segmentation more difficult. The brightness adjustment was set to zero for the dataset, but it could be increased or decreased if the need arose. Fig. 23(a) shows an example of a segmented region and Fig. 23(b) and (c) provide illustrations of the region when the brightness was adjusted by 20% and 40%, respectively.

3. Contrast adjustment

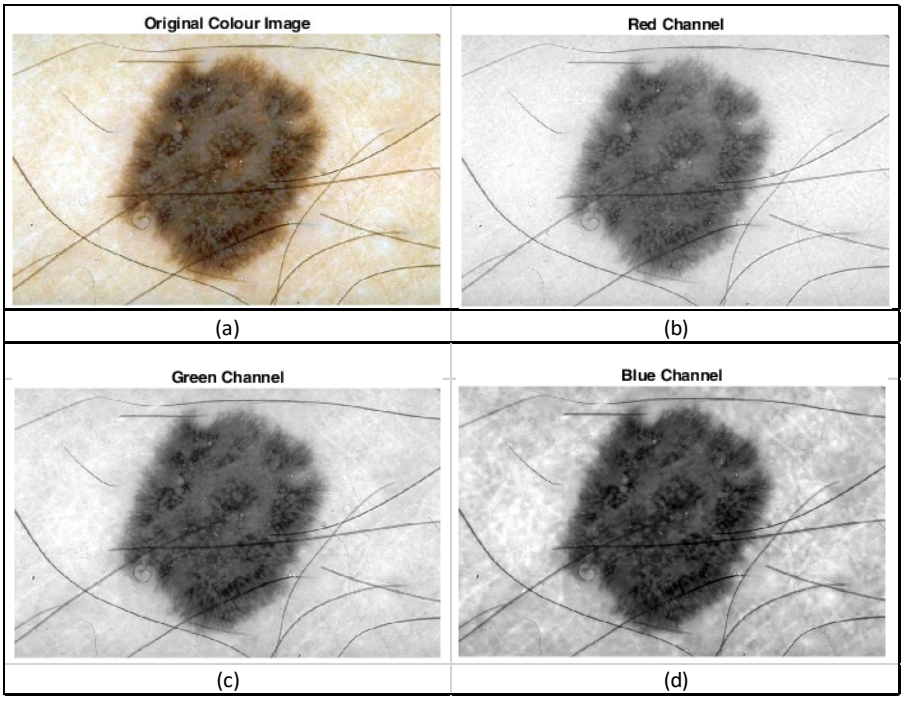


Fig. 24**: Individual RGB colour channels of dermoscopic images, a) Original RGB image, b) Red channel only, c) Green channel only, d) Blue channel only.**

The contrast of the image is increased using histogram stretching. The RGB colour channels were split, and the blue channel was chosen to reduce the computational processing of the algorithm.

From experimentation, the blue channel provided a better degree of contrast when compared to the other colour channels. Fig. 24(a) shows an example of a lesion in RGB, Fig. 24(b) and (c) provide illustrations of the red and green channels, respectively, and Fig. 24(d) shows the blue channel.

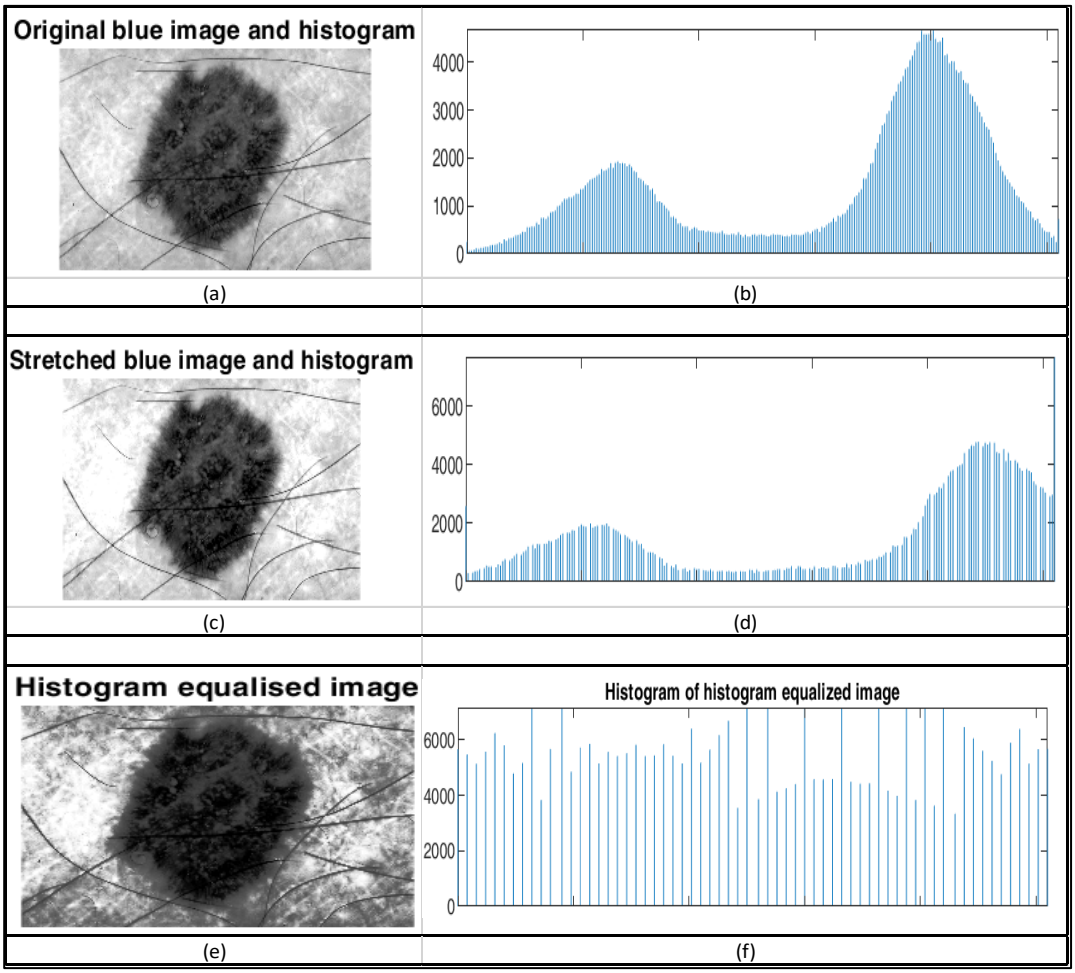


Fig. 25: Histogram stretching: a) Original blue channel image, b) histogram of original image, c) Stretched image, d) histogram of stretched image, e) histogram equalised image with tile size = 8, and f) histogram of equalised image.

Through experimentation, histogram stretching showed to be a more suitable method for image enhancement compared to histogram equalisation. Fig. 25(a) shows an example of a lesion using the blue channel and Fig. 25(b) illustrates its associated histogram. Fig. 25(d) shows the stretched histogram and Fig. 25(c) shows its associated result; in this case the lower 10% and the upper 10% of all pixel values

are saturated. Fig. 25(f) shows the equalised histogram and Fig. 25(e) shows its associated result; in this case the integrity of the border is diminished.

C. Image Segmentation

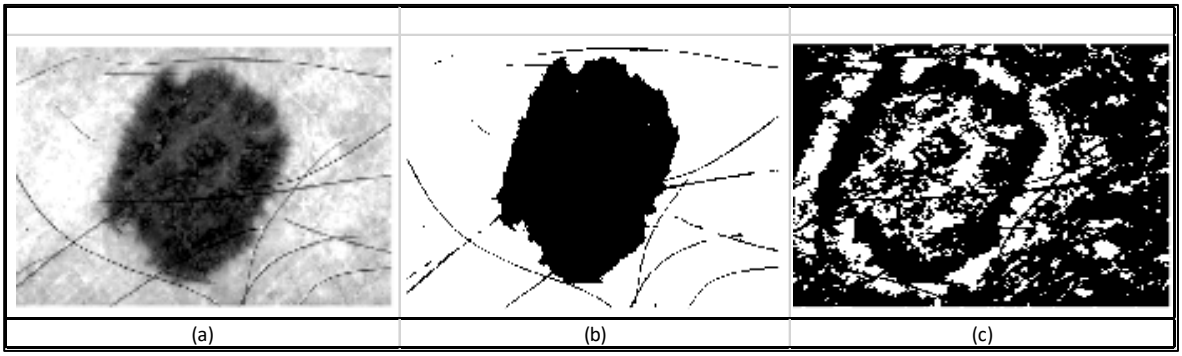


Fig. 26: a) Histogram equalised image b) Image binarization using Otsu's method and c) Image binarization using locally adaptive thresholding.

The salt and pepper noise are removed from the histogram equalised image shown in Fig. 26(a), using a 5x5 median filter. All images were successfully segmented when a 5x5 median filter was applied. The filtered image is binarized using Otsu's method by replacing the values above the global threshold with one and setting the remaining values to zero; this is illustrated in Fig. 26(b). This allows for the foreground to be easily separated from the background. From experiment, Otsu's method showed to be more suitable in converting from grayscale to binary than adaptive thresholding; the latter is shown in Fig. 26(c).

After binarization there may still exist some spurious edges or segments around the region-of-interest, as seen in Fig. 26(b); this may be a consequence of sporadic components such as hair or areas of discolouration. The image colour is inverted to have the white, 255, background and a black, 0, lesion. To remove these segments the binarized image undergoes further morphological processing using dilation and erosion. First, the binary image undergoes dilation where the white background areas are thickened and second, the dilated image undergoes erosion where the white background is thinned. From experiment, a disk structuring element of size 4 was chosen for the dilation and a disk structuring element of size 8 was chosen for the erosion.

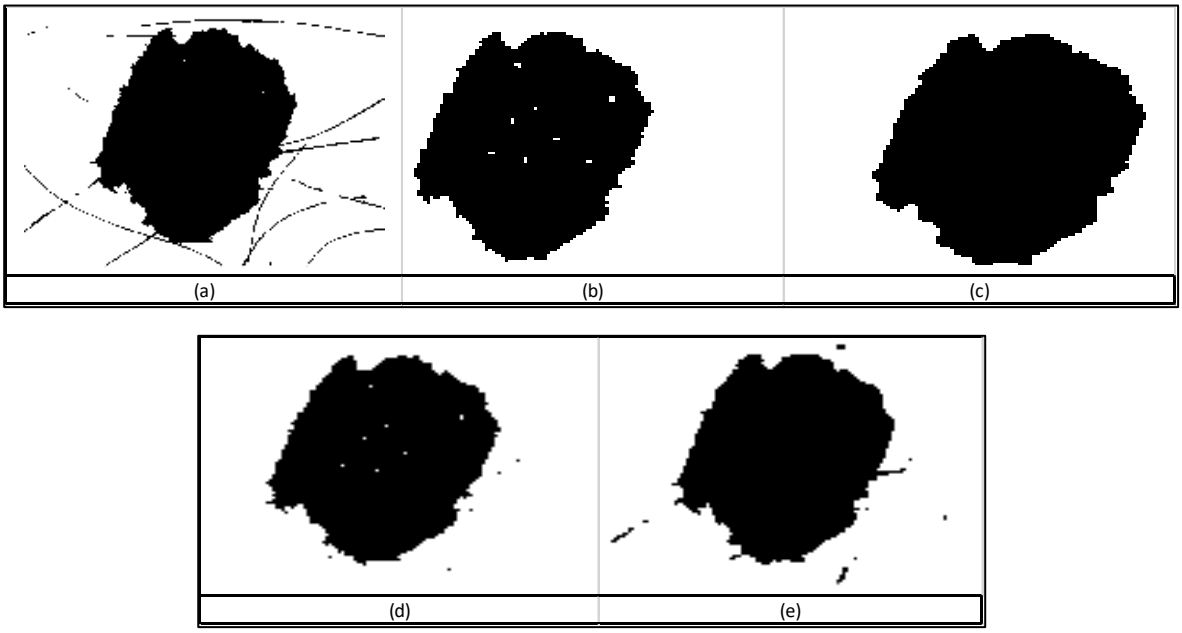


Fig. 27: Hair and noise removal using morphological processes, b) dilation (SE = 4), c) erosion (SE = 8), d) dilation (SE = 2), and e) erosion (SE = 4)

Fig. 27 shows the removal of spurious segments, such as hair, using various structuring elements. The structuring elements for dilation of size 2 and erosion of size 4 are not suitable as there were sporadic connected components still visible; this is illustrated in Fig. 27(d) and (e). This will hinder an accurate boundary detection and segmentation of the lesion. The structuring elements for dilation of size 4 and erosion of size 8 have preserved the boundary integrity satisfactorily, while removing unwanted spurious segments; this is illustrated in Fig. 27(b) and (c). Any larger structuring element has shown to break up weak boundaries in the erosion process and smoothen the edges in the dilation process, thereby altering the perimeter of the lesion. A disk-shaped structuring element was chosen as the structuring element as it is common practice to choose a structure resembling the object being processed.

The processed binary image is subsequently used a mask to delineate the ROI from the rest of the rest of the image; Fig. 28(a) shows the application of the binary mask to the original RGB image and Fig. 28(b) shows the final segmented image.

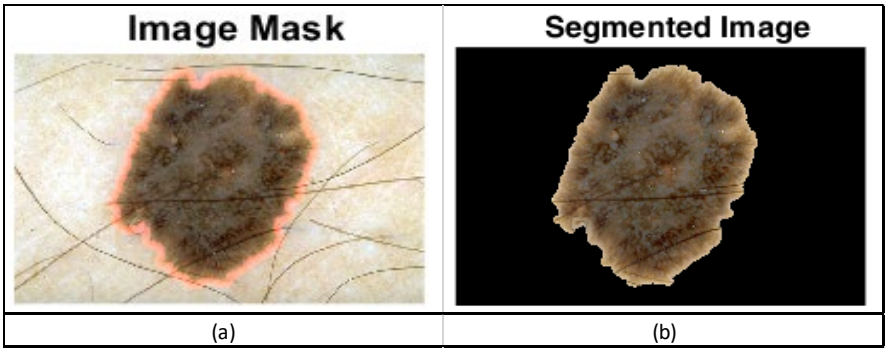


Fig. 28: Image segmentation, a) Boundary mask around the lesion and b) final segmented image

There are cases where there may exist several connected components after the binarization and morphological processing. In these cases, the largest connected component is only selected. Fig. 29 shows three separate binarized images with several connected components. The connected component with the largest area is assumed, with high probability, to be the ROI. Fig. 29c shows an image of a lesion with an unwanted measuring grid; the ROI again has been selected based on the connected component with the largest relative area.

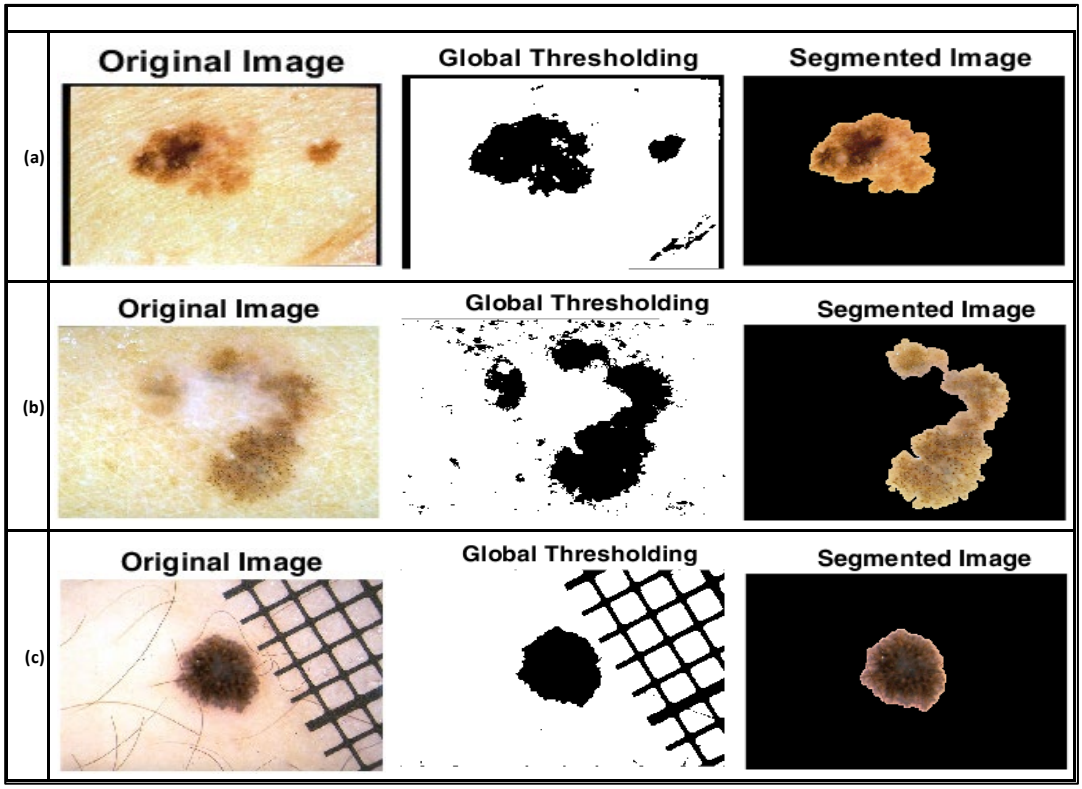


Fig. 29a-c: Examples showing the segmentation of the largest region while excluding the other smaller regions

Image segmentation using binarization and morphological processing have experimentally shown to be more suitable than the traditional edge detection algorithms. Figs. 30 and 31 show attempts at segmentation using some of the common edge detection techniques such as Roberts, Sobel, Prewitt, Canny, and Laplacian.

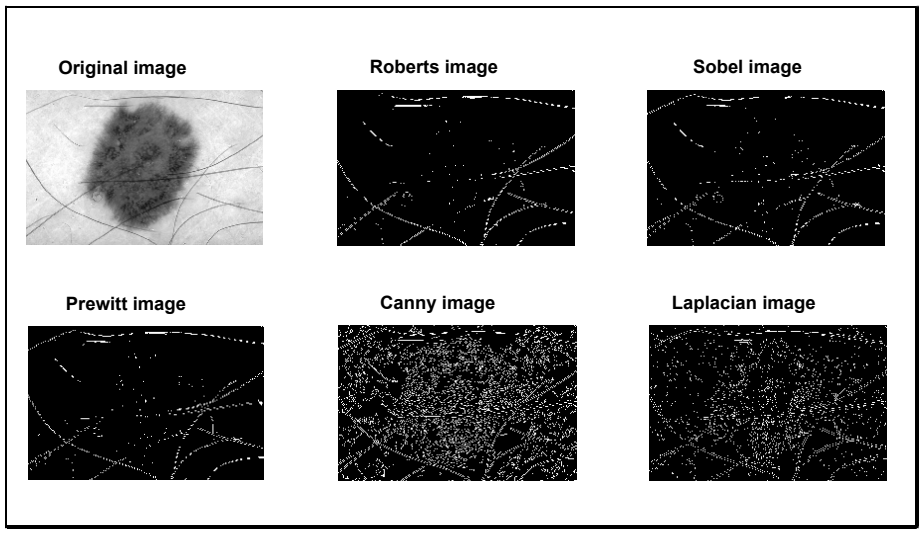


Fig. 30: Various edge detection methods for edge detection of a skin lesion, sample1

Accurate detection of the border is crucial as the border is used in the feature extraction calculation of asymmetry and border irregularity. Furthermore, the colour detection will be affected if the border detection includes too much of the background or removes excessive foreground during segmentation. Consequently, Otsu binarization combined with morphological processing has been chosen over the use of standard edge detection in this research.

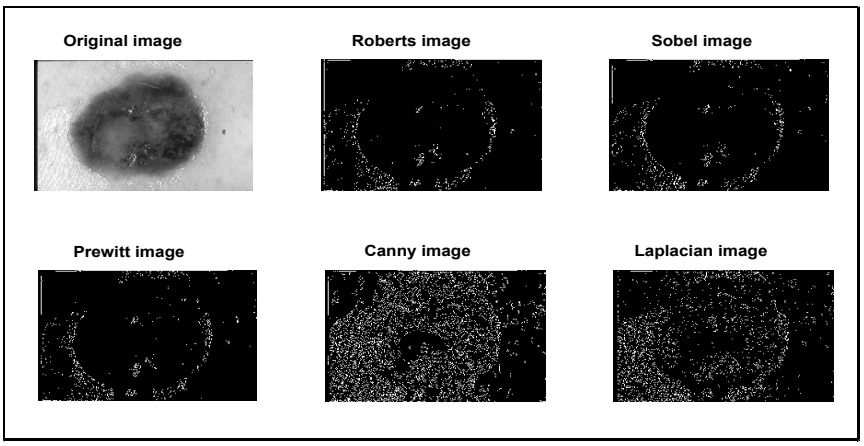


Fig. 31: Various edge detection methods for edge detection of a skin lesion, sample 2

D. Feature Extraction and Analysis

The features investigated in this study include the ABCD (asymmetry, border, colour, and diameter) rule for melanoma detection. These features are extracted and tabulated for each individual dermoscopic image.

1. Asymmetry

The geometric centre of an object is referred to as the centroid, c, of the object and it is the arithmetic mean position of all points of the object [114]. The coordinates of the centroid (c_x, c_y) are given by Eqs. (16) and (17):

$$c_x = \frac{1}{6A} \sum_{i=0}^{N-1} (x_i + x_{i+1})(x_i y_{i+1} - x_{i+1} y_i) \tag{14}$$

and

$$c_y = \frac{1}{6A} \sum_{i=0}^{N-1} (y_i + y_{i+1})(x_i y_{i+1} - x_{i+1} y_i), \tag{15}$$

where N is the number of vertices, as shown in Fig. 29, and A is the area represented by Eq. (18)

$$A = \frac{1}{2} \sum_{i=0}^{M-1} (x_i y_{i+1} - x_{i+1} y_i), \tag{16}$$

where M is the number of pixels in the object.

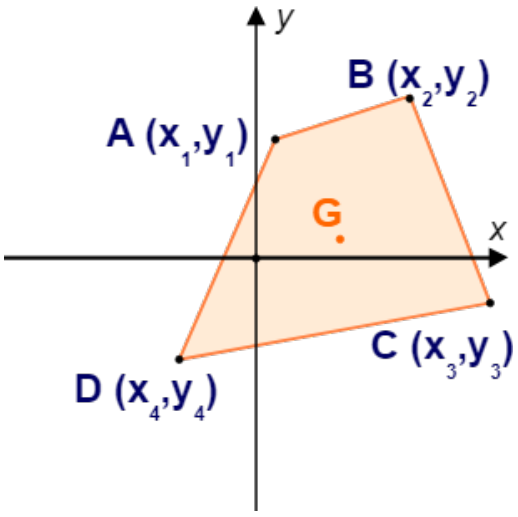


Fig. 32: Example of polygon for calculating the centroid [115]

Asymmetry is mathematically calculated using Eq. (19):

$$Asymmetry\ Index = \frac{A_x + A_y}{Total\ Area}, \tag{17}$$

where A_x represents the difference in area between the overlapping folds along the *x axis* and similarly A_y represents the difference in area between the overlapping folds in the *y axis* passing the centroid of the polygon. Bi-axial asymmetry produces a clearer representation of the degree of asymmetry that single-axis asymmetry.

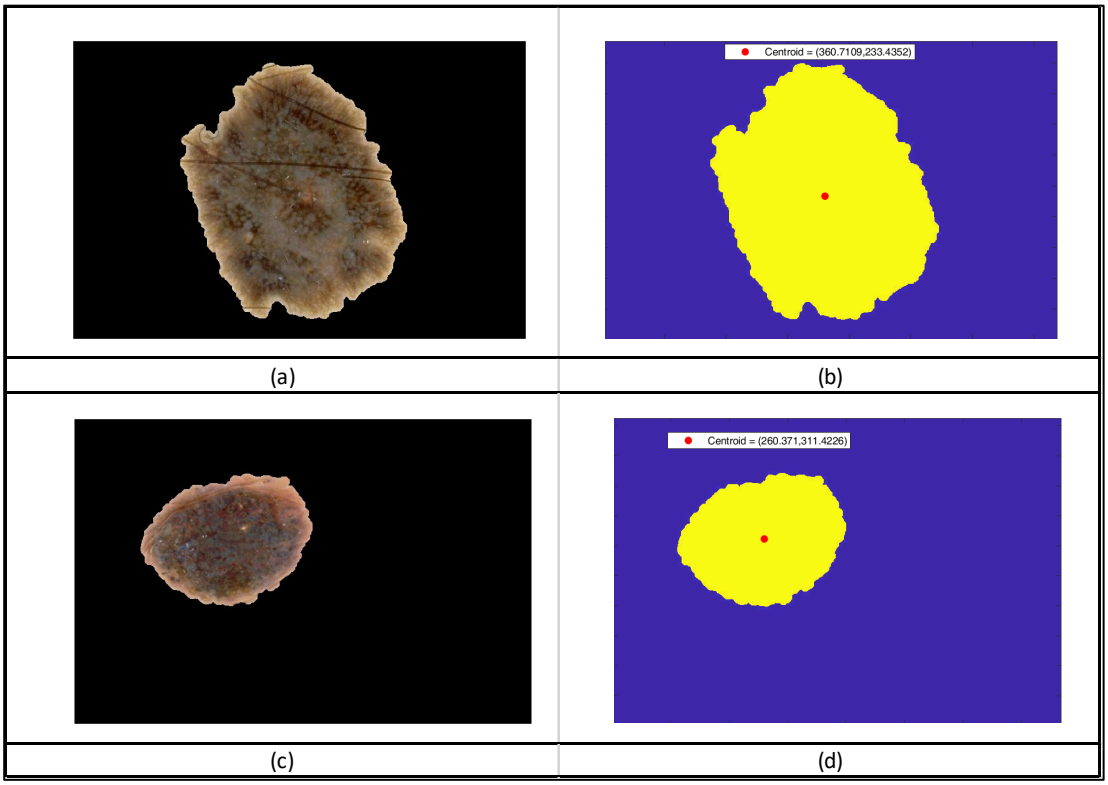


Fig. 33: Plotting of centroid for various segmented lesions

The geometric centre or centroid of the lesion is determined by applying Eqs. (16) to (17) to the segmented image (refer to Fig. 32); this is illustrated in Fig. 33. If the image is folded through the centroid and it overlaps completely with the other side, it is symmetrical. Conversely, if the segmented image is folded and does not overlap with the other side completely it may be asymmetrical [116]. The asymmetry index

in Eqs. (18) and (19) is used to quantify the degree of asymmetry of the segmented lesion. The *x* and *y* values of the centroid provide the respective axes for overlapping, allowing the asymmetry index calculation.

2. Border Irregularity

The outermost pixels of the segmented image represent the border. The roundness of an object is determined by calculating the circularity index, also called isoperimetric quotient; this is expressed by Eq. (20):

$$Circularity\ Index = \frac{4A\pi}{p^2} \quad . \tag{18}$$

The isoperimetric quotient measures the relationship between the area and perimeter where an index closer to zero indicates the lesion has a highly irregular border and if the index is closer to 1 the border is closer to a regular circle. Fig. 34 provides an illustration of the circularity index being calculated for various shapes.

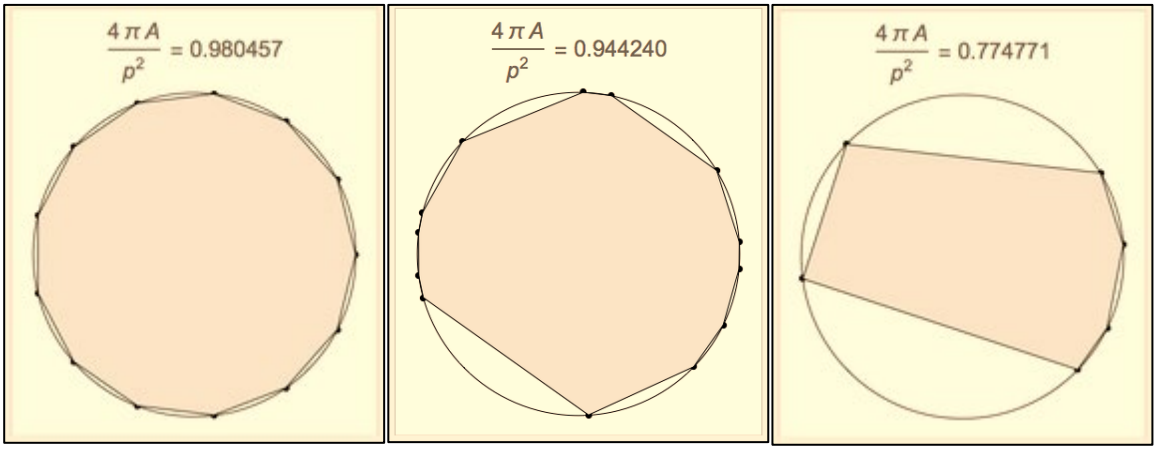


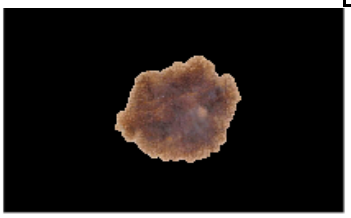
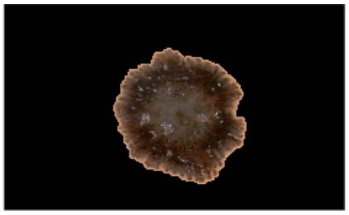
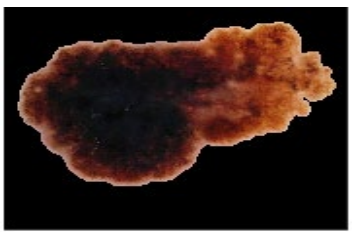
Fig. 34: Circularity Index calculation of various shapes [117]

Undulations in the perimeter creates structure irregularities along the border of the lesion. Melanoma is characterised by highly irregular borders in comparison to the non-melanoma lesions. The amount the border deviates from a circular shape needs to be quantified. Border irregularity was computed by determining the circularity index (isoperimetric quotient) of the lesion using Eq. (20). The circularity index is a widely accepted measure of compactness. The circularity index is a suitable measure as it is

dimensionless, and it does not vary as the image size changes. Furthermore, the circularity index is fast and easy to compute.

Table IV highlights the lower circularity index values for more irregular lesions and conversely, the higher the circularity index, the closer the lesions border resembles a circle.

TABLE IV:
SAMPLES OF CIRCULARITY CALCULATION RESULTS FOR MELANOMA AND NON-MELANOMA LESIONS

Melanoma	Circularity Index	Non- Melanoma	Circularity Index
	0.586		0.76
	0.692		0.809
	0.668		0.715

3. Colour

The eye perceives different colours by the various wavelengths of visible light. The colour of an object is determined by its reflective properties. An RGB image is constituted of three 8-bit layers viz. red, green, and blue. Fig. 36 shows an original RGB image of a lesion split into cyan (blue + green), magenta (red + blue), and yellow (red + green) bands, through setting of the red, green, and blue layer values, respectively, to zero.

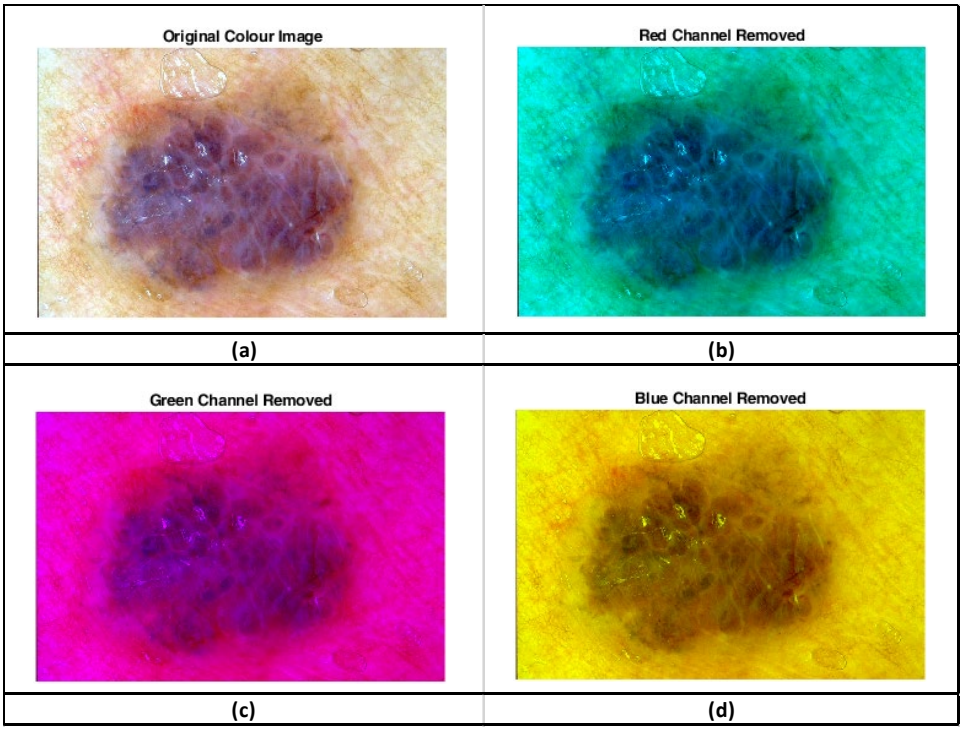


Fig. 35: Subtractive colour models of dermoscopic lesions: a) Original RGB image, b) Cyan, c) Magenta, and d) Yellow

The individual RGB colours may be extracted and interpreted to determine the ratios of RGB of interest for a particular segmented image. The ratio of RGB mix provides the various colours. There are six colours of interest when diagnosing melanoma [118]. Table V describes the six colours of interest and its origin.

TABLE IV:
THE 6 COMMON COLOURS OF INTEREST AND ITS LOCATION

Lesion colour	Cause
Red	Due to the level of inflammation
Dark brown	Melanin in epidermis/superficial dermis
Light brown	Melanin in epidermis/superficial dermis
Blue-grey	Melanin in the papillary dermis
Black	Melanin in the stratum corneum or granular layer of the dermis
White	Areas of regression

The segmented lesion is analysed by counting the number of pixels for each RGB channel. Melanoma usually has three or more of the colours of interest.

TABLE VI:
PIXEL COUNT FOR THE SIX COLOURS OF INTEREST

		RGB pixel count		
		Red (%)	**Green (%)**	**Blue (%)**
Colours of interest	**Red**	> 80	< 20	< 20
	Dark brown	20 - 60	5 - 45	0 - 35
	Light brown	60 - 95	30 - 70	5 - 45
	Blue-grey	< 20	30 - 70	35 - 75
	Black	< 20	< 20	< 20
	White	> 80	> 80	> 80

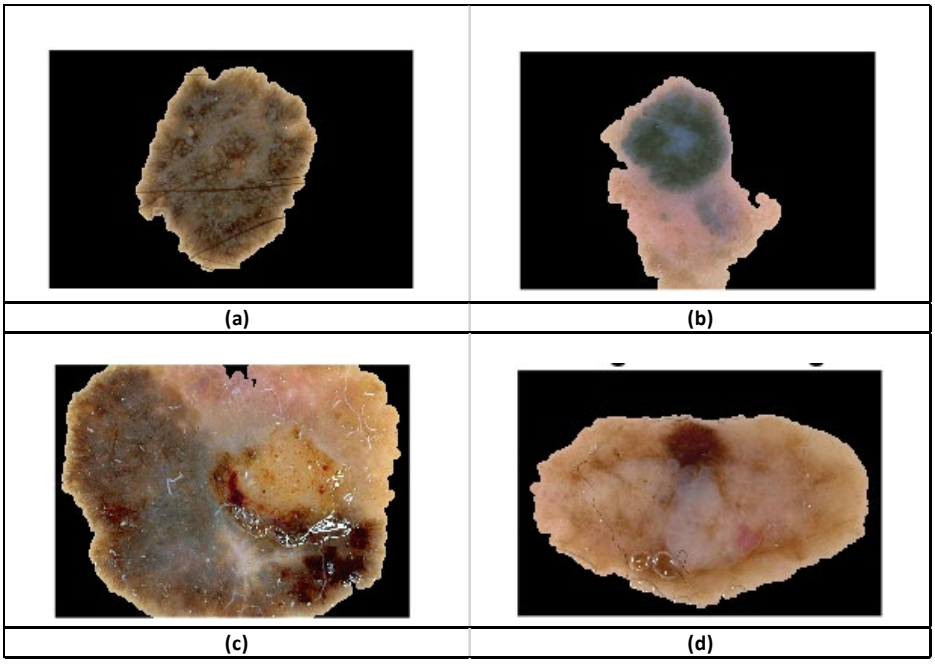


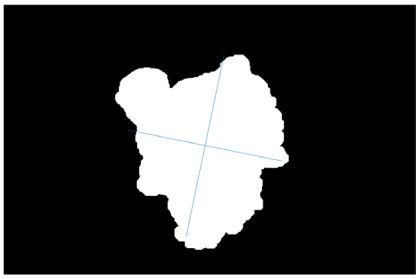
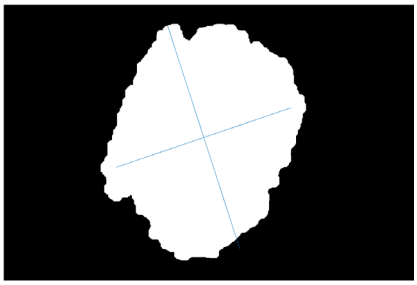
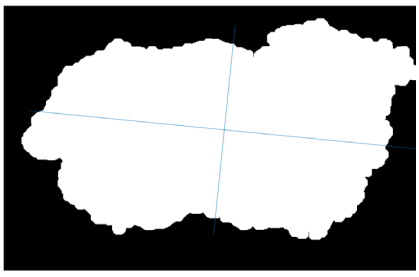
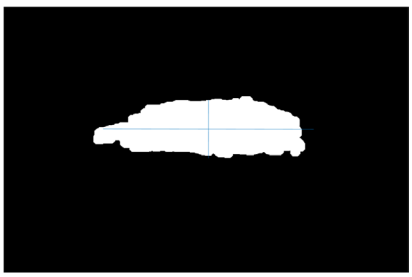
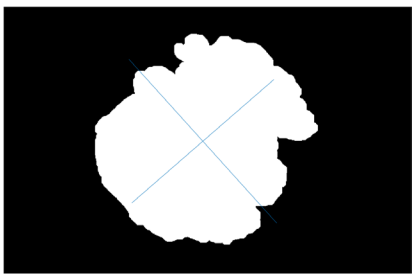
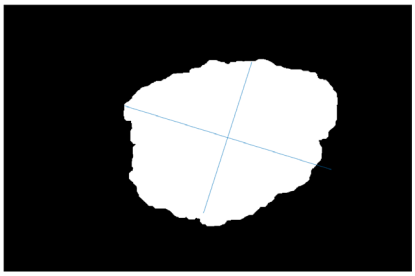
Fig. 36: Six colours of interest count for melanoma lesions: (a) Lesion sample 1 with colour count 2, (b) Lesion sample 2 with colour count 2, (c) Lesion sample 3 with colour count 3, and (d) Lesion sample 4 with colour count 3.

There is no standardisation of the colours of interest. Based on the research conducted by Majumder and Ullah [119] and together with experimental analysis of the training dataset, Table V was created to provide a representation of the six colours of interest. The values chosen provided the most accurate classification over the entire dataset. A resolution of 5% was used to ensure the pixel count was not overfitting the training data.

Each segmented image is analysed based on Table VI and the number of colours that meet the requirements for the colours of interest is added to provide a score from 0 to 6. Fig. 37 provides an illustration of the colour count for melanoma lesions. Figs. 37(a) and (b) have a colour count of 2 whereas Figs. 37(c) and (d) have a colour count of 2.

4. Diameter

TABLE VI:
DIAMETER CALCULATION FOR VARIOUS LESIONS

Melanoma	Diameter in mm	Non- Melanoma	Diameter in mm
	11.46		13.15
	19.5		9.04
	13.71		12.59

After the image has been segmented the diameter of the object within the image can be calculated without making physical measurements. The diameter of a circle is calculated by finding the centre of the circle and then measuring the distance from the edge of the circle, through the centre and to the opposite edge and is twice the radius.

The literature previously stated that melanoma lesions are usually greater than 6mm in diameter. The measurement of the diameter is found by the two pixels in the lesion that are the furthest away from each other and passing the centroid, namely the major axis [120]. The major axis is calculated using Eq. (21), where n is the number of pixels within the ROI and (x_i, y_i) is the coordinates of the ith pixel.

$$(x_c , y_c) = \left(\frac{\sum_{i=1}^{n} x_i}{n} , \frac{\sum_{i=1}^{n} y_i}{n}\right) \tag{19}$$

The resolution of the images in the dataset is 72dpi (or 2.83 pixels per mm) with a magnification of 10 [121]. Since the major axis was calculated in pixels, the diameter is given in pixels. The resolution and magnification are known variables; subsequently, the diameter, in millimetres, is calculated using Eq. (22):

$$d = \frac{major\ axis\ length}{resolution * magnification} \quad . \tag{20}$$

The major and minor axes are plotted on the segmented image as shown in Table VI. The major axis passes through the centroid of the lesion and the minor axis is perpendicular to the major axis.

E. Classification of Analysed Images

There is no generic machine learning algorithm for the multitude of types of source data. Consequently, several classifiers are tested to empirically validate the given data [122].

The four extracted features, viz. asymmetry, border, colour, and diameter, are used to generate the training and testing datasets. The training and test datasets comprise 178 and 40 images, respectively, of melanoma and non-melanoma lesions. The extracted features are tabulated to form a 178 x 5 training dataset as shown in appendix 7.1. The training datasets in turn are used to create the KNN, Naïve Bayes, and LSVM trained models. These models have been chosen as they are not data hungry nor computationally expensive

in comparison to deep learning methods. The classifiers are independently applied and the results are compared. In recent studies[123-127], KNN, Naïve Bayes and LSVM have been used to classify images of breast cancer (mammograms), plant leaf disease, brain lesions (MRI scans), food quality, Alzheimer's disease (MRI scan) and detection of Covid 19 in CT scan, showing their relevance in image classification.

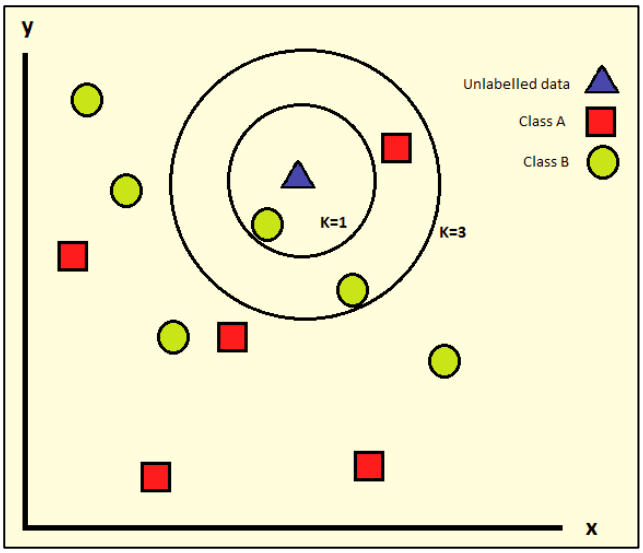


Fig. 37: KNN with two classes A and B with k=1 and k=3

The classification is determined by the majority of samples in the *k*-neighbours. KNN relies on the assumption that similar classes exist close together. The hyper-parameters, specifically the value of *k* and the distance metric, need to be selected to tune the algorithm.

Fig. 38 shows how value of *k* determines the number of neighbours used to classify an unknown data point. It is usually chosen to be an odd integer to avoid ties between two classes. The test sample is assigned to the class with the most data points for a given value of *k*. The Euclidean distance metric, which is based on the Pythagorean theorem, between two vectors x_s and y_t is given by the Eq. (23) below [128, 129]:

$$d_{st}^2 = (x_s - y_t)(x_s - y_t)' \tag{21}$$

The Euclidean distance metric is used to determine the training data points that are closest to the test sample in the neighbourhood, N_0. The probability of the classifier choosing a particular class, *j,* is based on Eq. (24):

$$Pr(Y = j \mid X = x_0) = \frac{1}{k} \sum_{i \in N_0} I(y_i = j) \quad , \qquad (22)$$

where $I(y_i = j)$ evaluates to 1 if the test sample, x_0, is of class *j*, and evaluates to 0 if not. X is the extracted features given as a matrix and Y is the class label. The value of *k* is determined by the data. If *k* is chosen to be too small, it may lead to overfitting and the predictions may be less accurate, conversely if *k* is too large under fitting of the training data occurs resulting in high error rates.

The KNN classifier used an initial value of k=10 and the Euclidean distance metrics to calculate the distance between the test and training sample. The distances are ranked from nearest to furthest and the top 10 are used. Tied values, equal to the 10th value, are also included. The value of *k* was incremented from 1 to 15 and a value of 10 provided the lowest false negative rate of the test samples. The test samples are classified based on the majority class of the 10 nearest neighbours. If a tie occurs between the 10 nearest neighbours, the class with the nearest neighbour is used.

Another common classifier is the Naïve Bayes; this is a probabilistic classifier that works well with a high number of input features [89]. Naïve Bayes assumes each feature is independent and equally contributes to the classification. Bayes theorem determines the probability of an event occurring given the probability of a prior event using the following equations:

$$Posterior = \frac{Likelihood\ x\ Prior}{Evidence} \qquad \text{and} \qquad (23)$$

$$P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)}, \qquad (24)$$

where $P(X|Y)$ is the probability of X given Y, $P(Y|X)$ is the probability of Y given X, $P(X)$ is the probability of X, and $P(Y)$ is the probability of Y. A Naïve Bayes classifier, which is centred on Bayes Theorem, naively assumes X and Y are mutually independent [130]. The probability of each class is calculated, and classification is based on the highest probability of *y* given the input features *x* for *n* number of *x* using the formula

$$y = argmax_y[P(y) * \prod_{i=1}^{n} P(x_i|y)] \qquad (25)$$

Naïve Bayes was implemented using a Gaussian (normal) distribution. The mean μ and standard deviation σ of the training data is calculated by estimating the normal distribution for each class shown by Eqs. (28) and (29):

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{26}$$

$$\sigma = \left[\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2\right]^{0.5} \tag{27}$$

The test data classification is predicted by calculating the posterior probability of a particular test sample belonging to each class. The test sample is then allocated to a class based on the highest posterior probability.

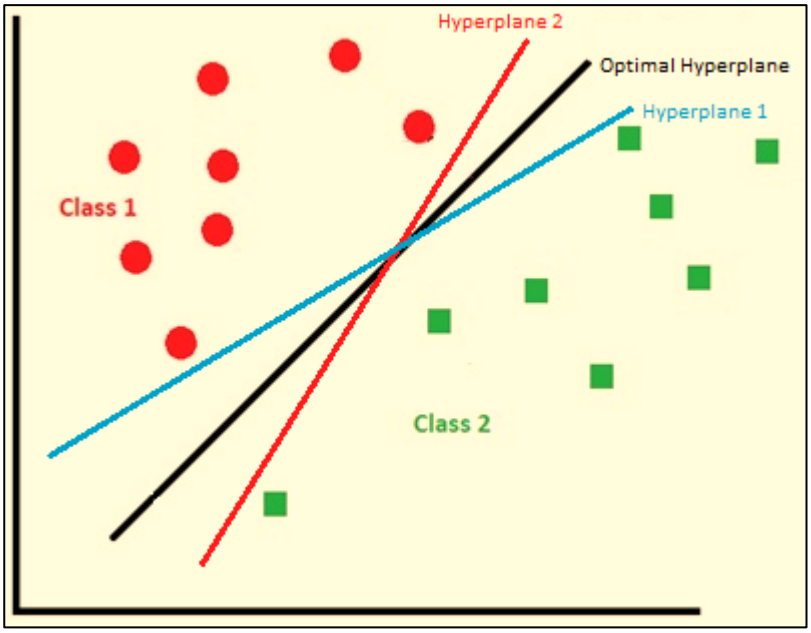


Fig. 38: Various hyperplanes that may separate the data into the 2 classes

A third familiar classifier is the support vector machine (SVM); this is a common choice for image classification in the biomedical field owing to its ease of use and flexibility [90]. SVM, which was originally developed as a binary classifier, separates data into classes using linear or non-linear classification depending on the type of data provided. Fig. 39 shows the lines or boundaries that may be used to separate the data into classes and is referred to as the hyperplane [93, 94]. All three hyperplanes separate the data but the margins

between the data and hyperplanes 1 and 2 are not maximised and could lead to an increase in misclassifications.

A support vector is the data point closest to the hyperplane. The optimal hyperplane maximises a margin between the support vectors; this is illustrated in Fig. 40.

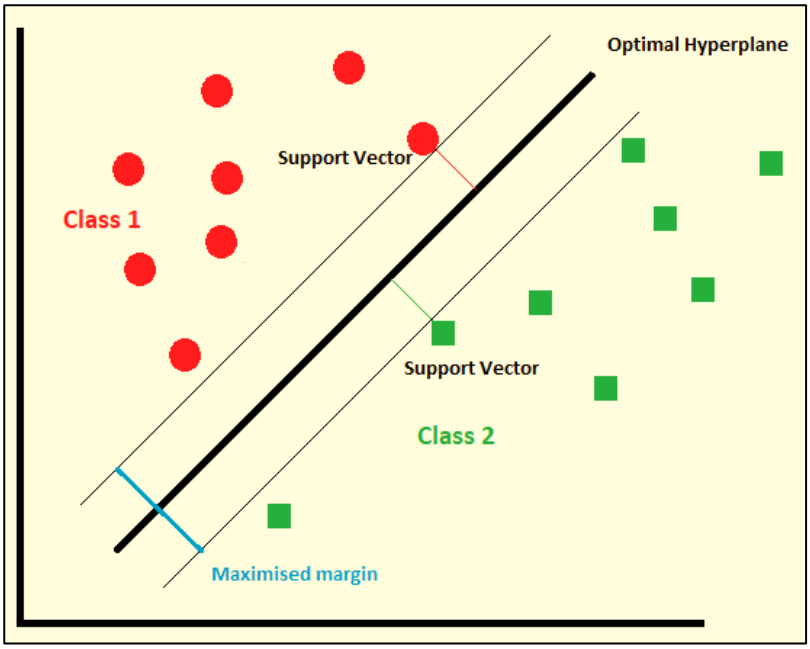


Fig. 39: Linear SVM classification of linearly separable data with maximised margins

Each extracted feature, as shown in Eq. (30) has a weight w^i. Prediction is based on the weighted sum.

$$f(x) = \sum_i w^i \; x^i = \vec{w} \bullet \vec{x} \tag{28}$$

The Class 1 support vector in Fig. 40 is given by the equation

$$\vec{w}.\vec{x} + b = -1 \tag{29}$$

and the Class 2 support vector is given by the equation

$$\vec{w}.\vec{x} + b = +1 \tag{30}$$

The equation for the hyperplane is

$$\vec{w}.\vec{x} + b = 0 \tag{31}$$

where, $\overrightarrow{w}$ is the weight vector, $\vec{x}$ is the input feature vector, and b is the bias.

Obtaining the largest possible margin will provide a better classification of the data and therefore reduce classification errors. The optimal hyperplane is found by maximising the margin (m); this is determined by the perpendicular distance between the origin and the hyperplane.

Using the distance formula, given by Eq. (34), to calculate the length of the perpendicular from the point (x_1, y_1) to the line $(ax_1 + by_1 + c = 0)$

$$Distance = \frac{|ax_1+by_1+c|}{\sqrt{a^2+b^2}} \quad . \tag{32}$$

The margin, m, is given by:

$$m = \frac{|\vec{w}\,.\,\vec{x}-b|}{\|\vec{w}\|_2} = \frac{1}{\|\vec{w}\|_2} \quad , \tag{33}$$

Where, $\vec{w}$ is a support vector and $\vec{x}$ is an input vector. The maximised margin is given by:

$$m = \max_{\vec{w}} \frac{1}{\|\vec{w}\|_2} = \min \frac{1}{2} \|\vec{w}\|^2 \tag{34}$$

The data is not perfectly separable, therefore LSVM uses a soft margin by introducing slack variables, ξ_j, with the box constraint or penalty factor C set to 1 to allow for misclassifications and maximise the margin to avoid overfitting as shown in Eq. (37). The kernel scale is set to a value of 2 to reduce bias towards features with higher magnitudes.

$$m = \min \frac{1}{2} \|\vec{w}\|^2 + C \sum_j \xi_j \tag{35}$$

The parallel coordinate plots of the extracted features are shown in Fig. 41. The parallel plot allows visualization of the four extracted features on a single two-dimensional plot. The incorrectly classified training data is represented with the dashed line. The parallel plot helps with selecting features that do not help in separating the classes. The concentration groupings of the orange and blue lines on each of the four parallel plots shows that all features contribute to the classification of the data.

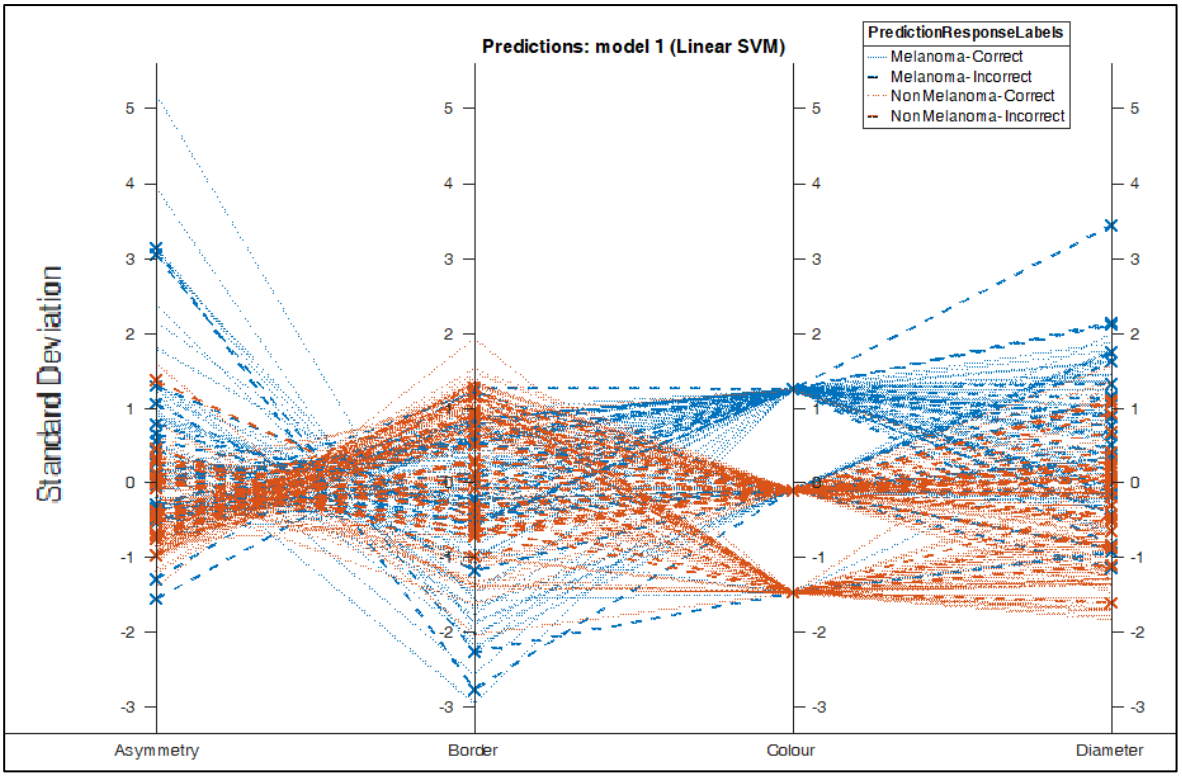


Fig. 40: Parallel coordinate plot for linear SVM

Finally, the test dataset is inputted into the trained models for classification. For this study, five-fold cross-validation was employed for the generation of the independent datasets. The testing set, which contains 40 images, is balanced, being divided evenly for melanoma and non-melanoma cases.

IV. RESULTS AND DISCUSSION

This chapter presents the results obtained from the proposed method described in chapter 4. In addition, this section discusses the results and provides a comparative analysis against previously proposed methods. The classification algorithms are measured by calculating the accuracy, sensitivity, specificity, and the false negative rate. The focus of this study is on minimising the false negative rate to ensure patients with melanoma are correctly diagnosed to alleviate delays in treatment.

A. Results of classification

The four extracted features, viz. asymmetry, border, colour, and diameter are used to generate the training and testing datasets. The training datasets in turn are used to create the KNN, Naïve Bayes, and LSVM models. Finally, the test datasets are inputted into the trained models for classification.

The evaluation of the classification algorithms is measured by calculating the accuracy, sensitivity, specificity and importantly the false negative rate; these are mathematically expressed by Eqs. (38)-(41), respectively [131]. Accuracy refers to the number of *correct* predictions over the *total* predictions made; this is expressed by Eq. (38):

$$Accuracy = \frac{t_p + t_n}{t_p + f_p + t_n + f_n}, \tag{36}$$

where t_p denotes a true positive, t_n refers to a true negative, f_p denotes a false positive, and f_n refers to a false negative. Sensitivity/Recall (true positive rate) refers to the proportion of actual *positives* that are correctly classified; this is given by Eq. (39):

$$Sensitivity = \frac{t_p}{t_p + f_n} \tag{37}$$

Specificity (true negative rate) is the proportion of actual *negative* samples that are correctly classified; this is expressed by Eq. (40):

$$Specificity = \frac{t_n}{t_n + f_p}. \tag{38}$$

False negative rate (FNR) is a type II error where positive samples have been incorrectly classified as negative; this is given by Eq. (41):

$$FNR = \frac{f_n}{t_p + f_n} = 1 - Sensitivity \quad . \quad (39)$$

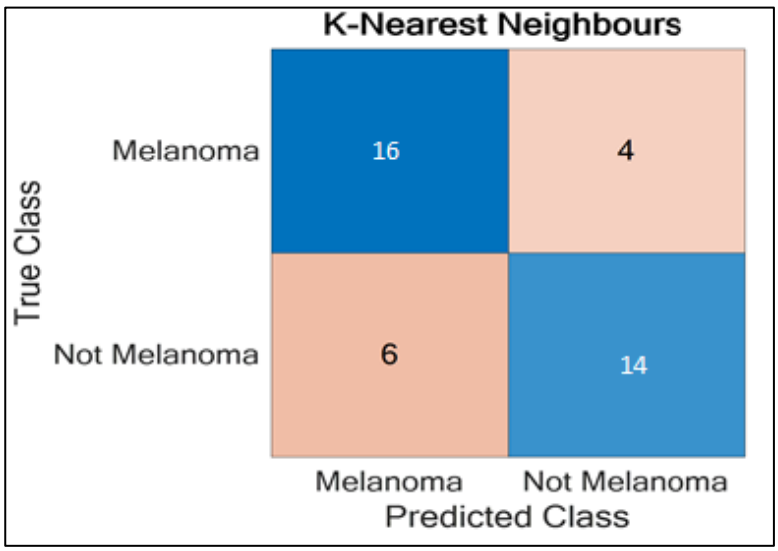


Fig. 41: Confusion Matrix for *k*-Nearest Neighbours for 40 test samples

The decision matrix for the KNN model is shown in Fig. 42. In this case, the model correctly classified 30 out of 40 images providing an accuracy of 75%, sensitivity of 80%, specificity of 70% and a false negative rate of 20%.

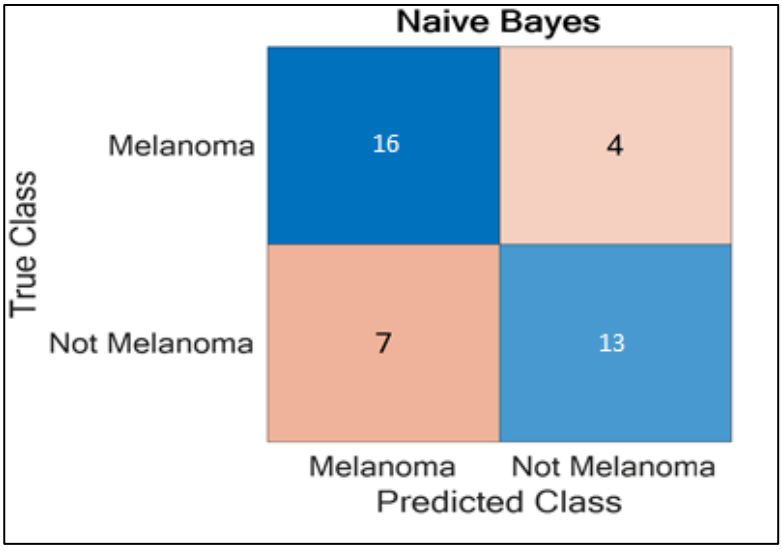


Fig. 42: Confusion Matrix for Naïve Bayes for 40 test samples

The decision matrix for the Naïve Bayes model is shown in Fig. 43. In this case, the model correctly classified 29 out of 40 images providing an accuracy of 72.5%, sensitivity of 80%, specificity of 65% and a false negative rate of 20%.

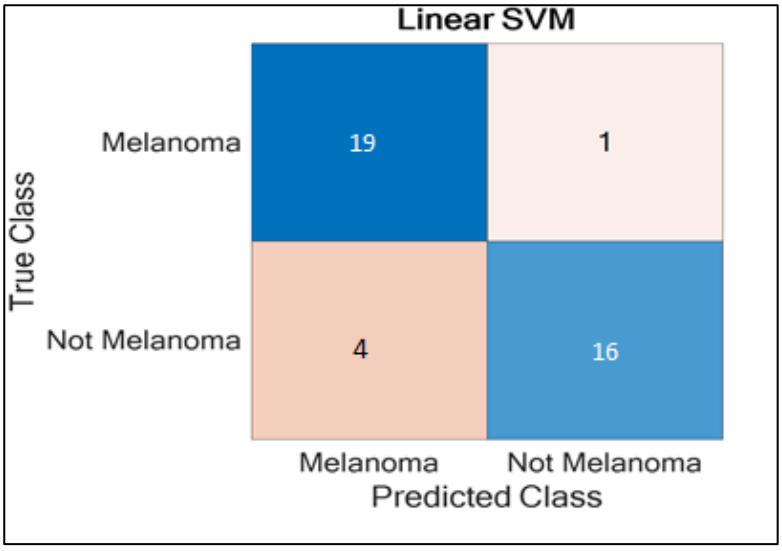


Fig. 43: Confusion Matrix for Linear SVM for 40 test samples

The decision matrix for the linear SVM model is shown in Fig. 44. In this case, the model correctly classified 35 out of 40 images providing an accuracy of 87.5%, sensitivity of 95%, specificity of 80% and a false negative rate of 5%.

TABLE VII:
STATISTICAL MEASUREMENT OF RESULTS

	Naïve Bayes	**KNN**	**Linear SVM**
Sensitivity	80%	80%	**95%**
Specificity	65%	70%	**80%**
Accuracy	72,5%	75%	**87.5%**
False Positive Rate	35%	30%	**20%**
False Negative Rate	20%	20%	**5%**

Table VII provides a summary of the measured results. The true negative rate test i.e., the specificity of SVM at 80% was notably greater than Naïve Bayes at 65% and KNN at 70%. A high specificity value will produce a low false-positive rate and is an indicator of the reliability of the melanoma classification result [122].

Similarly, the true positive rate test i.e., sensitivity value of linear SVM at 95%, clearly outperforms Naïve Bayes and KNN, both at 80%. Sensitivity is an indicator of how well the algorithms can correctly detect melanoma. A high sensitivity shows that the algorithm has a lower chance of providing a false negative result. High sensitivity is ideal for positively diagnosing a condition, namely melanoma. Therefore, a non-melanoma classification result would be reliable.

Linear SVM produced the lowest false negative rate at 5% compared to KNN and Naïve Bayes both at 20%. A low false negative rate shows that linear SVM missed the least number of positive melanoma cases when compared to Naïve Bayes and KNN. It is preferable to minimise false negatives in clinical diagnosis so treatment of the melanoma can begin as soon as possible limiting the risk of the cancer progressing to further stages.

Linear SVM handles outliers and unknown distributed data better than KNN and Naïve Bayes. This is done by using a soft margin to allow for misclassifications with a penalty. Naïve Bayes will fit the Gaussian distribution to both classes (Melanoma and non-melanoma) and thereafter perform the classification. KNN is biased to the data closest to the test sample [132].

SVM performs better with higher features and fewer training samples. Naïve Bayes requires a large dataset to perform well at classification. Furthermore, linear SVM is less prone to over-fitting data when compared to KNN and Naïve Bayes.

B. Analysis of SVM results

The trained SVM model had incorrectly classified 5 images; 1 was a false negative (this is shown in Fig. 45) and 4 were a false positive (these are shown in Fig. 46).

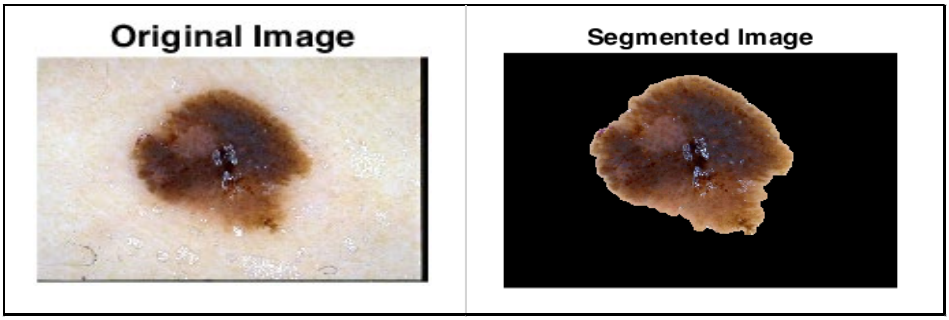


Fig. 44: Incorrect classification of a lesion (False Negative)

The lesion in Fig. 45 is well segmented. The high border index of 0.7592 indicates the image is closer to a regular shape and has 2 of the 6 colours of interest present. These factors could contribute to the lesion being incorrectly classified. Minimising false negatives is very important as a false negative misdiagnosis can be catastrophic owing to the possible late detection of melanoma and subsequent delay in the treatment.

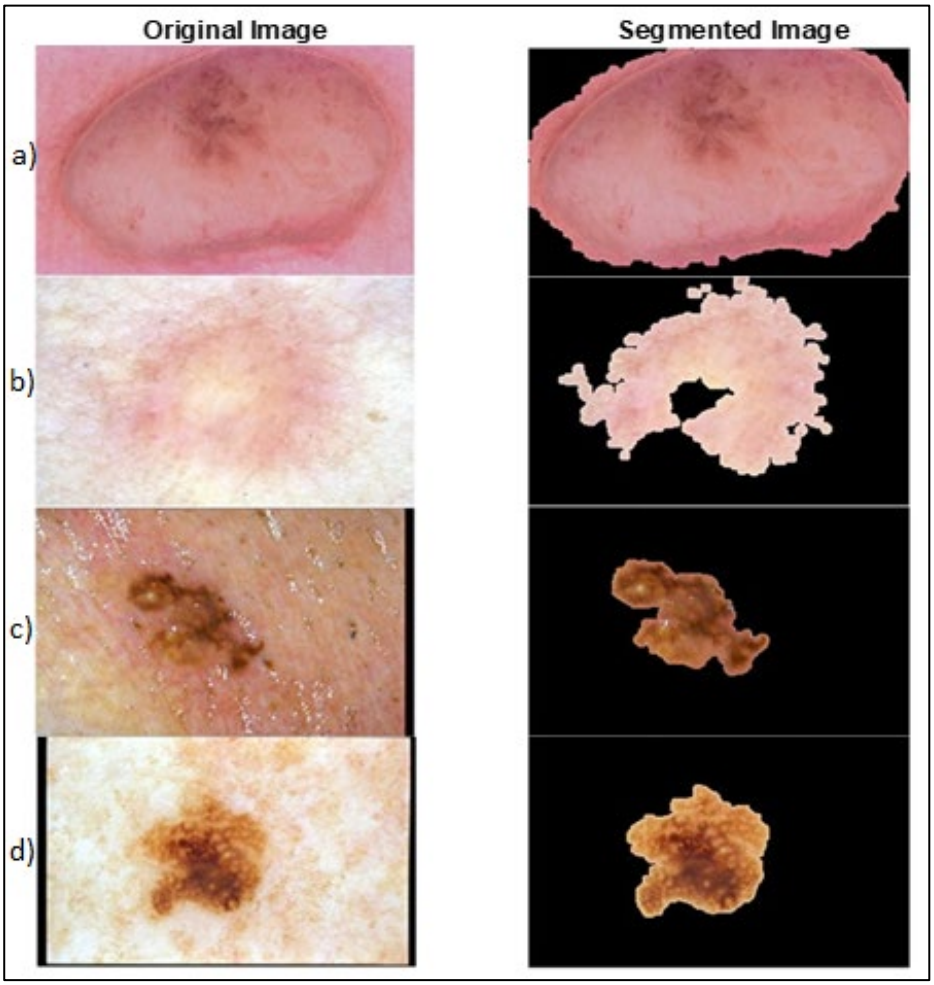


Fig. 45: Incorrectly classified lesions (false positive)

Fig. 46(a-d) shows the lesions that were incorrectly classified as melanoma lesions, producing 4 false positive results. Some images had very weak boundaries and this produced poor automated segmentation results thereby making the lesions appear more irregular than they were.

Fig. 46(a) shows a lesion that does not have a high contrast between the skin and lesion. This increases the difficulty to accurately detect the border and segment the lesion, producing an outlier for asymmetry calculation. Also, the lesion has a large diameter, which is typical of melanoma. Fig. 46(b) also has poor contrast between the lesion and the background skin, producing a very poor segmentation result. This incorrectly created a segmented lesion that is asymmetrical with highly irregular borders, which is typical of melanoma. Figs. 46(c) and (d) show lesions that are well segmented, asymmetrical and have irregular borders, which characterises a typical melanoma lesion.

The extracted ABCD features for each incorrectly classified lesion is shown in Table VIII. A false positive diagnosis would lead to unnecessary further testing and procedures. This could be costly, painful and time consuming. False positive also creates undue anxiety with the patients and family.

TABLE VIII:
LIST OF THE INCORRECT CLASSIFICATION OF LESIONS

Figure	Asymmetry	Border	Colour	Diameter	Clinical Classification	Predicted Classifier
Fig. 45	0.1338	0.7592	2	8.617	Melanoma	'Non-Melanoma'
Fig. 46(a)	0.0763	0.7308	2	16.3116	Non-Melanoma	'Melanoma '
Fig. 46(b)	0.472	0.1604	1	10.7404	Non-Melanoma	'Melanoma '
Fig. 46(c)	0.2919	0.5633	2	6.5735	Non-Melanoma	'Melanoma '
Fig. 46(d)	0.2712	0.6053	2	6.9333	Non-Melanoma	'Melanoma '

C. Comparison with similar studies

There have been many studies investigating the use of machine learning to classify dermoscopic images in the past 20 years. In 2016, the International Skin Imaging Collaboration (ISIC) created a challenge, using a dataset from multiple sources, to improve melanoma diagnosis. This has led to a significant increase in research in melanoma detection.

The classification results of proposed method indicate that the study is comparable to other research undertaken using the same EDRA database as shown in Table IX. The researchers do not state if the chosen datasets are exclusively of lesions of less than 0,76mm thick or lesions in situ nor the stage of cancer.

TABLE IX:
RESULTS OBTAINED FROM OTHER STUDIES USING THE INTERACTIVE ATLAS OF DERMOSCOPY (EDRA) DATABASE [133-136].

				Results %			
Researchers	**Year**	**Training/Test size**	**Classifiers used**	**Accuracy**	**Specificity**	**Sensitivity**	**FNR**
1[136]	2016	120/80	TDS	**94**	**96**	91	9
2[135]	2020	Pretrained InceptionV3 / 1011	Deep Neural Network	79			
3[134]	2016	29 / 340	Bag of Features	75	70	81	19
4[133]	2019	176/176	CNN	76	70	86	14
Proposed Method	2022	178/40	SVM	87,5	80	**95**	**5**

Kasmi and Mokrani [136] used a dataset of 200 images. This dataset was split into a training dataset of 120 images and a test dataset of 80 images. They employed the ABCD rule for feature extraction and a total dermoscopic score (TDS) for the classification. Their research produced an accuracy of 94%, specificity of 96% and a sensitivity of 91% giving a false negative rate of 9%. TDS is a simple to implement multi-criteria decision analysis tool that add up the weighted values of the extracted features. The sum of the features is referenced to a predetermined scale to classify as benign, suspicious, or malignant.

Tschandl et al. [133] trained a convoluted neural network(CNN), using Resnet50, with 888 images and tested the system with 176 images. The classification model produced an accuracy of 76%, specificity of 79% and sensitivity of 86% giving a false negative rate of 14 %. CNN does not require significant pre-processing of the images but requires a large amount of training data to be able to produce accurate results.

Barata et al. [134] only extracted the enhanced colour features using the bag of features method and *k*-means clustering. SVM was utilised to classify the images producing an accuracy of 75 %, specificity of 70% and sensitivity of 81%, giving a false negative rate of 19%. The enhanced colour features improve the colour constancy of images obtained from multiple sources.

V. CONCLUSION

This research presents a method for the automatic feature extraction and analysis of dermoscopic images of skin lesions towards early melanoma detection through the minimisation of false negative results.

A. Summary

This research has been conducted to show how modern computer imaging technology can be used as a support tool to assist medical practitioners in the early detection of melanoma. The patient survival rate is linked to the early detection of melanoma; therefore, the minimisation of false negative rate is significant and important. A low false negative rate will ensure the patient minimises delays in the necessary treatment. The dermoscopic images chosen are of melanoma lesions less than 0,76mm in thickness, which corresponds to the early stages of cancer.

The proposed methodology requires the acquisition, processing, segmentation, ABCD feature extraction and classification of skin lesions as either melanoma or non-melanoma and minimising the false negative rate using machine learning.

Three classifiers were chosen; these include *k*-nearest neighbours (KNN), Naive Bayes, and a linear support vector machine (LSVM). The performance of the selected classification algorithms is compared based on accuracy, sensitivity, specificity and false negative rate. All classifiers used in this study have been able to differentiate the images into their respective class with varying degrees of success.

The experimental results show that the Naïve Bayes algorithm produced an accuracy of 72.5%, specificity of 65% and sensitivity of 80%, and a false negative rate of 20%. The KNN algorithm produced an accuracy of 75%, specificity of 70% and sensitivity of 80% and obtained a false negative rate of 20%, which is similar to Naïve Bayes. The LSVM algorithm was the optimal classifier with an accuracy of 87.5%, specificity of 80% and sensitivity of 95%, and a false negative rate of 5%.

A high sensitivity was obtained to ensure the melanoma cases are classified correctly ensuring early treatment. Linear SVM produced the highest specificity of 80%. The classification of the lesions using classification algorithms may be considered as part of the decision methods employed by physicians. This has the potential to reduce unnecessary biopsies of benign lesions and speed up the diagnosis and treatment of melanoma. LSVM, which produced the lowest false negative rate of 5%, missed the least number of

positive cases. Currently the research presented is not intended to replace a health care physician but may be used as a decision support tool by providing a second opinion.

B. Recommendations

The uniform use of the dermoscope has proved to be a challenge as most images are not ideal. Training on proper capturing of the images along with meta data is needed to improve on the accuracy of the classification. Standard guidelines on the use of a dermoscope should be included along with the training of dermatologists. Thickness of lesion and the stage of cancer for confirmed melanoma should form part of the meta data. Darker skin types are not well represented in the datasets currently available.

C. Future work

Future work will include training and testing the models on significantly larger datasets. With improvement and standardization of the capturing of dermoscopic images the available images would increase, increasing the size and quality of the databases. Moreover, hybrid methods using the ABCD rule in conjunction with other rules such as Menzies and a seven-point checklist will increase the number of input features providing a more reliable classification model[137]. The ABCD model, by design, may not detect single colour or smooth shaped melanoma. Once a reliable and sufficiently large database of suitable dermoscopic images is created, deep learning techniques such as convolutional neural networks should be investigated for lesions less than 0,76mm thickness[138]. Furthermore, inclusion of metadata[139], such as age, location of lesion, thickness of lesion and evolution of lesion, together with the datasets, may be considered for the improvement in the metrics of the classifiers.

VI. REFERENCE LIST

[1] C. Y. Wright and M. Norval, "Health Risks Associated With Excessive Exposure to Solar Ultraviolet Radiation Among Outdoor Workers in South Africa: An Overview," (in English), *Frontiers in Public Health,* Review vol. 9, 2021-April-28 2021, doi: 10.3389/fpubh.2021.678680.

[2] D. Whitaker and W. Sinclair, *Guideline on the Management of Melanoma* (2008, no. 8). 2008.

[3] C. Conforti, R. Giuffrida, R. Vezzoni, F. S. S. Resende, N. di Meo, and I. Zalaudek, "Dermoscopy and the experienced clinicians," *Int J Dermatol,* vol. 59, no. 1, pp. 16-22, 2020, doi: https://doi.org/10.1111/ijd.14512.

[4] M. E. Celebi, N. Codella, A. Halpern, and D. G. Shen, "Guest Editorial Skin Lesion Image Analysis for Melanoma Detection," (in English), *Ieee J Biomed Health,* vol. 23, no. 2, pp. 479-480, Mar 2019, doi: 10.1109/Jbhi.2019.2897338.

[5] P. Carli *et al.*, "Pattern analysis, not simplified algorithms, is the most reliable method for teaching dermoscopy for melanoma diagnosis to residents in dermatology," (in eng), *Br J Dermatol,* vol. 148, no. 5, pp. 981-4, May 2003, doi: 10.1046/j.1365-2133.2003.05023.x.

[6] M. Boer, E. Duchnik, R. Maleszka, and M. Marchlewicz, "Structural and biophysical characteristics of human skin in maintaining proper epidermal barrier function," *Postepy Dermatol Alergol,* vol. 33, no. 1, pp. 1-5, Feb 2016, doi: 10.5114/pdia.2015.48037.

[7] H. Yousef, M. Alhajj, and S. Sharma, "Anatomy, Skin (Integument), Epidermis," in *StatPearls*. Treasure Island (FL), 2020.

[8] S. D. Lindsay M. Biga, Amy Harwell, Robin Hopkins, Joel Kaufmann, Mike LeMaster, Philip Matern, Katie Morrison-Graham, Devon Quick , Jon Runyeon, *ANATOMY & PHYSIOLOGY*, First ed. Corvallis, Oregon: OpenStax/Oregon State University, 20XX.

[9] M. Aida Maranduca *et al.*, "Skin - a vast organ with immunological function (Review)," (in eng), *Exp Ther Med,* vol. 20, no. 1, pp. 18-23, Jul 2020, doi: 10.3892/etm.2020.8619.

[10] S. Lawton, "Skin 1: the structure and functions of the skin," *Nursing Times,* vol. 115, no. 12, pp. 30-33, 2019.

[11] "Skin Lesions." American Society for Dermatologic Surgery. https://www.asds.net/skin-experts/skin-conditions/skin-lesions (accessed 21 January 2020.

[12] K. Holland. "What's Causing This Skin Lesion?" https://www.healthline.com/symptom/skin-lesion (accessed 28 January 2020.

[13] "Skin Lesion." Healthgrades Editorial https://www.healthgrades.com/right-care/skin-hair-and-nails/skin-lesion (accessed 21 January 2021.

[14] E. Ozgur, U. Kamiloglu, P. Temiz, and G. Eskiizmir, "Skin Cancers of the Auricle: A Retrospective Analysis of 41 Patients," (in English), *Turk Arch Otorhinol,* vol. 58, no. 3, pp. 169-173, Sep 2020, doi: 10.5152/tao.2020.5701.

[15] J. de Wet, M. Steyn, H. F. Jordaan, R. Smith, S. Claasens, and W. I. Visser, "An Analysis of Biopsies for Suspected Skin Cancer at a Tertiary Care Dermatology Clinic in the Western Cape Province of South Africa," *J Skin Cancer,* vol. 2020, p. 9061532, 2020/01/27 2020, doi: 10.1155/2020/9061532.

[16] WebMD. "Skin Cancer." eMedicineHealth. https://www.webmd.com/melanoma-skin-cancer/melanoma-guide/skin-cancer#1 (accessed 28 November 2020.

[17] "Cancer." World Health Organisation https://www.who.int/news-room/fact-sheets/detail/cancer (accessed 22 February 2020.

[18] "What is Skin Cancer." Centre for Disease Control. https://www.cdc.gov/cancer/skin/basic_info/what-is-skin-cancer.htm (accessed 12 December 2020.

[19] "Skin Cancer." Mayo Clinic. https://www.mayoclinic.org/diseases-conditions/skin-cancer/symptoms-causes/syc-20377605 (accessed 08 January 2021.

[20] "Basel cell carcinoma." Skin Cancer Foundation. https://www.skincancer.org/skin-cancer-information/basal-cell-carcinoma/ (accessed January 2021.

[21] E. Hale. "Squamous Cell Carcinoma." Skin Cancer Foundation. https://www.skincancer.org/skin-cancer-information/squamous-cell-carcinoma/ (accessed 15 August 2022.

[22] T. J. DiChiara. "The 5 Stages of Melanoma." https://www.verywellhealth.com/melanoma-staging-what-it-means-and-reveals-3010755 (accessed 20 August 2022.

[23] "Melanoma." Skin Vision. https://www.skinvision.com/melanoma/pictures/ (accessed January 2021.

[24] D. Ruiz, V. Berenguer, A. Soriano, and B. Sánchez, "A decision support system for the diagnosis of melanoma: A comparative approach," *Expert Syst Appl,* vol. 38, no. 12, pp. 15217-15223, 2011/11/01/ 2011, doi: https://doi.org/10.1016/j.eswa.2011.05.079.

[25] S. K. Kannan. Pigmented Lesions 2. Available: https://studylib.net/doc/15357901/pigmented-lesions-ii.

[26] "Melanoma: Diagnosis and treatment." Mayo Clinic. https://www.mayoclinic.org/diseases-conditions/melanoma/diagnosis-treatment/drc-20374888 (accessed 26 September 2020.

[27] A. Bruner and S. D. Schaffer, "Diagnosing Skin Lesions: Clinical Considerations for Primary Care Practitioners," *The Journal for Nurse Practitioners,* vol. 8, no. 8, pp. 600-604, 2012, doi: 10.1016/j.nurpra.2012.04.016.

[28] J. Benedetti. "Diagnostic Tests for Skin Disorders." https://www.msdmanuals.com/professional/dermatologic-disorders/approach-to-the-dermatologic-patient/diagnostic-tests-for-skin-disorders (accessed 28 October 2020.

[29] T. S. Tavares *et al.*, "Differential diagnoses of solitary and multiple pigmented lesions of the oral mucosa: Evaluation of 905 specimens submitted to histopathological examination," (in eng), *Head Neck,* vol. 43, no. 12, pp. 3775-3787, Dec 2021, doi: 10.1002/hed.26872.

[30] C. Guru. Dermatologist Inspecting Patient Skin Moles. Available: https://www.gettyimages.com/detail/photo/dermatologist-inspecting-patient-skin-moles-royalty-free-image/694181928.

[31] D. H. Thompson. "A comparison of polarised and nonpolarised dermoscopy." https://dermnetnz.org/topics/a-comparison-of-polarised-and-nonpolarised-dermoscopy/ (accessed 14 November 2020.

[32] R. Erol, "SKIN CANCER MALIGNANCY CLASSIFICATION WITH TRANSFER LEARNING," Master of Science, Department of Computer Science, Graduate School of University of Central Arkansas, 2018. [Online]. Available: https://uca.edu/computerscience/files/2020/02/Skin-Cancer-Malignancy-Classification-with-Transfer-Learning.pdf

[33] R. H. Johr, "Dermoscopy: alternative melanocytic algorithms-the ABCD rule of dermatoscopy, Menzies scoring method, and 7-point checklist," (in eng), *Clin Dermatol,* vol. 20, no. 3, pp. 240-7, May-Jun 2002, doi: 10.1016/s0738-081x(02)00236-5.

[34] A. C. Halpern. "Melanoma Warning Signs." https://www.skincancer.org/skin-cancer-information/melanoma/melanoma-warning-signs-and-images/ (accessed 4 February 2021.

[35] R. J. Friedman, D. S. Rigel, and A. W. Kopf, "Early detection of malignant melanoma: The role of physician examination and self-examination of the skin," *CA: A Cancer Journal for Clinicians,* vol. 35, no. 3, pp. 130-151, 1985, doi: https://doi.org/10.3322/canjclin.35.3.130.

[36] "Medical imaging & nuclear medicine," in "WHO list of priority medical devices for cancer management," World Health Organization, 2017. Accessed: 2021/02/18/. [Online]. Available: http://www.jstor.org/stable/resrep27923.12

[37] R. Lowe and D. J. Auson. "Medical Imaging." https://www.physio-pedia.com/Medical_Imaging (accessed 6 February 2021.

[38] T. Mendonça *et al.*, "PH2: A Public Database for the Analysis of Dermoscopic Images," 2015, pp. 419-439.

[39] J. Kawahara, S. Daneshvar, G. Argenziano, and G. Hamarneh, "Seven-Point Checklist and Skin Lesion Classification Using Multitask Multimodal Neural Nets," (in English), *Ieee J Biomed Health,* vol. 23, no. 2, pp. 538-546, Mar 2019, doi: 10.1109/Jbhi.2018.2824327.

[40] P. Tschandl, C. Rosendahl, and H. Kittler, "The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions," (in English), *Sci Data,* vol. 5, Aug 14 2018, doi: ARTN 180161
10.1038/sdata.2018.161.

[41] *SIIM-ISIC 202 Challenge dataset*, International Skin Imaging Collaboration, doi: https://doi.org/10.34970/2020-ds01.

[42] B. Cassidy, C. Kendrick, A. Brodzicki, J. Jaworek-Korjakowska, and M. H. Yap, "Analysis of the ISIC image datasets: Usage, benchmarks and recommendations," *Med Image Anal,* vol. 75, p. 102305, 2022/01/01/ 2022, doi: https://doi.org/10.1016/j.media.2021.102305.

[43] S. Krig, "Image Pre-Processing," in *Computer Vision Metrics: Survey, Taxonomy, and Analysis*. Berkeley, CA: Apress, 2014.

[44] S. Diana and R. B, "Skin cancer detection and stage prediction using image processing techniques," *2018,* Back Propagation; Feature Extraction; Histogram Equalization; Melanoma; Neural Network; Non-Melanoma; Noise Removal; Segmentation. vol. 7, no. 1, p. 6, 2018-02-27 2018, doi: 10.14419/ijet.v7i1.8643.

[45] M. Mahmoud, "Improving skin cancer (melanoma) detection : new method," 2014.

[46] M. Elgendy, *Deep Learning for Vision Systems*. Manning Publications, 2020.

[47] B. Kumar. "Image Preprocessing — Why is it Necessary?" Spider, the Research and Development Club of National Institute of Technology,. https://medium.com/spidernitt/image-preprocessing-why-is-it-necessary-8895b8b08c1d (accessed 10 May 2021.

[48] N. C. F. C. David Gutman, Emre Celebi, Brian Helba, Michael Marchetti, Nabin Mishra, Allan Halpern. *Skin Lesion Analysis toward Melanoma Detection: A Challenge at the International Symposium on Biomedical Imaging (ISBI) 2016*. [Online]. Available: https://challenge.isic-archive.com/data

[49] E. Vocaturo, E. Zumpano, and P. Veltri, *Image pre-processing in computer vision systems for melanoma detection*. 2018, pp. 2117-2124.

[50] N. Faruqui. "What is Image Enhancement?" https://www.nzfaruqui.com/what-is-image-enhancement/ (accessed 07/08/2021.

[51] S. Wang *et al.*, "Association between quantitative and qualitative image features of contrast-enhanced mammography and molecular subtypes of breast cancer," *Quantitative Imaging in Medicine and Surgery,* vol. 12, no. 2, pp. 1270-1280, 2021. [Online]. Available: https://qims.amegroups.com/article/view/82455.

[52] M. A. A. Moustaf and H. M. Ismaiel, "Quantitative and qualitative evaluations of image enhancement techniques," in *2003 46th Midwest Symposium on Circuits and Systems*, 27-30 Dec. 2003 2003, vol. 2, pp. 664-669 Vol. 2, doi: 10.1109/MWSCAS.2003.1562374. [Online]. Available: https://ieeexplore.ieee.org/document/1562374/

[53] S. A. A. Fazel, Y. Gal, Z. Yang, and V. Vegh, "Qualitative and Quantitave Analysis of Six Image Fusion Methodologies and Their Application to Medical Imaging," in *2011 International Conference on Digital Image Computing: Techniques and Applications*, 6-8 Dec. 2011 2011, pp. 308-313, doi: 10.1109/DICTA.2011.58.

[54] "Edge Detection." Mathworks. https://www.mathworks.com/discovery/edge-detection.html (accessed 05/08/2021.

[55] S. Jayaraman, S. Esakkirajan, and T. Veerakumar, *Digital image processing*. New Delhi: Tata McGraw Hill Education (in English), 2009.

[56] R. C. Gonzalez, R. E. Woods, and S. L. Eddins, *Digital image processing using MATLAB*, 3rd ed. Mcgraw Hill (in English), 2020.

[57] S. W. Smith, *The scientist and engineer's guide to digital signal processing*. California Technical Publishing, 1997.

[58] M. Kalra and S. Kumar, "Various image enhancement techniques for skin cancer detection using mobile app," in *2015 International Conference on Computer, Communication and Control (IC4)*, 10-12 Sept. 2015 2015, pp. 1-6, doi: 10.1109/IC4.2015.7375681. [Online]. Available: https://ieeexplore.ieee.org/document/7375681/

[59] E. Erwin, "Improving Retinal Image Quality Using the Contrast Stretching, Histogram Equalization, and CLAHE Methods with Median Filters," *International Journal of Image, Graphics and Signal Processing,* vol. 12, pp. 30-41, 04/08 2020, doi: 10.5815/ijigsp.2020.02.04.

[60] H. Kaur and N. Sohi, "A Study for Applications of Histogram in Image Enhancement," *The International Journal of Engineering and Science,* vol. 06, pp. 59-63, 06/01 2017, doi: 10.9790/1813-0606015963.

[61] A. Toet and T. Wu, "Efficient contrast enhancement through log-power histogram modification," *J Electron Imaging,* vol. 23, p. 063017, 12/17 2014, doi: 10.1117/1.JEI.23.6.063017.

[62] C. Solomon, T. Breckon, and T. Breckon, *Fundamentals of Digital Image Processing : A Practical Approach with Examples in Matlab*. New York, UNITED KINGDOM: John Wiley & Sons, Incorporated, 2011.

[63] V. Sharma, D. Soni, and D. Srivastava, "Filtration Based Noise Reduction Technique in an Image," in *2019 4th International Conference on Internet of Things: Smart Innovation and Usages (IoT-SIU)*, 18-19 April 2019 2019, pp. 1-6, doi: 10.1109/IoT-SIU.2019.8777623.

[64] A. C. Bovik, Bovik, and A. C. Bovik, *The Essential Guide to Image Processing*. San Diego, UNITED STATES: Elsevier Science & Technology, 2009.

[65] G. Gupta, "Algorithm for Image Processing Using Improved Median Filter and Comparison of Mean, Median and Improved Median Filter," 2011.

[66] S. Sengupta, N. Mittal, and M. Modi, "Improved skin lesion edge detection method using Ant Colony Optimization," (in English), *Skin Research and Technology,* vol. 25, no. 6, pp. 846-856, Nov 2019, doi: 10.1111/srt.12744.

[67] J. R. Parker, *Algorithms for Image Processing and Computer Vision*. Wiley Publishing, 2010.

[68] K. Deguchi and P. K. Ghosh, *Mathematics of Shape Description : A Morphological Approach to Image Processing and Computer Graphics*. Hoboken, SINGAPORE: John Wiley & Sons, Incorporated, 2008.

[69] E. S. Zbarsky. "Binary Morphology in Image Processing." https://github.com/MathWorks-Teaching-Resources/Morphology-in-Image-Processing (accessed 25 April 2022.

[70] N. Dhanachandra, K. Manglem, and Y. J. Chanu, "Image Segmentation Using K -means Clustering Algorithm and Subtractive Clustering Algorithm," *Procedia Computer Science,* vol. 54, pp. 764-771, 2015/01/01/ 2015, doi: https://doi.org/10.1016/j.procs.2015.06.090.

[71] O. Wong and P. Rajendran, "Image Segmentation Using Modified Region-Based Active Contour Model," *Journal of Engineering and Applied Sciences,* vol. 14, pp. 5710-5718, 05/01 2019, doi: 10.36478/jeasci.2019.5710.5718.

[72] R. J. Hemalatha, T. R. Thamizhvani, A. Dhivya, J. Joseph, B. Babu, and R. Chandrasekaran, "Active Contour Based Segmentation Techniques for Medical Image Analysis," 2018.

[73] Z. Ihab, "Diagnosis of Skin Lesions Based on Dermoscopic Images Using Image Processing Techniques," in *Pattern Recognition*, Z. Andrzej Ed. Rijeka: IntechOpen, 2019, p. Ch. 6.

[74] V. T. Y. Ng, B. Y. M. Fung, and T. K. Lee, "Determining the asymmetry of skin lesion with fuzzy borders," *Comput Biol Med,* vol. 35, no. 2, pp. 103-120, 2005/02/01/ 2005, doi: https://doi.org/10.1016/j.compbiomed.2003.11.004.

[75] A.-R. Ali, J. Li, and S. O'Shea, "Towards the automatic detection of skin lesion shape asymmetry, color variegation and diameter in dermoscopic images," *Plos One,* vol. 15, p. e0234352, 06/16 2020, doi: 10.1371/journal.pone.0234352.

[76] J. Jaworek-Korjakowska, "Novel Method for Border Irregularity Assessment in Dermoscopic Color Images," *Computational and Mathematical Methods in Medicine,* vol. 2015, p. 496202, 2015/10/29 2015, doi: 10.1155/2015/496202.

[77] Y. Zhou, M. Smith, L. Smith, and R. Warr, "A new method describing border irregularity of pigmented lesions," *Skin Research and Technology,* vol. 16, no. 1, pp. 66-76, 2010, doi: https://doi.org/10.1111/j.1600-0846.2009.00403.x.

[78] P. Rani, B. Aliahmad, and D. Kumar, "Curve irregularity index for quantification of roughness in non-fractal curves," *Physica A: Statistical Mechanics and its Applications,* vol. 528, p. 121435, 05/01 2019, doi: 10.1016/j.physa.2019.121435.

[79] S. H. Carbonetto and S. E. Lew, "Characterization of border structure using fractal dimension in melanomas," in *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*, 31 Aug.-4 Sept. 2010 2010, pp. 4088-4091, doi: 10.1109/IEMBS.2010.5627296.

[80] A. Halpern. "Melanoma warning signs." Skin Cancer Foundation. https://www.skincancer.org/skin-cancer-information/melanoma/melanoma-warning-signs-and-images/ (accessed January 2022.

[81] T. Mendonça, P. M. Ferreira, J. S. Marques, A. R. S. Marcal, and J. Rozeira, "PH2 - A dermoscopic image database for research and benchmarking," in *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 3-7 July 2013 2013, pp. 5437-5440, doi: 10.1109/EMBC.2013.6610779.

[82] C. Barata, M. E. Celebi, J. Marques, and J. Rozeira, "Clinically inspired analysis of dermoscopy images using a generative model," *Computer Vision and Image Understanding,* vol. 151, pp. 124-137, 10/31 2016, doi: 10.1016/j.cviu.2015.09.011.

[83] B. Amaliah, C. Fatichah, and M. R. Widyanto, "ABCD FEATURE EXTRACTION OF IMAGE DERMATOSCOPIC BASED ON MORPHOLOGY ANALYSIS FOR MELANOMA SKIN CANCER DIAGNOSIS," 2012.

[84] M. Mohammed, H. Cherifi, and A. Bessaid, *Segmentation and ABCD rule extraction for skin tumors classification*. 2021.

[85] M. Reshma and B. Priestly, "Novel ABCD Formula to Diagnose and Feature Ranking of Melanoma," *Int J Adv Comput Sc,* vol. 10, 01/01 2019, doi: 10.14569/IJACSA.2019.0100111.

[86] C.-N. Anagnostopoulos, D. Vergados, I. Anagnostopoulos, and P. Mintzias, *Total dermoscopy score calculation using quantitative measurements in digital dermoscopy*. 2014, pp. 6744-7.

[87] F. Nachbar *et al.*, "The Abcd Rule of Dermatoscopy - High Prospective Value in the Diagnosis of Doubtful Melanocytic Skin-Lesions," (in English), *J Am Acad Dermatol,* vol. 30, no. 4, pp. 551-559, Apr 1994, doi: Doi 10.1016/S0190-9622(94)70061-3.

[88] R. Troncy, B. Huet, and S. Schenk, *Machine Learning Techniques for Multimedia Analysis*. 2011, pp. 59-59-80.

[89] C. D. Manning, P. Raghavan, and H. Schütze, *Introduction to information retrieval / Christopher D. Manning, Prabhakar Raghavan, Hinrich Schütze*. 2008.

[90] D. A. Pisner and D. M. Schnyer, "Chapter 6 - Support vector machine," in *Machine Learning*, A. Mechelli and S. Vieira Eds.: Academic Press, 2020, pp. 101-121.

[91] T. Kim and J. S. Lee, "Exponential Loss Minimization for Learning Weighted Naive Bayes Classifiers," *Ieee Access,* vol. 10, pp. 22724-22736, 2022, doi: 10.1109/ACCESS.2022.3155231.

[92] R. S. Kharisma, Muttafi'ah, and A. Dahlan, "Comparison of Naïve Bayes Algorithm Model Combinations with Term Weighting Techniques in Sentiment Analysis," in *2021 4th International Conference on Information and Communications Technology (ICOIACT)*, 30-31 Aug. 2021 2021, pp. 160-163, doi: 10.1109/ICOIACT53268.2021.9563999.

[93] M. Somvanshi, P. Chavan, S. Tambade, and S. V. Shinde, "A review of machine learning techniques using decision tree and support vector machine," in *2016 International Conference on Computing Communication Control and automation (ICCUBEA)*, 12-13 Aug. 2016 2016, pp. 1-7, doi: 10.1109/ICCUBEA.2016.7860040. [Online]. Available: https://ieeexplore.ieee.org/document/7860040/

[94] J. Bell, *Machine Learning: Hands-On for Developers and Technical Professionals*. John Wiley & Sons Inc, 2014, p. 408.

[95] A. Goel and S. K. Srivastava, "Role of Kernel Parameters in Performance Evaluation of SVM," in *2016 Second International Conference on Computational Intelligence & Communication Technology (CICT)*, 12-13 Feb. 2016 2016, pp. 166-169, doi: 10.1109/CICT.2016.40.

[96] S. Jayatilake and G. U. Ganegoda, "Melanoma Skin Cancer Detection from Dermoscopic Images using Computer Vision," in *2022 2nd International Conference on Image Processing and Robotics (ICIPRob)*, 12-13 March 2022 2022, pp. 1-6, doi: 10.1109/ICIPRob54042.2022.9798723.

[97] K. Chaiyakhan, P. Chujai, R. Chanklan, and N. Kerdprasop, "Classification of Melanoma Considering the Adaptive Additional Shape Feature," *International Journal of Future Computer and Communication,* vol. 8, pp. 114-118, 12/01 2019, doi: 10.18178/ijfcc.2019.8.4.551.

[98] D. N. H. Thanh, V. B. S. Prasath, L. M. Hieu, and N. N. Hien, "Melanoma Skin Cancer Detection Method Based on Adaptive Principal Curvature, Colour Normalisation and Feature Extraction with the ABCD Rule," (in eng), *J Digit Imaging,* vol. 33, no. 3, pp. 574-585, Jun 2020, doi: 10.1007/s10278-019-00316-x.

[99] D. Gautam and M. Ahmed, "Melanoma detection and classification using SVM based decision support system," in *2015 Annual IEEE India Conference (INDICON)*, 17-20 Dec. 2015 2015, pp. 1-6, doi: 10.1109/INDICON.2015.7443447.

[100] S. A.S*, Graceline.H, H., & Jey J B, A., "Highly Accurate Melanoma Detection from Skin Images using Multiple Feature Extraction and Support Vector Machine (SVM)." *International Journal of Recent Technology and Engineering (IJRTE),* vol. 8, no. 5, 2020. [Online]. Available: https://doi.org/10.35940/ijrte.d9150.018520.

[101] S. Majumder and M. A. Ullah, "Feature extraction from dermoscopy images for melanoma diagnosis," *Sn Appl Sci,* vol. 1, no. 7, p. 753, 2019/06/20 2019, doi: 10.1007/s42452-019-0786-8.

[102] E. M. Senan and M. E. Jadhav, "Analysis of dermoscopy images by using ABCD rule for early detection of skin cancer," *Global Transitions Proceedings,* vol. 2, no. 1, pp. 1-7, 2021/06/01/ 2021, doi: https://doi.org/10.1016/j.gltp.2021.01.001.

[103] S. Oukil, R. Kasmi, K. Mokrani, and B. García-Zapirain, "Automatic segmentation and melanoma detection based on color and texture features in dermoscopic images," *Skin Research and Technology,* vol. 28, no. 2, pp. 203-211, 2022, doi: https://doi.org/10.1111/srt.13111.

[104] A. Amoabedini, M. S. Farsani, H. Saberkari, and E. Aminian, "Employing the Local Radon Transform for Melanoma Segmentation in Dermoscopic Images," (in eng), *J Med Signals Sens,* vol. 8, no. 3, pp. 184-194, Jul-Sep 2018, doi: 10.4103/jmss.JMSS_40_17.

[105] A. P. Chakkaravarthy and A. Chandrasekar, "Anatomical region segmentation method from dermoscopic images of pigmented skin lesions," (in English), *Int J Imag Syst Tech,* vol. 30, no. 3, pp. 636-652, Sep 2020, doi: 10.1002/ima.22404.

[106] W. Al-Zyoud, A. A. Helou, E. AlQasem, and N. A. Rawashdeh, "Visual feature extraction from dermoscopic colour images for classification of melanocytic skin lesions," *EurAsian Journal of BioSciences,* vol. 14, pp. 1299-1307, 2020.

[107] M. MESSADI, S. MAHMOUDI, and A. BESSAID, "ANALYSIS OF SPECIFIC PARAMETERS FOR SKIN TUMOR CLASSIFICATION," *Journal of Mechanics in Medicine and Biology,* vol. 22, no. 04, p. 2250017, 2022, doi: 10.1142/s0219519422500178.

[108] Z. Ma and J. M. R. S. Tavares, "Effective features to classify skin lesions in dermoscopic images," (in English), *Expert Syst Appl,* vol. 84, pp. 92-101, Oct 30 2017, doi: 10.1016/j.eswa.2017.05.003.

[109] K. Abhishek, J. Kawahara, and G. Hamarneh, "Predicting the clinical management of skin lesions using deep learning," (in eng), *Sci Rep-Uk,* vol. 11, no. 1, pp. 7769-7769, 2021, doi: 10.1038/s41598-021-87064-7.

[110] I. Zalaudek *et al.*, "Three-point checklist of dermoscopy: an open internet study," *Brit J Dermatol,* vol. 154, no. 3, pp. 431-437, 2006, doi: https://doi.org/10.1111/j.1365-2133.2005.06983.x.

[111] M. Pérez-Ortiz, A. Sáez, J. Sánchez-Monedero, P. A. Gutiérrez, and C. Hervás-Martínez, "Tackling the ordinal and imbalance nature of a melanoma image classification problem," in *2016 International Joint Conference on Neural Networks (IJCNN)*, 24-29 July 2016 2016, pp. 2156-2163, doi: 10.1109/IJCNN.2016.7727466.

[112] B. Lee, J. Choi, K. Yun, and J. Y. Choi, "Gradient preserving RGB-to-gray conversion using random forest," in *2015 IEEE International Conference on Image Processing (ICIP)*, 27-30 Sept. 2015 2015, pp. 3170-3174, doi: 10.1109/ICIP.2015.7351388.

[113] "Convert RGB image or colormap to grayscale." Mathworks Help Centre. https://www.mathworks.com/help/matlab/ref/rgb2gray.html (accessed 07 March 2022.

[114] J. Chaki and N. Dey, *A Beginner's Guide to Image Shape Feature Extraction Techniques*. 2019.

[115] L. Ma, L. Cheng, W. Han, L. Zhong, and M. Li, "Cultivated land information extraction from high-resolution unmanned aerial vehicle imagery data," *Journal of Applied Remote Sensing,* vol. 8, pp. 083673-1, 02/20 2014.

[116] M. E. Celebi *et al.*, "A methodological approach to the classification of dermoscopy images," *Comput Med Imag Grap,* vol. 31, no. 6, pp. 362-373, 2007/09/01/ 2007, doi: https://doi.org/10.1016/j.compmedimag.2007.01.003.

[117] C. Boucher. "Isoperimetric Inequality for Polygons." http://demonstrations.wolfram.com/IsoperimetricInequalityForPolygons/ (accessed 06 June 2022.

[118] Q. Jin. "ABCDEFG of melanoma." https://dermnetnz.org/topics/abcdes-of-melanoma (accessed 01 May, 2022).

[119] S. Majumder and M. Ullah, "A Computational Approach to Pertinent Feature Extraction for Diagnosis of Melanoma Skin Lesion," *Pattern Recognition and Image Analysis,* vol. 29, pp. 503-514, 07/01 2019, doi: 10.1134/S1054661819030131.

[120] N. Katariya, D. D. G. Thakore, and D. U. K. Jaliya, "A Survey on Image Processing Techniques to Detect Melanoma," *International Journal for Research in Applied Science and Engineering Technology (IJRASET),* vol. 6, no. IV, pp. 669-673, 2018.

[121] B. Erkol, R. H. Moss, R. J. Stanley, W. V. Stoecker, and E. Hvatum, "Automatic lesion boundary detection in dermoscopy images using gradient vector flow snakes," (in eng), *Skin research and technology : official journal of International Society for Bioengineering and the Skin (ISBS) [and] International Society for Digital Imaging of Skin (ISDIS) [and] International Society for Skin Imaging (ISSI),* vol. 11, no. 1, pp. 17-26, 2005, doi: 10.1111/j.1600-0846.2005.00092.x.

[122] F. Ian, G. Rayid, S. J. Ron, K. Frauke, and L. Julia, *Big Data and Social Science : A Practical Guide to Methods and Tools*, 2017. [Online]. Available: https://search.ebscohost.com/login.aspx?direct=true&db=nlebk&AN=1353316&site=eds-live. Accessed on: 2017-01-01.

[123] C. Kumar and V. Kumar, "Vegetable Plant Leaf Image Classification Using Machine Learning Models," in *Proceedings of Third International Conference on Advances in Computer Engineering*

and Communication Systems, Singapore, A. B. Reddy, S. Nagini, V. E. Balas, and K. S. Raju, Eds., 2023// 2023: Springer Nature Singapore, pp. 31-45.

[124] V. D. P. Jasti *et al.*, "Computational Technique Based on Machine Learning and Image Processing for Medical Image Analysis of Breast Cancer Diagnosis," *Security and Communication Networks,* vol. 2022, p. 1918379, 2022/03/09 2022, doi: 10.1155/2022/1918379.

[125] R. Li and S. Li, "Multimedia Image Data Analysis Based on KNN Algorithm," *Comput Intel Neurosc,* vol. 2022, p. 7963603, 2022/04/12 2022, doi: 10.1155/2022/7963603.

[126] S. Lahmiri, "Integrating convolutional neural networks, kNN, and Bayesian optimization for efficient diagnosis of Alzheimer's disease in magnetic resonance images," *Biomed Signal Proces,* vol. 80, p. 104375, 2023/02/01/ 2023, doi: https://doi.org/10.1016/j.bspc.2022.104375.

[127] S. Nivetha and H. H. Inbarani, "Neighborhood Rough Neural Network Approach for COVID-19 Image Classification," *Neural Process Lett,* vol. 54, no. 3, pp. 1919-1941, 2022/06/01 2022, doi: 10.1007/s11063-021-10712-6.

[128] K. Chomboon, P. Chujai, P. Teerarassammee, K. Kerdprasop, and N. Kerdprasop, *An Empirical Study of Distance Metrics for k-Nearest Neighbor Algorithm*. 2015, pp. 280-285.

[129] K. Taunk, S. De, S. Verma, and A. Swetapadma, "A Brief Review of Nearest Neighbor Algorithm for Learning and Classification," in *2019 International Conference on Intelligent Computing and Control Systems (ICCS)*, 15-17 May 2019 2019, pp. 1255-1260, doi: 10.1109/ICCS45141.2019.9065747.

[130] Y. Huang and L. Li, "Naive Bayes classification algorithm based on small sample set," in *2011 IEEE International Conference on Cloud Computing and Intelligence Systems*, 15-17 Sept. 2011 2011, pp. 34-39, doi: 10.1109/CCIS.2011.6045027.

[131] A. Tharwat, "Classification assessment methods," *Applied Computing and Informatics,* vol. 17, no. 1, pp. 168-192, 2021, doi: 10.1016/j.aci.2018.08.003.

[132] H. Bhavsar and A. Ganatra, "A Comparative Study of Training Algorithms for Supervised Machine Learning," *International Journal of Soft Computing and Engineering (IJSCE),* vol. 2, 01/01 2012.

[133] P. Tschandl, G. Argenziano, M. Razmara, and J. Yap, "Diagnostic accuracy of content-based dermatoscopic image retrieval with deep classification features," *Brit J Dermatol,* vol. 181, no. 1, pp. 155-165, 2019, doi: https://doi.org/10.1111/bjd.17189.

[134] C. Barata, M. E. Celebi, and J. Marques, "Toward a Robust Analysis of Dermoscopy Images Acquired under Different Conditions," 2015, pp. 1-22.

[135] T. Nedelcu, M. Vasconcelos, and A. Carreiro, *Multi-Dataset Training for Skin Lesion Classification on Multimodal and Multitask Deep Learning*. 2020.

[136] R. Kasmi and K. Mokrani, "Classification of malignant melanoma and benign skin lesions: Implementation of automatic ABCD rule," *Iet Image Process,* vol. 10, 01/17 2016, doi: 10.1049/iet-ipr.2015.0385.

[137] M. Styła and T. Giżewski, "The Study of Usefulness of a Set of Fractal Parameters to Build Classes of Disease Units Based on Images of Pigmented Skin Lesions," *Diagnostics,* vol. 11, p. 1773, 09/26 2021, doi: 10.3390/diagnostics11101773.

[138] K. Das *et al.*, "Machine Learning and Its Application in Skin Cancer," (in eng), *Int J Environ Res Public Health,* vol. 18, no. 24, Dec 20 2021, doi: 10.3390/ijerph182413409.

[139] D. N. A. Ningrum *et al.*, "Deep Learning Classifier with Patient's Metadata of Dermoscopic Images in Malignant Melanoma Detection," (in eng), *J Multidiscip Healthc,* vol. 14, pp. 877-885, 2021, doi: 10.2147/jmdh.S306284.

VII. APPENDIX

A. Training dataset

Type	Filename	Asymmetry	Border	Colour	Diameter	Classification
Melanoma in situ	FBL026.jpg	0.1395	0.8464	2	7.573	Melanoma
Melanoma in situ	gzl28.jpg	0.1248	0.7778	3	11.7478	Melanoma
Melanoma in situ	Nal020.jpg	0.8306	0.2827	2	9.3843	Melanoma
Melanoma in situ	Nal047.jpg	0.2025	0.5279	2	10.2245	Melanoma
Melanoma in situ	Nal022.jpg	0.0762	0.8452	2	9.0816	Melanoma
Melanoma in situ	Nal073.jpg	0.2511	0.4648	2	8.7944	Melanoma
Melanoma in situ	Nal075.jpg	0.5818	0.2509	2	14.1093	Melanoma
Melanoma in situ	Nal077.jpg	0.357	0.4617	3	11.4205	Melanoma
Melanoma in situ	Nal080.jpg	0.3327	0.343	2	9.6509	Melanoma
Melanoma in situ	Nal084.jpg	0.1469	0.6247	2	9.2079	Melanoma
Melanoma in situ	Nal088.jpg	0.5709	0.3603	1	6.4723	Melanoma
Melanoma in situ	Nal090.jpg	0.3358	0.2684	3	11.9828	Melanoma
Melanoma in situ	Nal099.jpg	0.3094	0.3256	3	13.6911	Melanoma
Melanoma in situ	Nbl001.jpg	0.1696	0.6092	3	11.0111	Melanoma
Melanoma in situ	Nbl022.jpg	0.1549	0.6325	2	14.0826	Melanoma
Melanoma in situ	Ngl002.jpg	0.1319	0.8013	2	9.5688	Melanoma
Melanoma in situ	NGL010.JPG	0.1383	0.6623	3	9.5207	Melanoma
Melanoma in situ	Ngl068.jpg	0.1361	0.6378	2	13.939	Melanoma
Melanoma in situ	Nil031.jpg	0.1039	0.8155	1	5.9267	Melanoma
Melanoma in situ	Nil084.jpg	0.1039	0.8056	2	11.0154	Melanoma
Melanoma in situ	Adl301.jpg	0.229	0.5869	2	6.6572	Melanoma
Melanoma in situ	Ael496.jpg	0.1109	0.8698	1	4.5143	Melanoma
Melanoma in situ	Fbl028.jpg	0.3154	0.1847	3	12.0113	Melanoma
Melanoma in situ	Fcl040.jpg	0.1497	0.78	3	8.9367	Melanoma
Melanoma in situ	FCL058.jpg	0.2955	0.5859	2	8.5943	Melanoma
Melanoma in situ	FDL036.jpg	0.2273	0.5899	3	7.1952	Melanoma
Melanoma in situ	Fgl018.jpg	0.1375	0.6819	3	8.9788	Melanoma
Melanoma in situ	Fil004.jpg	0.2452	0.5088	3	9.671	Melanoma
Melanoma in situ	Fil070.jpg	0.1078	0.595	3	10.7296	Melanoma
Melanoma in situ	Fll122.jpg	0.1118	0.7756	2	12.129	Melanoma
Melanoma in situ	Gal240.jpg	0.2143	0.71	3	8.4328	Melanoma
Melanoma in situ	Ggl036.jpg	0.1376	0.628	2	13.759	Melanoma
Melanoma <0.76mm	Adl305.jpg	0.2353	0.566	1	5.9404	Melanoma

Melanoma <0.76mm	Ael490.jpg	0.5783	0.3067	3	8.3057	Melanoma
Melanoma <0.76mm	FBL046. jpg	0.1805	0.7354	3	10.5572	Melanoma
Melanoma <0.76mm	Fbl050.jpg	0.1128	0.692	2	9.311	Melanoma
Melanoma <0.76mm	FBL058. jpg	0.141	0.6444	2	7.8299	Melanoma
Melanoma <0.76mm	Fcl100.jpg	0.1607	0.6217	3	12.3949	Melanoma
Melanoma <0.76mm	Ffl092.jpg	0.194	0.5806	2	9.6118	Melanoma
Melanoma <0.76mm	Fgl082.jpg	0.2667	0.1164	3	13.6683	Melanoma
Melanoma <0.76mm	Fil006.jpg	0.2164	0.6022	2	10.0036	Melanoma
Melanoma <0.76mm	Fil032.jpg	0.1506	0.7315	3	8.0021	Melanoma
Melanoma <0.76mm	Fll071.jpg	0.1896	0.5479	3	14.4103	Melanoma
Melanoma <0.76mm	Fll108.jpg	0.2105	0.5119	2	9.8176	Melanoma
Melanoma <0.76mm	Fll132.jpg	0.1489	0.7272	2	8.4823	Melanoma
Melanoma <0.76mm	Gal113.jpg	0.1296	0.7228	3	13.0095	Melanoma
Melanoma <0.76mm	Gal211.jpg	0.1579	0.7928	3	10.0051	Melanoma
Melanoma <0.76mm	Gal257.jpg	0.2143	0.6708	3	10.7108	Melanoma
Melanoma <0.76mm	Gcl094.jpg	0.6767	0.2344	3	8.8989	Melanoma
Melanoma <0.76mm	Gdl014.jpg	0.2251	0.7797	2	8.3926	Melanoma
Melanoma <0.76mm	Gdl021.jpg	0.4584	0.4819	1	10.5353	Melanoma
Melanoma <0.76mm	Ggl058.jpg	0.2209	0.6906	2	11.9471	Melanoma
Melanoma <0.76mm	gzl04.jpg	0.2	0.6969	3	9.5041	Melanoma
Melanoma <0.76mm	Gal008.jpg	0.1034	0.6623	2	7.6834	Melanoma
Melanoma <0.76mm	gzl76.jpg	0.2043	0.6841	2	14.7417	Melanoma

Melanoma <0.76mm	Nal018.jpg	0.1413	0.6864	1	7.2112	Melanoma
Melanoma <0.76mm	Nal025.jpg	0.3634	0.5345	1	8.2003	Melanoma
Melanoma <0.76mm	Nal035.jpg	0.1353	0.6521	3	12.9267	Melanoma
Melanoma <0.76mm	Nal049.jpg	0.356	0.374	3	14.1148	Melanoma
Melanoma <0.76mm	nal095.jpg	0.4897	0.3106	2	9.6393	Melanoma
Melanoma <0.76mm	Nbl006.jpg	0.416	0.5281	2	8.1709	Melanoma
Melanoma <0.76mm	Nbl017.jpg	0.1387	0.7509	3	11.6391	Melanoma
Melanoma <0.76mm	Nbl028.jpg	0.2106	0.6641	3	8.7136	Melanoma
Melanoma <0.76mm	Nbl030.jpg	0.1965	0.5488	2	9.6622	Melanoma
Melanoma <0.76mm	Nbl034.jpg	0.3089	0.5347	3	13.1307	Melanoma
Melanoma <0.76mm	newl014.jpg	0.1853	0.7294	2	6.5695	Melanoma
Melanoma <0.76mm	Nfl063.jpg	0.2146	0.7035	2	8.4237	Melanoma
Melanoma <0.76mm	nfl072.jpg	0.2299	0.5233	2	9.8134	Melanoma
Melanoma <0.76mm	Ngl008.jpg	0.2569	0.4635	3	9.7298	Melanoma
Melanoma <0.76mm	Ngl013.jpg	0.1692	0.651	3	10.9227	Melanoma
Melanoma <0.76mm	Ngl017.jpg	0.2959	0.4488	3	9.2911	Melanoma
Melanoma <0.76mm	Ngl019.jpg	0.3475	0.3808	3	11.9645	Melanoma
Melanoma <0.76mm	Ngl021.jpg	0.1044	0.7313	3	11.7752	Melanoma
Melanoma <0.76mm	Ngl027.jpg	0.1991	0.492	3	10.961	Melanoma
Melanoma <0.76mm	Ngl039.jpg	0.1082	0.7573	2	11.7068	Melanoma
Melanoma <0.76mm	Ngl055.jpg	0.199	0.5133	2	9.1497	Melanoma
Melanoma <0.76mm	Ngl100.jpg	0.1556	0.6895	3	13.2728	Melanoma

Melanoma <0.76mm	Nil066.jpg	0.1309	0.6934	2	12.2295	Melanoma
Melanoma <0.76mm	Nil078.jpg	0.197	0.5143	3	11.2821	Melanoma
Melanoma <0.76mm	Nil096.jpg	0.1596	0.551	3	12.6329	Melanoma
Melanoma <0.76mm	Nll002.jpg	0.3005	0.5964	3	8.8935	Melanoma
Melanoma <0.76mm	Nml077.jpg	0.3345	0.5748	2	8.6225	Melanoma
Melanoma <0.76mm	Nml079.jpg	0.1289	0.8159	3	7.0701	Melanoma
Benign	Aal004.jpg	0.1963	0.7498	2	7.749	Non Melanoma
Benign	Aal006.jpg	0.1921	0.7443	3	6.6897	Non Melanoma
Benign	Aal010.jpg	0.2077	0.7182	2	7.3635	Non Melanoma
Benign	Aal056.jpg	0.1275	0.6671	1	4.704	Non Melanoma
Benign	Aal062.jpg	0.1328	0.5101	2	7.1483	Non Melanoma
Benign	Aal086.jpg	0.3542	0.557	2	5.8217	Non Melanoma
Benign	Ggl002.jpg	0.1968	0.7362	2	12.3949	Non Melanoma
Benign	Abl189.jpg	0.153	0.663	3	8.6962	Non Melanoma
Benign	FDL016.JPG	0.1312	0.604	3	12.2615	Non Melanoma
Benign	FDL022.JPG	0.1504	0.6254	3	11.5858	Non Melanoma
Benign	Acl263.jpg	0.1924	0.386	2	13.5276	Non Melanoma
Benign	Adl331.jpg	0.1228	0.6284	2	12.9831	Non Melanoma
Benign	Fgl020.jpg	0.1361	0.6499	3	12.8445	Non Melanoma
Benign	Fgl035.jpg	0.1746	0.351	2	11.4278	Non Melanoma
Benign	Ggl008.jpg	0.0982	0.6924	2	8.0895	Non Melanoma
Benign	Aal030.jpg	0.2066	0.3923	1	8.7528	Non Melanoma

Benign	Aal038.jpg	0.1961	0.6153	1	7.4628	Non Melanoma
Benign	Acl231.jpg	0.0947	0.5663	2	11.4514	Non Melanoma
Benign	Acl233.jpg	0.203	0.4483	2	12.3175	Non Melanoma
Benign	Ael415.jpg	0.1899	0.5289	2	8.9026	Non Melanoma
Benign	Fil064.jpg	0.1229	0.8472	1	5.7763	Non Melanoma
Benign	newl010.jpg	0.5677	0.2378	1	6.4331	Non Melanoma
Benign	newl018.jpg	0.1287	0.8223	1	4.3038	Non Melanoma
Benign	Acl225.jpg	0.0705	0.7792	1	5.8588	Non Melanoma
Benign	Acl295.jpg	0.1314	0.8768	1	4.6827	Non Melanoma
Benign	Adl358.jpg	0.1579	0.7164	1	5.3985	Non Melanoma
Benign	Adl395.jpg	0.0782	0.8909	1	6.0471	Non Melanoma
Benign	Adl399.jpg	0.1481	0.7643	1	6.9103	Non Melanoma
Benign	Ael401.jpg	0.095	0.8474	1	7.1653	Non Melanoma
Benign	Ael405.jpg	0.1875	0.8501	1	3.9041	Non Melanoma
Benign	Ael407.jpg	0.1122	0.8309	1	5.253	Non Melanoma
Benign	Ael421.jpg	0.3911	0.2796	1	6.3466	Non Melanoma
Benign	Ael465.jpg	0.1663	0.6512	1	4.2573	Non Melanoma
Benign	gzl14.jpg	0.1591	0.7885	3	8.5664	Non Melanoma
Benign	gzl58.jpg	0.0882	0.781	2	8.709	Non Melanoma
Benign	Ncl020.jpg	0.096	0.8765	2	5.6491	Non Melanoma
Benign	Ncl032.jpg	0.1431	0.6297	1	5.4114	Non Melanoma
Benign	Ncl093.jpg	0.2155	0.6524	1	5.4432	Non Melanoma

Benign	Ncl095.jpg	0.0685	0.8814	2	5.4749	Non Melanoma
Benign	Ndl050.jpg	0.005	0.7802	3	18.8436	Non Melanoma
Benign	Ndl052.jpg	0.0542	0.8739	2	5.4662	Non Melanoma
Benign	Ndl054.jpg	0.1257	0.8354	2	5.8923	Non Melanoma
Benign	Ndl056.jpg	0.0906	0.8021	2	6.7619	Non Melanoma
Benign	Ndl058.jpg	0.0766	0.8466	2	10.7181	Non Melanoma
Benign	Ndl060.jpg	0.0811	0.9128	2	8.8032	Non Melanoma
Benign	NDL063.JPG	0.253	0.4346	2	7.3922	Non Melanoma
Benign	Ndl068.jpg	0.0771	0.9013	2	4.9049	Non Melanoma
Benign	Acl227.jpg	0.579	0.1484	2	9.9523	Non Melanoma
Benign	Acl229.jpg	0.2466	0.5395	3	11.2468	Non Melanoma
Benign	Fdl058.jpg	0.1694	0.456	2	10.4613	Non Melanoma
Benign	Fhl048.jpg	0.2919	0.5633	2	6.5735	Non Melanoma
Benign	Fll104.jpg	0.2259	0.5382	3	10.1004	Non Melanoma
Benign	Gal134.jpg	0.075	0.8795	2	8.0426	Non Melanoma
Benign	Ggl018.jpg	0.2078	0.5434	3	11.6388	Non Melanoma
Benign	Ggl046.jpg	0.0372	0.8654	3	10.8471	Non Melanoma
Benign	Ggl066.jpg	0.1531	0.8581	2	13.6761	Non Melanoma
Benign	Ggl012.jpg	0.1176	0.786	1	7.6653	Non Melanoma
Benign	Ggl014.jpg	0.269	0.5759	1	7.3651	Non Melanoma
Benign	Nel068.jpg	0.2944	0.3895	1	6.5337	Non Melanoma
Benign	Nel075.jpg	0.1295	0.7357	3	15.1686	Non Melanoma

Benign	Aal012.jpg	0.1389	0.7675	2	7.0855	Non Melanoma
Benign	Aal044.jpg	0.276	0.5992	2	6.59	Non Melanoma
Benign	Aal046.jpg	0.1023	0.8648	1	4.1965	Non Melanoma
Benign	Aal048.jpg	0.2302	0.6409	1	5.6158	Non Melanoma
Benign	Aal052.jpg	0.1883	0.6017	2	6.1512	Non Melanoma
Benign	Fbl030.jpg	0.2633	0.579	3	15.0692	Non Melanoma
Benign	FBL042.JPG	0.2166	0.8209	2	4.9785	Non Melanoma
Benign	Ffl042.jpg	0.0685	0.8039	3	8.5418	Non Melanoma
Benign	Abl197.jpg	0.1717	0.4668	1	8.9882	Non Melanoma
Benign	Acl211.jpg	0.2089	0.4116	1	7.1779	Non Melanoma
Benign	gzl48.jpg	0.3253	0.4308	2	10.4891	Non Melanoma
Benign	gzl60.jpg	0.2677	0.3986	1	6.8921	Non Melanoma
Benign	newl006.jpg	0.1591	0.6838	1	5.2237	Non Melanoma
Benign	Ael484.jpg	0.1018	0.7593	2	6.5689	Non Melanoma
Benign	Fbl018.jpg	0.1168	0.8468	1	5.4325	Non Melanoma
Benign	Fbl040.jpg	0.0755	0.8059	2	9.4223	Non Melanoma
Benign	Fel074.jpg	0.0941	0.8575	2	7.5133	Non Melanoma
Benign	Fel080.jpg	0.049	0.86	2	9.7819	Non Melanoma
Benign	Ffl034.jpg	0.243	0.3991	2	9.2826	Non Melanoma
Benign	Ffl084.jpg	0.0831	0.8229	2	8.6719	Non Melanoma
Benign	Fgl023.jpg	0.085	0.8798	1	5.227	Non Melanoma
Benign	Fgl037.jpg	0.0897	0.7879	2	8.4746	Non Melanoma

Benign	Nbl088.jpg	0.1028	0.7944	2	8.5985	Non Melanoma
Benign	Nbl100.jpg	0.1188	0.8308	1	4.1559	Non Melanoma
Benign	Fgl074.jpg	0.1404	0.7533	3	7.8813	Non Melanoma
Benign	Fgl087.jpg	0.0737	0.7978	2	8.7315	Non Melanoma
Benign	Fll077.jpg	0.2019	0.7374	2	8.6367	Non Melanoma
Benign	Gal222.jpg	0.1602	0.7742	1	3.9965	Non Melanoma
Benign	Gal238.jpg	0.1117	0.8677	1	4.521	Non Melanoma
Benign	Ggl010.jpg	0.1084	0.7505	2	6.4535	Non Melanoma
Benign	Ggl026.jpg	0.1889	0.6464	2	9.9892	Non Melanoma
Benign	Ggl086.jpg	0.1982	0.5872	2	14.0318	Non Melanoma
Benign	gzl30.jpg	0.1095	0.7915	3	7.8226	Non Melanoma
Benign	Ggl016.jpg	0.0248	0.9857	1	6.7839	Non Melanoma
Benign	Nel066.jpg	0.2192	0.4623	1	12.8814	Non Melanoma

B. Test dataset

Asymmetry	Border	Colour	Diameter
0.1705	0.5313	3	15.974
0.3184	0.5646	2	9.3773
0.2082	0.5173	3	10.0787
0.1407	0.7394	3	7.1387
0.1651	0.6278	3	7.1029
0.1492	0.6991	3	9.4484
0.0924	0.7292	3	15.5919
0.2784	0.1226	3	11.4648
0.3024	0.3858	3	12.7929
0.1614	0.5766	3	15.5005
0.2494	0.4938	2	10.066
0.1494	0.6987	3	10.3665
0.1338	0.7592	2	8.617
0.14	0.6442	3	10.2849
0.4627	0.34	3	15.1252
0.2197	0.7031	3	13.46
0.1371	0.6988	3	8.9995
0.2402	0.6117	3	15.8078
0.3213	0.129	3	11.778
0.2233	0.5619	3	11.58
0.1944	0.5557	1	4.4462
0.1085	0.8497	2	4.9437
0.2566	0.5883	1	3.9568
0.4106	0.3738	1	8.0816
0.0763	0.7308	2	16.3116
0.472	0.1604	1	10.7404
0.2468	0.2417	1	10.1954
0.1822	0.7395	1	4.143
0.3334	0.7162	1	3.4507
0.1584	0.8023	1	3.533
0.1176	0.7574	2	8.2296
0.1414	0.8537	1	3.5904
0.0972	0.8019	2	6.5105
0.2919	0.5633	2	6.5735
0.1313	0.5135	2	12.1429
0.2712	0.6053	2	6.9333
0.0665	0.9228	1	4.4366
0.0682	0.9704	1	3.8516
0.0545	0.8952	2	6.3379
0.2721	0.4708	1	7.929