Computer-aided Diagnosis of Melanoma

Rahil Garnavi

Submitted in total fulfillment of the requirements
of the degree of
Doctor of Philosophy

August, 2011

Department of Electrical and Electronic Engineering
The University of Melbourne
Australia

Produced on archival quality paper

ABSTRACT

Malignant melanoma is one of the most common and the deadliest type of skin cancer. In Australia, it represents 10% of all cancers and its incidence is four times higher than in Canada, the UK and the US. The worldwide steady increase in incidence of melanoma in recent years, its high mortality rate and the massive medical cost has made its early diagnosis a continuing priority of public health. Melanoma survival rate depends highly on its stage and thickness; advanced and thickened melanoma is still incurable, whereas early detection of thin melanoma and immediate surgical excision of the lesion shows effective prognosis. However, early diagnosis of melanoma is not trivial even for experienced dermatologists. Despite the use of advanced imaging technologies such as dermoscopy and existing clinical diagnostic algorithms such as the ABCD rule of dermoscopy, clinical diagnosis of melanoma is still challenging and its accuracy has been an issue of concern (estimated to be about 75–85%) especially with equivocal pigmented lesions.

A computer-based melanoma diagnostic system is able to provide quantitative and objective evaluation of the skin lesion versus the subjective clinical assessment. It allows for reproducible diagnosis by diminishing the inter- and intra-observer variabilities found in dermatologists' examinations. It also automates the analysis, thereby mitigates the amount of repetitive and tedious tasks to be done by physicians. Numerous groups have developed automated and semi-automated computer-based methods to analyse skin lesions and diagnose melanoma. However, the diversity of existing problems makes any further contributions greatly appreciated. Moreover, it is widely acknowledged that much higher accuracy is required for computer-based system to be considered reliable and trustworthy enough by clinicians, therefore be adopted routinely in their diagnostic process. With the aim of improving some of existing methods and developing new techniques to facilitate accurate, fast and reliable computer-based diagnosis of melanoma, this thesis makes contributions in various stages of development of a computer-aided diagnostic system of melanoma; namely, image segmentation or border detection, feature extraction, feature selection, and classification.

One of the contributions of this thesis is to develop a fast and accurate border detection method. Variegation of the colour within the lesion and the diversity found in human skin colour makes border detection a challenging process. To tackle this challenge, a comprehensive colour optimisation procedure is conducted in this thesis. It incorporates a number of colour spaces and a variety of dermoscopy images with respect to colour, texture and shape, and demonstrates that the X colour channel from the XYZ colour space, and XoYoR, which integrates the X and Y colour channels (from the XYZ colour space) with the R colour channel (from the RGB colour space), exhibit the most optimal separation between the lesion and background skin. The above-mentioned process is the main building block of our initial border detection method, called the Global method. The Global method applies global histogram thresholding on the optimal colour channels derived

from the colour optimisation process. This is followed by mathematical morphology and connected component analysis which results in automated border-detection of the lesions with an accuracy of approximately 97%. Comparative study carried out in this thesis, involving two other state-of-the-art automated border-detection methods, demonstrates that with a proper choice of colour channels the proposed Global border detection method, in spite of its simplicity, is highly competitive with well-known skin lesion segmentation methods. Furthermore, our proposed method is potentially faster as it mainly involves scalar processing as opposed to vector processing performed in those methods.

Considering the practical nature of melanoma diagnosis which calls for extreme caution when excluding portions of the image, in practice dermatologists choose the outmost detectable pigment to minimise the risk of incorrect diagnosis. Accordingly, the Global method is further improved in this thesis to emerge into the Hybrid border-detection method, which is shown to reduce the discrepancy between the manual and automatic borders. The Hybrid method is comprised of two stages; in the first stage the Global method is applied on a given lesion to detect an initial boundary, in the second stage the local colour information around the initial boundary is used to expand the border and form the final segmentation result. The proposed method is designed to increase specificity in the first stage (global thresholding) and sensitivity in the second stage (adaptive local thresholding). The performance of the method is quantitatively analysed and shown to be highly effective with an accuracy of 98% on a set of 85 high-resolution dermoscopy images. Taking manual borders drawn by an experienced dermatologist as the ground truth, experimental results show that border drawn by the Hybrid method are more accurate than borders drawn by a less-experienced dermatologist. Moreover, comparative study carried out in this thesis, involving three other state-of-the-art border-detection methods, shows that the proposed Hybrid method is superior than (with respect to some of the evaluation metrics) or highly competitive with those methods (with respect to other metrics).

Another contribution of this thesis is to provide a customised and comprehensive evaluation metric for border-detection method in dermoscopy images. This need arises from the fact that no objective performance assessment of the border-detection methods in dermoscopy images, which can genuinely reflect dermatologists' perspective in the melanoma diagnosis process, has been thoroughly investigated in the literature. Moreover, in using the standard metrics (*e.g.* sensitivity, specificity, border error, etc.), comparative studies can yield in paradoxical results, *i.e.*, a border detection method is superior to other method with respect to some metrics but inferior with respect to the others. A problem therefore arises in determining which border-detection method yields better overall results. To tackle this problem, two quantitative metrics of Weighted Performance Index (WPI) and Optimised Weighted Performance Index (OWPI) have been proposed in this thesis. The two indices are useful means of assessing performance of a given border detection method, which also facilitates comparison between different border detection methods. They take into account the existing standard metrics and have been designed such that they accommodate dermatologists' trends in defining the borders of melanoma lesions. WPI and OWPI, both are weighted average of six standard metrics of sensitivity, specificity, accuracy, border error, similarity and precision, with weights being embedded

within the six metrics. The weights are attached to the parameters of True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN) that exist within the six metrics. In WPI the weights are fixed and adjusted based on dermatologists' perspective and their emphasis on each of the TP, TN, FP and FN parameters. In OWPI, optimal weights are determined by formulating a constrained nonlinear multivariable optimisation problem, with constrains designed based on dermatologists' perspective. The optimisation is done in order to provide weights which genuinely reflect the importance of each parameter (*e.g.* TP) and metric (*e.g.* sensitivity) in the evaluation process. The optimal weights are calculated by applying trust-region and interior-point optimisation methods on five recent border detection methods using a set of 55 high-resolution dermoscopy images, and the union of four sets of dermatologist-drawn borders as the ground truth. The results are also confirmed by a grid-search procedure which results in same set of weights to be attached to the four parameters of TP, TN, FP and FN.

Furthermore, a novel, comprehensive and highly effective feature extraction method is proposed in this thesis, which involves a guided generation of categorised features, extracted from wavelet-based texture analysis, geometrical measurements, and time-series analysis. The texture-based features are extracted from a tree-structured wavelet transform, which uses four-level decomposition as opposed to three-level decomposition used in the existing wavelet-based methods. It also uses four colour channels of red, green, blue and luminance as opposed to only luminance in existing texture-based studies and applies various statistical measurements and ratios on all the nodes of the wavelet-tree. The border-based features are extracted by applying some of the standard geometrical measurements such as area and irregularity index, complemented by new ones proposed in this thesis, drawn from discussions with expert dermatologists, to analyse the overall shape of the lesion. Moreover, we have proposed a time-series model of the lesion border and have analysed it in spatial and frequency domains, using histogram, wavelet and statistical analysis. Time-series has been widely applied in statistics, signal processing, data mining, mathematical finance and economy. Yet, to the best of our knowledge it has not been previously applied to the application of melanoma diagnosis.

The next step involves feature selection and optimal integration of features, wherein the Gain Ratio feature selection method is applied. As a filter-based method, it is known to be computationally efficient, which is a vital factor considering the high dimensional feature vector constructed in the feature extraction phase of this study, and we have shown its superiority over a sample set of well-known filter-based feature selectors. Gain Ratio has been successfully applied in various applications, yet to the best of our knowledge, it has not been applied in previous melanoma-diagnosis studies. This thesis establishes that the Gain Ratio feature selection method can deliver substantial improvement to the system performance compared to using raw data (the whole of the constructed feature vector). It yields in 12% increment in the classification accuracy, as well as a considerable reduction in feature space dimension (by a factor of 214) and computation time (50 times).

Classification is the final step in the diagnosis process, wherein the optimal feature set determined in the feature selection process, are utilised to differentiate melanoma from

benign lesions. In this thesis, four advanced classes of classification techniques; namely, Support Vector Machine, Random Forrest, Logistic Model Tree and Hidden Naive Bayes are employed. Support Vector Machine has been very popular and shown to be efficient in computer-based melanoma diagnosis studies in the literature. The other three classifiers are applied to the melanoma classification problem for the first time, in this thesis. All these classifiers and more particularly the Random Forest classifier are shown to be highly effective in the classification of melanoma lesions. The proposed diagnostic system, which combines border detection, feature extraction, feature selection and classification, is applied on a set of 289 images (114 malignant, 175 benign) partitioned into train, validation and test image sets. The system achieves remarkable results of 93.21% accuracy, 90.90% sensitivity and 94.92% specificity, which compare favorably with and outperform most state-of-the-art melanoma diagnosis systems reported in the literature. An important finding is the clear advantage gained in optimally complementing texture with border and geometry features compared to using texture information only. Another interesting finding of this study is that in the melanoma classification application texture features play the dominant role and make much higher contribution than border-based features. Among border features, which include geometry and time series features, the latter has been shown to be more informative and obtains higher evaluation ranks (Gain Ratio values) than geometry in the melanoma diagnosis application.

List Of Author's Publications

Book Chapter:

- R. Garnavi, M. Aldeen, S. Finch, and G. Varigos. Global versus hybrid thresholding for border detection in dermoscopy images. In A. Elmoataz, O. Lezoray, F. Nouboud, D. Mammass, and J. Meunier, editors, Image and Signal Processing, volume 6134 of *Lecture Notes in Computer Science*, pages 531-540. Springer Berlin/Heidelberg, 2010.

Journal Papers:

- R. Garnavi, M. Aldeen, M. E. Celebi, S. Finch, and G. Varigos. Border detection in dermoscopy images using hybrid thresholding on optimised colour channels. *Computerised Medical Imaging and Graphics, Special Issue: Advances in Skin Cancer Image Analysis*, 35:105-115, 2011.
- R. Garnavi, M. Aldeen, and M. E. Celebi. Weighted performance index for objective evaluation of border detection methods in dermoscopy images. *Skin Research and Technology*, 17:35-44, 2011.
- R. Garnavi, M. Aldeen, M. E. Celebi, A. Bhuiyan, C. Dolianitis, and G. Varigos. Automatic segmentation of dermoscopy images using histogram thresholding on optimal colour channels. *International Journal of Medicine and Medical Sciences*, 1(2):126-134, 2010.

Submitted Journal Papers:

- R. Garnavi and M. Aldeen. Optimised weighted performance index for evaluation of border detection methods. *IEEE Transactions on Information Technology in Biomedicine*, submission: July 2010, revision: June 2011.
- R. Garnavi, M. Aldeen, and J. Bailey. Computer-aided Diagnosis of Melanoma Using Shape, Time Series and Wavelet-based Texture Analysis. *IEEE Transactions on Information Technology in Biomedicine*, submission: June 2011.

Conference Papers:

- R. Garnavi, M. Aldeen, and J. Bailey. Classification of melanoma lesions using wavelet-based texture analysis. In *Digital Image Computing: Techniques and Applications (DICTA2010)*, pages 75-81, 2010.

- R. Garnavi, M. Aldeen, M. E. Celebi, A. Bhuiyan, C. Dolianitis, and G. Varigos. Skin lesion segmentation using colour channel optimisation and clustering-based histogram thresholding. In *International Conference on Machine Vision, Image Processing, and Pattern Analysis (MVIPPA09)*, Academy of Science, Engineering and Technology, volume 60, pages 549-557, 2009.

Symposiums:

- R. Garnavi, M. Aldeen, G. Varigos, C. Dolianitis, G. Ross, E. Upjohn, A. Tuxen, A. Rudd. Automatic Border Detection in Dermoscopy Images of Melanoma Lesions. *Melbourne Health Research Week 2010*, Royal Melbourne Hospital, Melbourne, Australia, Jun 2010.

- R. Garnavi, M. Aldeen, G. Varigos, C. Dolianitis, G. Ross, E. Upjohn, A. Tuxen, A. Rudd. Computerised Analysis of Symmetry in Dermoscopy Images. *Melbourne Health Research Week 2010*, Royal Melbourne Hospital, Melbourne, Australia, Jun 2010.

This is to certify that

1. The thesis comprises only my original work towards the PhD,
2. due acknowledgement has been made in the text to all other material used,
3. the thesis is less than 100,000 words in length, exclusive of tables, maps, bibliographies and appendices.

Rahil Garnavi

August 2011

Acknowledgements

I would like to sincerely thank my supervisor, Dr. Mohammad Aldeen (from Department of Electrical and Electronic Engineering, University of Melbourne) for his unstinted continuous support and valuable guidance throughout the thesis.

I would also like to thank my advisory committee member, Dr. James Bailey (from Department of Computer Science and Software Engineering, University of Melbourne) for his useful inputs and discussions.

My thanks go to Dr. George Varigos and his great team (from Department of Dermatology, Royal Melbourne Hospital) for providing a good number of dermoscopy images, as well assisting us with their useful medical input at various stages of the project.

I would like to thank Dr. Emre Celebi from the Department of Computer Science, Louisiana State University, for his valuable collaborations and for acquiring some of the data used in this thesis. Also, useful advices from Dr. Les Kitchen and Dr. Alauddin Bhuiyan (from Department of Computer Science and Software Engineering, University of Melbourne) at early stages of the thesis is gratefully appreciated.

I would also like to express my gratitude to my dear friend Dr. William Withers for his valuable feedback on my writing.

The support from Department of Electrical and Electronic Engineering, University of Melbourne, and National ICT Australia (NICTA) in offering me working facilities, scholarships and travel grants is gratefully acknowledged.

Last, but definitely not the least, I would like to thank my friends and most particularly my family for all their support and encouragement.

To my beloved mom, Gity.

To my beautiful family,
Effat, Neda, Leila, Javad, Mohammad, and to my little ones Yasaman, Hooman & Athena.

And to the memory of my dear dad, Reza.

Contents

List of Figures

List of Tables

List of Abbreviations

ALM	Acral Lentiginous Melanoma
AUC	Area Under Curve
CI	Compactness Index
CFS	Correlation-based Feature Selection
CSLM	Confocal Scanning Laser Microscopy
CSM	Continuous Symmetry Measure
CT	Computed Tomography
EBCE	Element Based Conditional Entropy
FD	Fractal Dimension
GR	Gain Ratio
GRFS	Gain Ratio-based Feature Selection
GVF	Gradient Vector Flow
HNB	Hidden Naive Bayes
II	Irregularity Index
LM	Lentigo Maligna
LMM	Lentigo Maligna Melanoma
LMT	Logistic Model Tree
NM	Nodular melanoma
PI	Performance Index
OWPI	Optimised Weighted Performance Index
OOB	Out Of Bag
PET	Positron Emission Tomography
QP	Quadratic Programming
RBF	Radial Basis Function

ROC	Receiver Operating Characteristic
ROI	Region Of Interest
RF	Random Forest
RMH	Royal Melbourne Hospital
SD	Symmetry Distance
SFD	Structural Fractal Dimension
SIA	Spectrophotometric Intracutaneous Analysis
SMO	Sequential Minimal Optimisation
SR	Sigma Ratio
ST	Symmetry Transform
SVM	Support Vector Machine
SSM	Superficial Spreading Melanoma
TEM	Transmission Electron Microscopy
UV	Ultra-Violet
WPI	Weighted Performance Index

Chapter 1

Introduction

1.1 Motivation

Malignant melanoma is the deadliest form of skin cancer. According to recent figures [4], the American Cancer Society reported 68,130 new cases of melanoma in the United States in 2010, with 8,700 melanoma deaths (compared to 8,420 estimated deaths in 2009). In Australia, melanoma is the most common cancer in people aged 15–44 years. It represents 10% of all cancers and its per-capita incidence is four times higher than in Canada, the UK and the US, with more than 10,000 cases diagnosed and around 1250 deaths annually [5, 6]. The worldwide steady increase in incidence of melanoma in recent years [7], its high mortality rate and the massive respective medical cost has made its early diagnosis a continuing priority of public health.

Early diagnosis of melanoma is particularly important for two reasons [8, 9, 10]: First, the prognosis of melanoma patients depends highly on tumor thickness. If melanoma is detected at an early stage, when the tumor thickness is less than 1 mm, it is highly curable, with a 10-year survival rate between 90 and 97%. However, thickened melanoma is lethal and diagnosis at more advanced stage decreases the five-year survival rate to 10–15%. Second, melanoma is a skin cancer that in a majority of cases is localised to the skin and is therefore detectable by simple examination. However, it is more likely to metastasise and spread to other organs compared to other skin tumours. It has been shown [7] that early detection campaign may lead to mortality reduction more quickly than sun protection efforts. Nevertheless, the early diagnosing of melanoma is not always trivial even for experienced dermatologists and more particularly for primary care physicians and less-experienced dermatologists [11, 12, 13].

Much effort has been made in the last two decades to improve the clinical diagnosis of melanoma. These include alternative imaging technologies such as dermoscopy [11] and several diagnostic algorithms such as pattern analysis [14], the ABCD rule of dermoscopy [15], Menzies method [16], 7-point checklist [17], and the CASH algorithm [18].

Meta-analysis of studies conducted before 2001 showed that using the mentioned diagnstic algorithms along with dermoscopy improves the diagnosis of melanoma compared to simple naked-eye examination or traditional clinical examinations by 5–30%, depending on the type of skin lesion and the experience of the dermatologist [7, 11]. However, it has been demonstrated that dermoscopy can reduce the diagnostic accuracy in the hands of inexperienced dermatologists [19] or sometimes make no improvements compared with clinical examinations [7]. Furthermore, clinical diagnosis of melanoma is inherently subjective, highly reliant on clinical experience and visual perception of individuals, and its accuracy has been an issue of concern, especially with equivocal pigmented lesions [11]. Despite the use of dermoscopy, the accuracy of expert dermatologists in diagnosing melanoma is still estimated to be about 75–85% [20]. This issue highlights the demand for receiving a non-invasive *in vivo* second opinion, which increases the accuracy of diagnosis as well as decreasing the number of false excision of benign lesions. The former leads to saving more lives and the latter lessons the financial and emotional costs imposed on patients.

Furthermore, it is highly desirable and advantageous for dermatologist to have a diagnostic system which could provide quantitative and objective evaluation of the skin lesion, versus the subjective clinical assessment. Such a system would allow for reproducible diagnosis by diminishing the inter-observer and intra-observer variabilities that could be found in dermatologists' diagnosis of lesions, especially the equivocal ones. It can also automate the analysis, and thereby reduces the amount of repetitive and tedious tasks to be done by physicians.

1.2 Research Objectives

Due to enhancements in skin imaging technology and image processing techniques in recent decades, there has been a significant increase in interest in the area of computer-aided diagnosis of melanoma. The aim of such systems is to remove subjectivity and uncertainty from the diagnostic process, provide a reliable second opinion to dermatologists and overcome the low reproducibility [21] found in clinical diagnosis. Through digital representation of dermoscopy images and in-depth mathematical and statistical analysis of colour and intensity of the lesion, it is possible that a computer-based diagnostic system recognises features not detectable by human eyes, and therefore improves the diagnostic accuracy over even well trained and experienced clinicians [22].

Several studies have developed automated and semi-automated computer-based systems to analyse skin lesions and diagnose melanoma [12]. The potential benefits of the computerised solution to melanoma application and a number of other medical problems are incalculable. However, the difficulties involved are also very great, and new contributions are thus greatly appreciated and eagerly sought [23]. Moreover, it is widely acknowledged that much higher accuracy is required for computer-based systems to be reliable and trustworthy enough to be adopted routinely in the diagnostic process by clinicians [24, 25, 26, 27, 28]. The research reported in this thesis is an attempt to improve some of the existing algorithms and develop new techniques to facilitate accurate, fast and

reliable computer-based diagnosis of melanoma.

A computer-aided diagnosis of melanoma generally comprises several components: image acquisition, image segmentation or border detection, feature extraction, feature selection, and classification, as illustrated in Figure 1.1.

Image acquisition is a basic yet vital component for a computer-aided diagnosis system of melanoma. In order to develop a reliable computer-based diagnostic system, it is required to establish a unified systematic protocol under which a reasonably large collection of good quality and well-illuminated skin images is acquired. In this study, we have engaged a group of dermatologists, with different levels of expertise, from Royal Melbourne Hospital (RMH), Melbourne, Australia, to establish such a unified and consistent image acquisition procedure. In this respect, every reasonable effort has been made to capture images under unified zooming and lighting conditions so that the acquired images serve as good-quality input data to our diagnosis system. In the early stages of this research, high resolution dermoscopy images were taken by professional photographers from Medical Illustration, Royal Melbourne Hospital, using a Canon EOS 450D camera. However, at some later stages of the work, due to time limitation and lack of sufficient data, we sought to obtain images from other sources [29], whose contributions are gratefully appreciated. It is worth commenting that the collection of images from different sources has been advantageous as training a diagnostic system with images collected from various sources and obtaining highly accurate results is an indication that the proposed diagnosis system provides a robust and reliable performance in real-world application. Our preference of using dermoscopy images over clinical images stems from the well established facts that: (i) dermoscopy imaging reveals the internal structure of the skin lesion which is not visible through naked eye [7, 11, 20], (ii) using dermoscopy images improve the diagnostic accuracy by dermatologists [30], and (iii) computer systems using dermoscopy images obtain higher diagnostic accuracy than those using clinical images [31].

The purpose of the *image segmentation* or *border detection* component is to separate the region covered by the lesion from the surrounding normal skin. The output of this procedure is a binary image with lesion pixels in black and background pixels in white, or alternatively a vector encompassing the coordinates of pixels along the boundary of the lesion. The determined lesion is the region of interest (ROI) to be analysed in the subsequent components of the diagnostic system. After design and development of the border detection algorithm proposed in this study, and testing the method on a sufficiently large set of dermoscopy images taken from RMH, another data acquisition challenge at this stage is to obtain a gold standard. The gold standard is a reference segmented image required to validate and assess the performance of the proposed border detection method. It was obtained for each image by holding a number of meetings with experienced dermatologists as well as dermatology registrars[1] from the RMH and have them manually draw borders for the set of acquired images, using a Wacom Intuos A4 size Tablet PC.

[1]A Specialist Registrar or SpR is a doctor in the United Kingdom, Republic of Ireland, and Australia who is receiving advanced training in a specialist field of medicine in order eventually to become a consultant. In surgery, it is also referred to as Higher Surgical Trainee or HST.

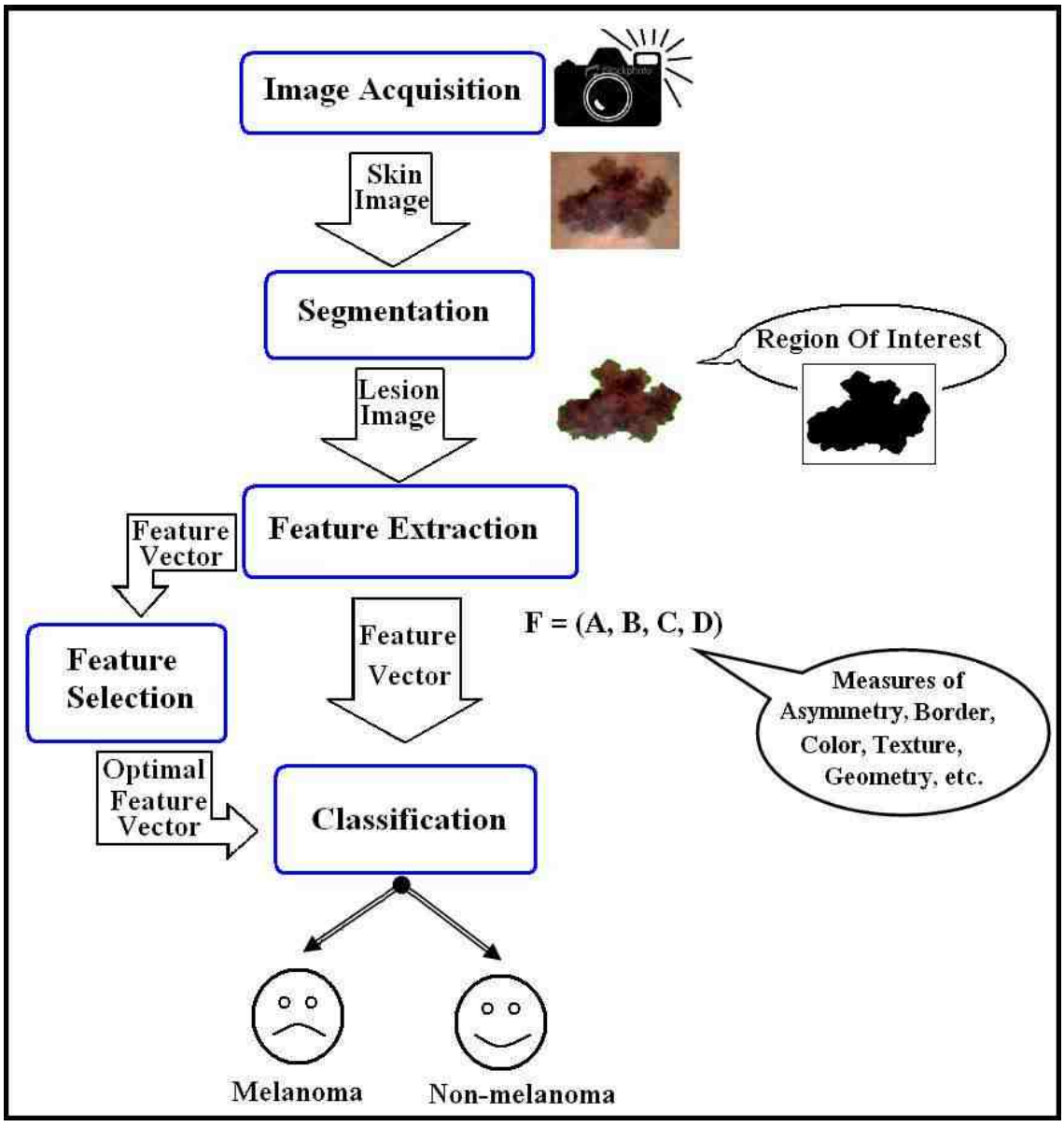

Figure 1.1: Schematic diagram for a computer-aided diagnosis of melanoma.

The role of the *feature extraction* component is to extract various features from the image of a skin lesion. These features are later utilised in the classification component to decide whether the lesion is malignant or benign. In most computer vision systems, an intermediate step is embedded between feature extraction and classification; *i.e.*, *feature selection*. The purpose of feature selection is to select an optimal subset of features which provides the highest discriminating power when employed by the *classification* component. In order to train the system and examine the performance of the proposed methods, the images used for these three steps are required to have histopathological results which are determined by biopsy or excision.

In the development of a computer-aided diagnosis of melanoma, as we proceed through each of the above-mentioned stages of the diagnostic system, the following issues arise:

1. A dermoscopy image of the skin generally comprises two main regions; a mole (the skin lesion) and the surrounding skin. Our region of interest (ROI) in this application is the skin lesion, over which the subsequent components of the diagnostic system would operate and analyse. The issue that arises here is what is the best technique to extract the ROI from a given dermoscopy image? This issue is addressed in the segmentation (border detection) component of the diagnostic system proposed in this thesis.

2. Once the borders of the lesions are drawn by using the border-detection method developed in the previous step, the question that remains is how can we objectively and quantitatively assess the correctness of the determined borders in reference to a given gold standard? The accuracy of the detected border is essential for accurate implementation of the subsequent parts of the system, *i.e.*, feature extraction and classification.

3. After performing the segmentation and determining the region of interest (the lesion) on each dermoscopy image of the skin and validating its correctness, we need to look into the lesion and define the indicative features which best characterise the skin lesion. These features, which are generally derived from clinical criteria used by dermatologists in their diagnosis practice, then need to be translated into computer language. In other words, we need to find mathematical (objective) measurements that capture the visual cues physician adopt in their visual (subjective) assessments. This issue is addressed in the feature extraction component of the diagnostic system proposed in this study.

4. The diagnostic features determined in the previous step, although useful, can sometimes be redundant and have counter-effects when combined. In other words, it is possible to reduce or eliminate redundancies to arrive at an optimal set of features which produces the most accurate results. This issue is addressed in the feature selection component of the diagnostic system proposed in this study.

5. After defining the optimal set of features for characterising skin lesions we need to

find a classifier which can effectively classify the given lesions into malignant and benign classes. Furthermore, we need to quantitatively assess the performance of the classification scheme. This issue is addressed in the classification component of the diagnostic system proposed in this thesis.

As outlined above, this thesis aims to develop a fast and accurate border detection method which produces borders very close to gold standards (*i.e.*, those drawn by dermatologists). The ultimate goal is to provide a practical, simple and easily reproducible solution to the problem of border detection in dermoscopy images, and avoid complications as far as possible. This may be achieved by benefiting from basic yet useful image processing techniques and integrating them in a systematic way to obtain highly accurate results. In doing so, we have proposed a thresholding-based segmentation algorithm and have demonstrated that this approach, even though fairly straightforward, outperforms three state-of-the-art border detection methods in dermoscopy image (issue 1). Moreover, we have developed a systematic approach to quantify the value and usefulness of a detected border by an automated border detection method. Such objective evaluation provides a basis for quantitative comparison between different border detection methods (issue 2).

Furthermore, after extensive meetings with a number of dermatologists and observing their approach to clinical assessment of different lesions and diagnosing melanoma lesions in various stages, we have generated guided lists of various features which can capture as much information about the lesion as possible (issue 3). Other objectives of this study include determining the optimal subset of features with the highest contributions in the recognition of melanoma (issue 4). This has been achieved by applying an entropy-based feature selection scheme. Lastly, through the application of four advanced classification techniques, this work achieves a highly accurate recognition of melanoma lesions which can classify dermoscopy images with high measures of sensitivity and specificity (issue 5). In the next section, thesis structure, we provide an insight into how each of the above issues has been implemented.

1.3 Thesis Structure

The rest of the thesis is organised as follows:

Chapter 2 - "Background and Literature Review" provides background information about melanoma and its stages, dermoscopy technology and clinical algorithms using which dermatologists make diagnose. This is followed by presenting a survey of existing computer-based melanoma diagnosis systems covering all main components of such systems such as segmentation, feature extraction, feature selection and classification.

Chapter 3 - "Segmentation or Border Detection" presents two automated methods for skin lesion segmentation (border detection) in dermoscopy images; Global and Hybrid method, the latter being an enhancement of the former. As a perquisite for performing the segmentation, this chapter applies colour-space analysis to determine the most effective

and discriminative colour channels for detecting borders in dermoscopy images. The chapter also includes comparative studies between each of the proposed segmentation methods (Global and Hybrid) and the state-of-the-art border detection methods. Moreover, an analytical framework including statistical analysis, optimisation and cross validation, is provided in this chapter to determine the optimal parameters for the proposed Hybrid border detection method. The two proposed border detection methods; Global and Hybrid, are then compared through a newly introduced evaluation metric, Performance Index. The last two sections of the chapter explore the objective evaluation of the border detection results in dermoscopy image analysis, and provide a comprehensive evaluation metric which reflects dermatologists' perspective in the melanoma recognition process. This is done by introducing two performance indices of Weighted Performance Index and Optimised Weighted Performance Index; the former directly uses expert opinion, whereas the latter formulate a constrained nonlinear multivariable optimisation problem embedding the dermatologists perspective into constraints of optimisation.

Chapter 4 - "Feature Extraction, Selection and Classification" presents a new, comprehensive and highly effective feature extraction method which combines features derived from textural, border and geometry properties of the melanoma lesion. The textural features are derived from tree-structure wavelet-decomposition; the border features are derived from constructing a time-series model of the lesion border and analysing it in spatial and frequency domains; and the geometry features are derived from shape indices. The chapter then discusses the process of feature optimisation through applying the computationally efficient Gain-Ratio feature selection method. Moreover, the application of four different classes of techniques for classification of melanoma lesion is discussed; Random Forest, Logistic Model Tree and Hidden Naive Bayes, which have not been previously used in existing computer-based melanoma recognition systems, and the Support Vector Machine classifier, which is very popular in melanoma related studies. Lastly, various streams of feature optimisation are conducted and the results are compared.

Chapter 5 - "Summary and Conclusion" concludes this thesis with a review of our contributions along with considerations of future work in the area of computerised melanoma recognition and dermoscopy image analysis.

1.4 Research Contributions

1. Colour-channel optimisation for border detection in dermoscopy images.

2. Fully automated Global border-detection method in dermoscopy images based on colour-space analysis and global histogram thresholding which exhibits high performance in detecting the borders of melanoma lesions.

3. Comparative study with two other state-of-the-art skin lesion segmentation methods which shows the superiority of the proposed Global method.

4. Fully automated Hybrid border-detection method based on adaptive histogram thresh-

olding which is demonstrated to be highly effective in detecting the border of lesions in dermoscopy images. It improves the Global border-detection method as shown via statistical analysis.

5. Analytical framework including statistical analysis, optimisation and cross validation, in order to determine the optimal parameters for the proposed Hybrid border detection method.

6. Comparative study with three other state-of-the-art skin lesion segmentation methods which shows the superiority of the proposed Hybrid method.

7. Objective evaluation of border-detection methods in dermoscopy images by proposing evaluation metrics of Performance Index, Weighted Performance Index and Optimised Weighted Performance Index.

8. Guided generation of ternary list of features based on texture, geometry and time-series information extracted from the interior pattern, overall shape and boundary of the lesion.

   (a) The use of four-level wavelet decomposition, incorporating four colour channels as well as applying various statistical measurement and ratios on all nodes of the fully extended wavelet tree for texture-based feature extraction. The proposed texture-based method is shown to overcome three other wavelet-based methods.

   (b) The use of time-series analysis in spatial and frequency domains and new geometrical measures for border-based feature extraction.

9. The use of the computationally efficient Gain Ratio method for feature selection.

10. The use of four different highly efficient classifiers; Random Forest, Logistic Model Tree and Hidden Naive Bayes, which have not been previously used in existing computer-based melanoma recognition systems, and the Support Vector Machine classifier, which is very popular in melanoma related studies.

11. The use of various effective ways of optimising the feature vector.

12. The proposed system achieves remarkable results of 93.21% accuracy, 90.90% sensitivity and 94.92% specificity, which compare favourably with and outperform most state-of-the-art melanoma diagnosis systems as reported in the literature.

Chapter 2

Background and Literature Review

In this chapter we provide a concise background about the skin structure, formation of melanoma cancer and the clinical diagnostic approach taken by dermatologists to detect it. This is followed by a synopsis of existing computer-aided diagnostic systems of melanoma, including image processing and machine learning based methods used in those systems.

2.1 Skin Structure and Melanoma

Human skin comprises three principal layers; namely, stratum corneum, epidermis, and dermis (including papillary dermis and reticular dermis, as shown in Figures 2.1 and 2.2). Each layer has a particular function and optical properties; (i) Stratum corneum, the top layer, is a protective layer of keratin-impregnated cells, and varies in thickness. This layer is optically neutral and only has a scattering effect to the incident light on the skin. (ii) The epidermis layer is composed of connective tissue. It also contains melanocytes in its basal layer. Melanocytes produce the skin pigment, melanin, which provides the tan or brown colour of the skin. Melanocytes act as a filter and protect the skin from harmful Ultra-Violet (UV) sun rays by increasing the production of melanin. Hence, the epidermis layer strongly absorbs blue and UV light, and the extent of absorption depends on concentration of melanocytes. Within the epidermal layer there is very little scattering, with the small amount that occurs being forward directed. The result is that all light not absorbed by melanin can be considered to pass into the dermis. (iii) Dermis consists of collagen fibres, sensors, receptors, blood vessels and nerve ends. It has two sub layers: papillary dermis and reticular dermis. Papillary dermis with collagen fibres scatters backward any incident light, and more particularly the red spectrum, toward the surface of the skin. As infrared is not absorbed by melanin and blood, this part of the spectrum is best for assessing the thickness of papillary dermis. In the reticular dermis, the large size of collagen fibres causes forward scattering, and any light which gets to this layer is passed deeper into the

skin and does not contribute to the spectrum remitted from the skin [12, 32].

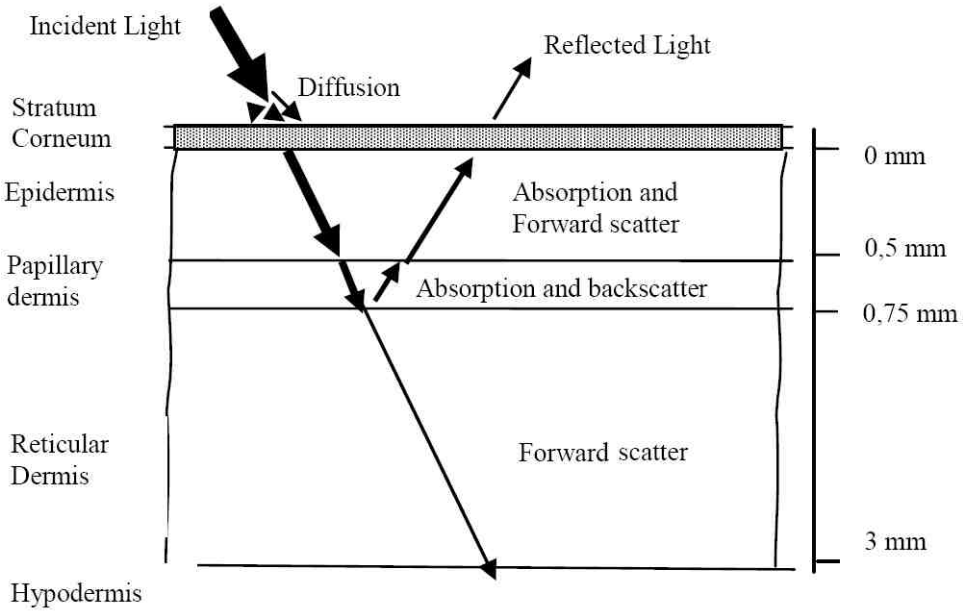

Figure 2.1: Multi-layered skin model [32]

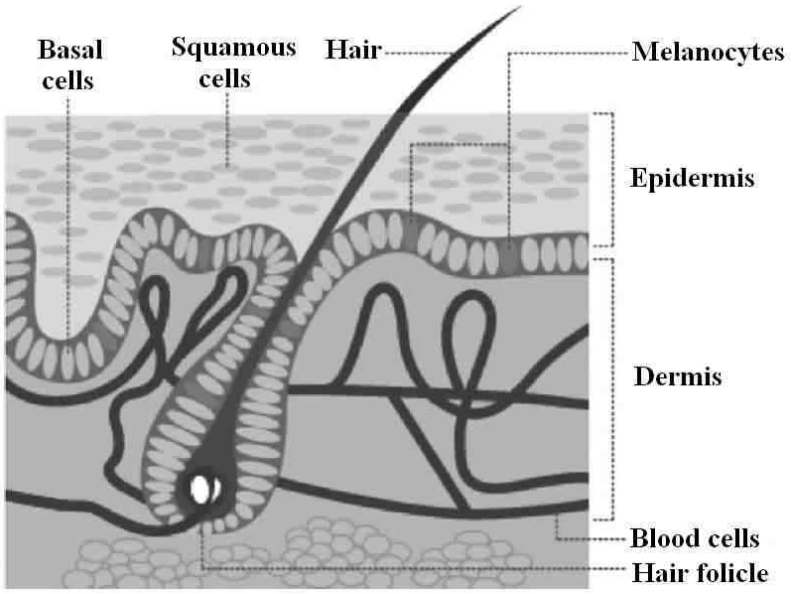

Figure 2.2: Normal skin lesions and main components [12]

Pigmented skin lesions appear as patches of darker colour on the skin, caused by excessive concentration of melanin. In benign lesions such as common nevi, melanin deposits

are normally found in the Epidermis layer. Malignant change of melanocytes results in cutaneous melanoma, wherein the melanocytes reproduce melanin at an abnormally high rate (see Figure 2.3).

Melanoma is the deadliest type of skin cancer. Its incidence and mortality rates have increased steadily worldwide over the past several decades [7, 33, 34]. The melanoma survival rate depends highly on its stage and thickness. If the abnormally produced melanin remains in the epidermis, melanoma is called *in-situ*. At this stage, it is not life-threatening and early detection with immediate surgical excision of the lesion result in a good prognosis [11, 35]. However, the early diagnosis of melanoma is not very straightforward and there are always equivocal cases; the optical and visual properties of *in-situ* melanoma resemble to those of normal but highly pigmented skin. When malignant melanocytes penetrate into the Dermis, they leave melanin deposits there, thus resulting in changes in skin colouration. This leads to advanced and thickened melanoma which may be incurable. Table 2.1 shows different stages of melanoma.

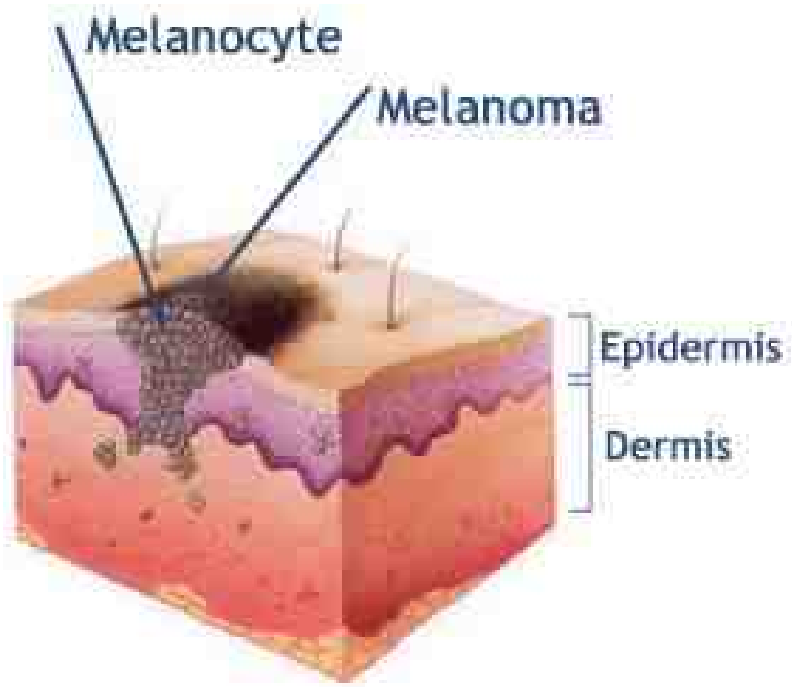

Figure 2.3: Melanocytes and Melanoma on skin [36]

According to microscopic characterisations, there are four main types of melanoma; namely, Superficial spreading melanoma (SSM), Nodular melanoma (NM), Lentigo maligna/lentigo maligna melanoma (LM/LMM), and Acral lentiginous melanoma (ALM). SSM, which occurs mostly in middle-aged people, is considered the most common type of melanoma, and accounts for about 70% of diagnosed cases [33]. However, a recent study challenged this notion and found lentigo maligna to be the most common subtype

Table 2.1: Various stages of melanoma.

Stage	Description	Survival Rate
Stage 0	Cancer cells are found only in the epidermis, the outer layer of the skin.	virtually 100%
Stage I	Cancer has spread to the top of the dermis, but not to lymph nodes. The tumor is less than 1.5 millimetres thick.	over 90%
Stage II	Cancer has spread to the bottom part of the dermis, but not to tissue below the skin or to lymph nodes. The tumor is larger, with a Breslow thickness ranging from 1.5 millimeters to 4 millimeters.	about 80%
Stage III	Cancer has spread to lymph nodes near the tumor site.	about 50%
Stage IV	Cancer has spread beyond nearby lymph nodes to internal organs.	about 20%

of melanoma among the cases studied [37].

Although the presence of melanin in the dermis is the most significant sign of melanoma, it cannot be used as the only diagnosis measure as *in-situ* melanomas have no dermal melanin. Furthermore, some benign nevi have dermal deposits, although their spatial patterns tend to be more regular than in melanoma. In clinical diagnosis of malignant melanoma the configuration of gross morphologic features within the lesion is assessed. These features include shape, size, colouration, border and symmetry of the pigmented lesion. The initial pattern of all types of melanoma, except for nodular melanoma, is radial growth. Melanomas are diagnosed mostly when the diameter exceed 10 mm, while the ones with diameter less than 6 mm are less frequently diagnosed. The most suspicious sign of melanoma is changes in size and colour of the lesion. Both darkening and lightening (loss of pigmentation) are alarming signs. Moreover, itching, tenderness, bleeding, and elevation are symptoms of a more advanced melanoma. Other signs, some of which can be indicative of melanoma in-situ, are thickening of the collagen fibres in the Papillary dermis (fibrosis), increased blood supply at the lesion periphery (erythematic reaction), and lack of blood within the lesion in the areas destroyed by cancer. The colours associated with skin, which has melanin deposits in the dermis, normally show characteristic hues not found in any other skin conditions, which provides an important diagnostic cue for clinicians. If the clinical visual examination indicates a suspicion of skin cancer, histology is required to make precise diagnosis [12, 33, 38]. In traditional clinical practice, dermatologists diagnose melanoma accurately in 65%–80% of cases [39, 40].

2.2 Enhancement in clinical diagnosis of melanoma

Owing to the urgency and importance of early and accurate diagnosis of melanoma, a lot of effort has been made to enhance the clinical diagnosis of melanoma in the last two decades. In doing so, dermoscopy technique and digital monitoring, as well as several clinical approaches such as such as pattern analysis [14], ABCD rule of dermoscopy [15], Menzies method [16], 7-point checklist [17], and the CASH algorithm [18] have been introduced. More details about dermoscopy and clinical algorithms are presented in the following.

2.2.1 Dermoscopy

Dermoscopy (also known as epiluminescence microscopy or ELM, dermatoscopy, and surface microscopy) introduced in 1987, is a non-invasive *in vivo* clinical examination method. This useful technique allows for a magnified and clear visualisation of the morphological structures of the skin not visible to the naked eye. It is based on application of a liquid, *i.e.*, gel or oil or alcohol, to the skin and analysing the pigmented lesion with a hand-held device, called dermatoscope. The liquid reduces the light reflection and makes the epidermis transparent, rendering the underlying structures visible. In expert hands and trained general practitioners, dermoscopy has been shown to improve both the sensitivity and specificity[1] for the diagnosis of melanoma, elevating the diagnostic accuracy by 5%–30%, which is also reflected in lower benign melanoma excision ratios and decreased excision rates [30]. The improvement rate in diagnosis depends on the type of skin lesion and the experience of the dermatologist [11]. Due to evidence of misdiagnosis in naked-eye examinations, Menzies [30] suggests that dermoscopy should be considered as the minimum requirement for general dermatologists to clinically assess pigmented skin lesions. Besides, the meta-analysis of all studies conducted before 2001 shows that higher performance is obtained by dermoscopy employed by expert physicians, compared with traditional clinical examination. However, literature reveals that using dermoscopy is less common in the USA compared with Europe and Australia [7, 11, 20, 30].

Applying dermoscopy discloses different characteristics of the lesion, which are not detectable by the naked eye. colours, blood vessels and various dermoscopic structures are among these characteristics. Dermoscopic structures include pigment network, dots, globules, branched streaks, streaks, structureless areas, blotches, regression, blue-white veil, milia-like cysts, comedo-like openings, fingerprint-like structures, moth-eaten border, fissure ridges or brain or leaf-like areas, spoke-wheel-like structures, large blue-grey ovoid nests, and multiple blue-grey globules [41]. Figures 2.4 and 2.5 illustrate some of these dermoscopic patterns. According to the agreement of the Consensus Net Meeting on

[1]Sensitivity and specificity are statistical measures of the performance of a binary classification test. Sensitivity (also called recall rate) measures the proportion of actual positives which are correctly identified as such (e.g. the percentage of melanoma lesions which are correctly identified as being malignant). Specificity measures the proportion of negatives which are correctly identified (e.g. the percentage of benign lesions which are correctly identified as being normal). See equations 3.9 and 3.10.

Dermoscopy [20], a two-step procedure was defined to classify pigmented skin lesions:

1. Distinguishing between melanocytic and non-melanocytic lesions.
2. For those labelled as melanocytic, deciding the benignity or malignancy of the lesion.

As the title of this thesis implies, the second step is of interest to this study. To differentiate between malignant and benign lesions, there exist different dermoscopic algorithms; namely, pattern analysis [14], the ABCD rule of dermoscopy [15], the Menzies method [16], the 7-points checklist [17], and the CASH algorithm [18]. A brief outline of each is given in the following.

2.2.2 Pattern Analysis

The pattern analysis method was reportedly established by Pehamberger *et al.* [14] in 1987. It differentiates between benign and malignant lesions by analysing the morphological features within the lesion. These features include specific patterns, colours, pigmentation intensity, configuration and regularity of both the margin and the surface of the pigmented skin lesion. It qualitatively evaluates the defined dermoscopic criteria, and is claimed to exhibit the highest diagnostic accuracy compared to other clinical algorithms (such as ABCD rule of dermoscopy, Menzies method, 7-points checklist, and CASH algorithm), when applied by experienced dermatologists [7, 43]. Over a virtual gathering of 40 experienced dermatologists analysing 108 pigmented skin lesions, the consensus diagnosis (the diagnosis agreed upon by over 50% of observers) showed a sensitivity of 100% and a specificity of 88% [20].

2.2.3 ABCD Rule of Dermoscopy

The ABCD rule of dermatoscopy [15] is a semi-quantitative scoring system which classifies the pigmented skin lesions into benign, suspicious and malignant classes. To analyse the lesion the four features; asymmetry (A), border (B), colour (C), and differential structure (D) are assessed over the lesion. The features' values are then multiplied by particular weighting factors and finally summed up to result in the Total Dermatoscopy Score (TDS), given in Equation 2.1. A TDS value of less than 4.75 represents a benign lesion, a value between 4.75 and 5.45 indicates a suspicious lesion and a value greater than 5.45 is highly suspicious for melanoma.

$$TDS = [(A \times 1.3) + (B \times 0.1) + (C \times 0.5) + (D \times 0.5)] \tag{2.1}$$

Asymmetry

In the ABCD rules of dermatoscopy [21], to quantify the asymmetry, the lesion is bilaterally segmented by two perpendicular axes positioned such that the resultant asymmetry score

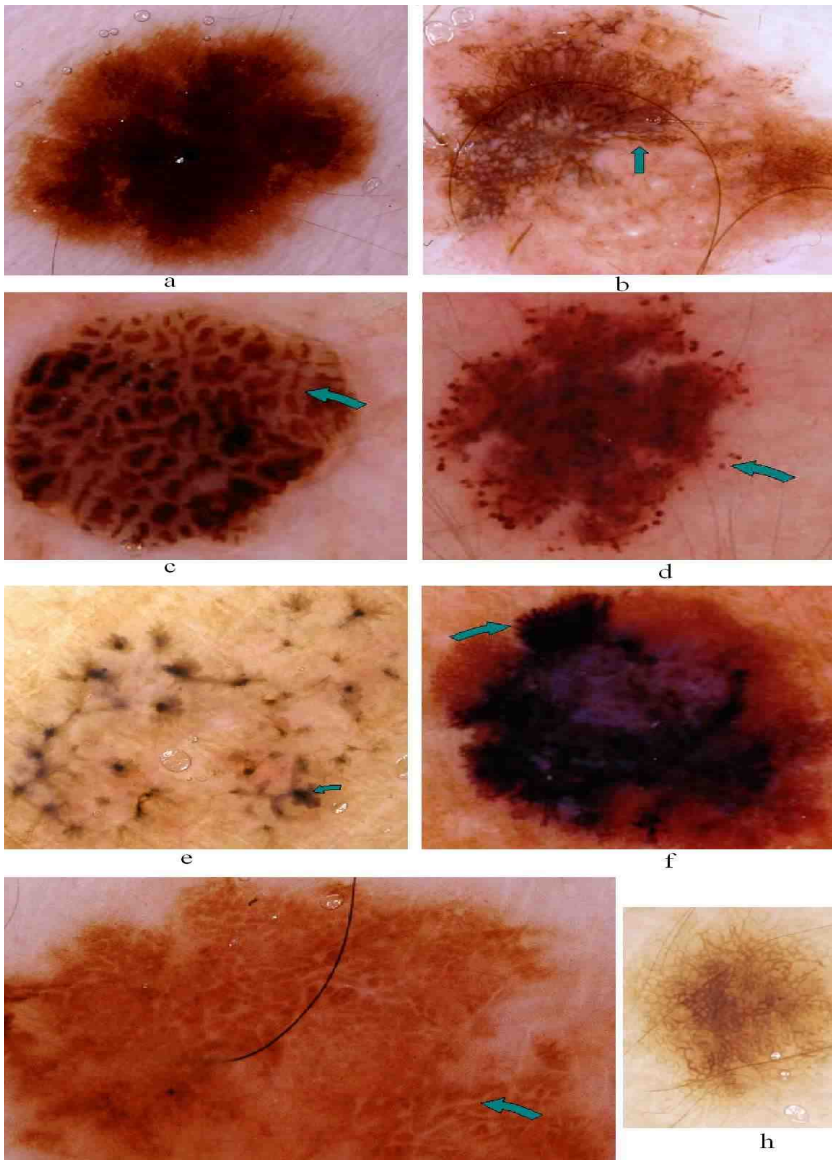

Figure 2.4: Dermoscopic structures: (a) Blotchy area, (b) Branched streaks, (c) Fissures, (d) Globules, (e) Leaf-like structures, (f) Radial streaming, (g) Negative network, (h) Pigmented network [42].

is minimum. The asymmetry score is calculated according to the contour (shape), colour distribution and internal structure of the lesion on either side of each axis. The asymmetry score is 2, if both axes show asymmetry. Asymmetry on only one axis results in score of 1 and if there is no asymmetry on either of the axes, the score is 0. Stolz *et al.* [21] note that in their experiment 96% of malignant melanoma cases had an asymmetry score of

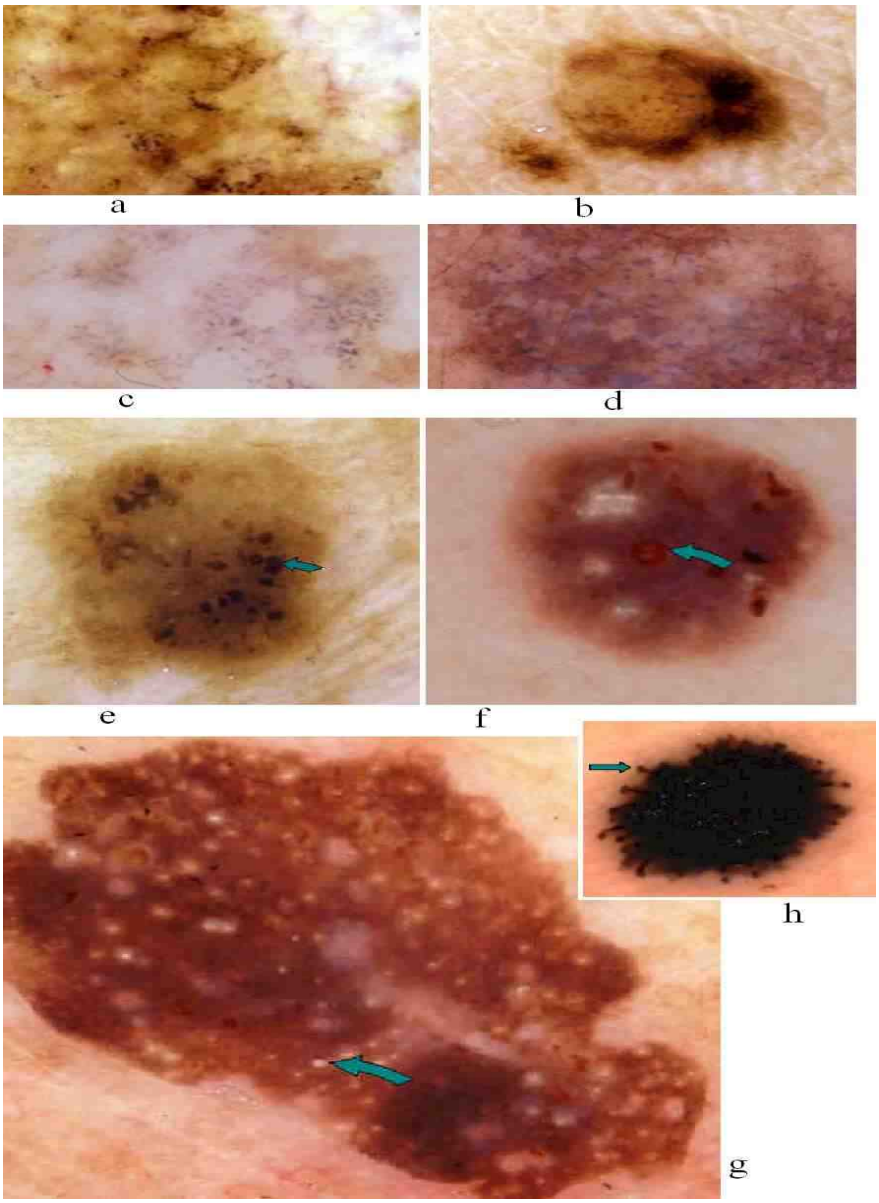

Figure 2.5: Dermoscopic structures: (a) Black dots, (b) Brown dots, (c) Grey granules, (d) Blue granules, (e) and (f) Comedone-like structures, (g) Milia-like cysts, (h) Pseudopods [42].

2, whereas only 24.2% of benign melanocyte nevi had such score. Since equivocal lesions have symmetrical shapes, incorporating colour and structural asymmetry is very crucial when scoring the symmetry of a lesion.

Border

Stolz *et al.* [21] evaluate the border score by investigating the sharpness of the lesion border. To do this, the lesion is divided into eight sections. In each section, a sharp, abrupt cut-off of pigment pattern at the periphery of the lesion yields a score of 1, whereas a gradual, indistinct cut-off within gets a score of 0. Thus, the maximum border score is 8, and the minimum score is 0. Experiments show that the border score in melanocytic nevi is very low (0 in 60% and 4 only in 10%), whereas in malignant melanomas, it is preponderantly between 3 and 8. However, Stolz *et al.* [21] recognise the low reproducibility of border evaluation and notice the difficulty in determining whether the border cut-off is abrupt or not.

Colour

With the use of a dermatoscope, a wider colour spectrum of melanocytic lesions is detectable, than with the naked eye. According to the ABCD rules of dermatoscopy [21], to determine the colour score, a total number of six different colours; namely, white, red, light-brown, dark-brown, blue-grey, and black, are considered. White colour should be selected only if the area is lighter than the neighbouring skin. With the presence of all of these six colours, the maximum colour score is 6, while the minimum score is 1. Their experiments show 56% of melanocytic nevi with 2 colours, 29% with 3 colours, and 10% with more than 3 colours. On the other hand, malignant melanomas manifest 3 or more colours in 85% of the cases and 5 or 6 colours 40% of the time.

Differential Structure

Five structural features have been nominated by Stolz *et al.* [21] to evaluate the differential structures (also known as dermoscopic structure): namely, a pigment network, structureless or homogeneous areas, streaks, dots, and globules. Presence of more structures increases the probability of malignancy of the lesion. In more than 90% of melanocytic nevi, 3 or fewer structural components were observed, while in more than 70% of malignant melanomas, 4 or more of them were found. Structural components can be detected via dermatoscope by expert dermatologist but rarely by the naked eye. To calculate the D score in the ABCD rule, structureless or homogenous areas must be larger than 10% of the lesion. Streaks and dots are considered only when more than two are observed. Finally, globules have to be more than one to be counted as a structural component. Thus the maximum and minimum values for D score are 5 and 1, respectively.

Stolz *et al.* [15] recognise several differences between their proposed ABCD rule and Friedman's ABCD rule [44]. For example, Friedman's ABCD rule emphasises shape symmetry, while the main focus of Stolz's algorithm is on colour and structure symmetry. Moreover, Friedman investigates the border irregularity, while Stolz analyses the abrupt pigment pattern at the border. Finally, while the D feature is differential structure in

Stolz's ABCD rule, it is diameter in Friedman's method. There is also another interpretation for D parameter of ABCD rules of dermatoscopy, which is diameter greater than 6 mm [45].

The ABCD algorithm showed consensus sensitivity and specificity of about 96% and 70% in the above-mentioned virtual consensus meeting [20]. The classic ABCD acronym for diagnosis of melanoma has recently been revised (ABCDE [46]), where *Evolving* has been added to the ABCD acronym. That is because changes of size, shape, shades of colour, symptoms (itch, tenderness), and surface (*e.g.*, bleeding, crusting, scaling) can allow detection of smaller, nodular, and morphologically banal melanoma [30].

2.2.4 Menzies Method

In the Menzies method [16], two groups of negative and positive features are defined. Negative features include presence of a single colour, as well as axial symmetry of pigmentation, which refers to symmetry of pattern around any axis through the center of lesion. There are nine positive features; namely, blue-white veil, multiple brown dots, pseudopods, radial streaming, scar-like depigmentation, peripheral black dots/globules, multiple colours (five or six), multiple blue/grey dots, and broadened network. For diagnosing melanoma, both negative features must be missing and at least one positive feature must be found. In this approach, colours identified within the lesion include black, grey, blue, dark brown, tan and red. The Menzies method obtained 96% and 73% consensus sensitivity and specificity [20].

2.2.5 7-point Checklist

In the 7-point checklist method [17], three major criteria (atypical pigment network, blue-white veil, atypical vascular pattern), each with a score of 2 are employed. As well, four minor criteria (irregular streaks, irregular pigmentation, irregular dots/globules, regression structures), each with a score of 1, are identified. A minimum total score of 3 is required for the lesion to be diagnosed as melanoma. The vascular pattern as a criterion to investigate the malignancy of the lesion was reportedly suggested for the first time by the 7-point checklist algorithm [41]. The consensus sensitivity and specificity for the 7-point checklist were reported as 96% and 73%, respectively [20].

2.2.6 CASH Algorithm

CASH is a simplified version of the pattern analysis algorithm and is aimed to be suitable for less experienced examiners. It provides a scoring system, similar to that of ABCD, which semi-quantitatively evaluates colour (C), architecture (A), symmetry (S) and homogeneity (H) of the skin lesion. Its accuracy is considered to be comparable to that of other algorithms in distinguishing melanoma from benign lesion [18].

2.3 Computer-based Diagnosis of Melanoma

Even with the use of technologies such as dermoscopy and applying the above-mentioned algorithms, clinical diagnosis is still challenging and its accuracy is considered to be limited, especially with equivocal pigmented lesions [11]. Despite the use of dermoscopy, the accuracy of expert dermatologists in diagnosing melanoma is estimated to be about 75-84% [20]. Moreover, clinical diagnosis of melanoma is inherently subjective and suffers from inter- and intra-observer variabilities. This issue highlights the demand for receiving an *in vivo* second opinion, which (i) increases the accuracy of diagnosis, thus saving more lives, (ii) decreases the number of false excision of benign lesions, hence reducing the medical and emotional costs imposed on individuals by unnecessary surgeries. Image processing and computer vision techniques have been applied in an attempt to solve this problem since 1985. A computer-aided diagnosis of melanoma provides quantitative and objective evaluation of the skin lesion, versus the subjective clinical assessment. It allows for reproducible diagnosis by diminishing the inter-observer and intra-observer variabilities that could be found in dermatologists' examinations. It also automates the analysis, and thereby reduces the amount of repetitive and tedious tasks to be done by clinicians.

Due to enhancements in skin imaging technology and image processing techniques in the recent years, and owing to the repetitive patterns in skin lesions, which makes it an intriguing problem to work on, there has been a significant increase in interest in development of computer-based diagnostic systems for melanoma. Various groups around the world have been trying to develop effective systems and devices to assist in the evaluation of pigmented skin lesions, some of which include:

SolarScan [47] developed by Polartechnics Ltd, Sydney, Australia, analyses dermoscopy images of the pigmented skin lesion according to 103 variables based on colour, pattern, and geometry of the lesion. The system was tested [48] on an image set of 2430 lesions, among which 382 were melanomas with median Breslow thickness of 0.36 mm. Seven specialist referral centres and two general practice skin cancer clinics from three continents were engaged in the study and the SolarScan diagnosis results were compared with those of 13 clinicians, including three dermoscopy experts, four dermatologists, three trainee dermatologists, and three general practitioners. Applying statistical analysis and calculating the AUC (area under the ROC[2] curve), the system obtained a sensitivity of 86%–96% and specificity of 64%-72%.

DermoGenius-Ultra [49], wherein the ABC features were calculated to approximate the A, B and C parameters of the ABCD clinical algorithm, was developed by LINOS Photonics Inc. Instead of the D parameter in the ABCD rule of dermoscopy which identifies the dermoscopic structures, in this system D was to approximate the degree of heterogeneity in the lesion. The system was tested on 187 patients at risk for melanoma, resulted in identifying 52 lesions to be removed, which were labelled as unsuspicious in the clinical examination by physicians. Biopsy of these lesions showed that nine of them were potentially dangerous and the removal was required.

[2]Receiver Operator Characteristic

The DBDermo-MIPS system [50], developed at the University of Siena, Italy, analysed 147 pigmented skin lesions (selected from an archive of more than 10,350 images collected between 1991 and 2000, with 3220 of them being excised). The clinical criteria for inclusion of lesions in the study were asymmetrical, pigmented, impalpable lesions with variegated colour, with diameter of 0.4 ± 1 cm. Using an Artificial Neural Networks (ANN) classifier the obtained diagnostic accuracy was about 93%.

The DANAOS expert system [51] also applied ANN, and the system was tested during a multi-centre study in 13 dermatology centres in nine European countries in 1997. The resulting performance of the system on 2218 skin lesions was shown to be similar to that of dermatologists as published in the literature, which was dependent on the size and quality of the images.

While the above-mentioned systems used dermoscopy images, Melafind [52, 53, 54] (developed by Electro-Optical Sciences Inc., Irvington, New York) and SIA[3]scope [55] (developed by Astron Clinica, Cambridge, U.K.) have applied a different technology to capture skin images, using multi-spectral narrow bands ranging from 400 to 1000 nm. Because of the multi-layered structure of skin (as previously explained in Section 2.1 of the thesis) light of different wave-lengths penetrates the skin to different depths. MelaFind takes 10 images of each lesion using 10 different narrow spectral bands from 430 to 950 nm. Eight of these 10 images were used for each lesion, and wavelet analysis along with a linear classifier was applied. The classification accuracy on a difficult image set, including dysplastic and congenital nevi, was reported to be 89% [52].

SIAscope produced eight narrow-band spectrally filtered images of the skin with radiation ranging from 400 to 1000 nm. The purpose of the analysis was to identify the amount, distribution, and depth of certain critical features including collagen, melanin, and haemoglobin within the epidermis and papillary dermis layers of skin. Over an image set of 348 pigmented lesions with lesions greater than 6 mm, the system obtained sensitivity and specificity of about 83% and 80%, respectively [55].

Some other computerised image analysis systems available in the market or under development include [56]: MoleMax II developed by Derma Instruments L. P., MicroDerm developed by VisioMED, NevusScan by Romedix, FotoFinder by Derma Edge Systems Corp., and VideoCap100 by DS Medica. It is acknowledged that [31] the cost efficiency of the existing digital imaging systems (computer-aided diagnosis systems of melanoma) is still an issue of concern. Some of these devices are undergoing clinical trials before final approval, and are not for sale.

A computer-aided diagnosis of melanoma generally comprises five main components; namely, image acquisition, image segmentation (or border detection), feature extraction, feature selection, and classification. A summary of the methods applied to implement each of these components in some of the studies is provided in the remainder of this chapter.

[3]Spectrophotometric Intracutaneous Analysis

2.3.1 Image Acquisition

The first step in a computer-based diagnostic system of melanoma is the acquisition of the digital image of the lesion. The main techniques used for this purpose in the literature include [12]: digitised colour slides which were used in early melanoma analysis studies [57], acquisition of clinical images using still or video cameras, ELM (or dermoscopy previously discussed in Section 2.2.1) which captures detailed information about the surface of the lesion and has been widely utilised in the literature [12, 24, 51, 58, 59, 60, 61], transmission electron microscopy or TEM [62] which is useful for studying growth and inhibition of melanoma. Nevoscope [63, 64, 65], introduced by Dhawan in 1985, is a non-invasive trans-illumination based imaging modality for analysis and diagnosis of melanoma. It involves three modes of illumination: (i) surface illumination which is achieved by directing light perpendicular to the skin surface through fiber optic cables, (ii) trans-illumination performed by directing light into the skin at 45 degree through a ring light source, (iii) epi-illumination obtained from the combination of surface illumination and trans-illumination. The main advantage of trans-illumination is its sensitivity to imaging increased blood flow and vascularisation and also to viewing the subsurface pigmentation in a nevus.

Confocal Scanning Laser Microscopy (CSLM) [66, 56] is a non-invasive imaging technology used for the *in vivo* examination of skin. It uses laser light to focus on a specific spot within the tissue and captures high-resolution images of cellular structures of skin lesions which are comparable to detailed histologic images. The imaging depth is limited by the condition of the stratum corneum and the wavelength of the laser. Currently, the maximum depth of imaging is 200-300 μm at the level of papillary dermis. Confocal images obtained from melanomas are clearly different from those of melanocytic nevi [66]. In addition to exhibiting the morphologic differences between melanoma and melanocytic nevi, CSLM technology has been applied to determining the border of melanoma lesions [67]. Currently, application of the CSLM to the diagnosis of melanoma is in the research stage [30, 56].

Ultrasound is also applied to clinical dermatology, more in Europe but less in the United States [56]. The technology works based on the acoustic properties of skin tissue. There are three modes of ultrasound scan [56]: (i) A-mode is one-dimensional and shows the amplitude of the intensity at different levels in the skin tissue, (ii) B-mode, commonly used in clinical settings, creates two-dimensional images from the brightness level of multiple A-mode scans (iii) C-mode, which is still at an experimental stage, creates a three-dimensional display using computer assistance. The image resolution and tissue penetration are largely dependent on the frequency of the scanner, *i.e.*, an ultrasound scanner with a high frequency has a short wavelength, and has less tissue penetration but higher image resolution. Even though ultrasound alone cannot provide a reliable means of differentiating between melanomas and common melanocytic nevi [68], it is shown to be potentially useful for differentiating melanoma from seborrheic keratosis [69]. Moreover, an important application of the ultrasound technology is the *in vivo* assessment of melanoma thickness [68, 70]. Other imaging modalities used for dermoscopy image analysis include computed tomography (CT) [71], positron emission tomography (PET)

employing fluorodeoxyglucose [72] which has shown high sensitivity and specificity in diagnosing the staging of melanoma, MRI [73], multifrequency electrical impedance [74] and Raman spectra [75].

As outlined previously, image acquisition is a basic yet vital component for a computer aided diagnosis system of melanoma. Development of a reliable computer-based diagnostic system, requires establishing a unified systematic protocol under which a reasonably large collection of good quality and well-illuminated skin images is acquired. In this study, we have applied high-resolution dermoscopy images obtained from Royal Melbourne Hospital, Melbourne, Australia, taken by professional photographers using a Canon EOS 450D camera. At some later stages of the work, we have also applied images from other resources [29], whose contributions are gratefully acknowledged.

2.3.2 Segmentation or Border Detection

Image segmentation is the process of dividing an image into disjoint and homogeneous partitions with respect to some characteristics such as colour, texture, etc. [76]. Alternatively, it is the process of locating the boundaries between the regions, called border detection [77]. Segmentation is a prerequisite in the development of any computer vision system. Similarly, it is the first step towards the automated analysis and evaluation of medical images in a computer-aided diagnostic system. The accuracy of the segmentation result is of high importance due to the bias it can impose on the subsequent steps of the diagnostic system, *i.e.*, in the feature extraction and the ultimate classification result. In computer-aided diagnosis of melanoma the purpose of segmentation is to detect the border of the lesion, in order to separate the lesion, the region of interest (ROI), from the background skin. The accuracy of the detected border is crucial as exclusion of any part of the lesion my lead to loss of dermoscopic patterns, colour, and texture-based information that can be extracted from the interior of the lesion. Moreover, the geometric shape of the lesion and structural properties of the border have diagnostic importance, all of which depends on the detected border (the segmentation results). Low contrast between the lesion and the background skin, irregularity and blurriness of the lesion border, image artifacts such as black frames, skin lines, blood vessels, hairs, and air bubbles, and scars, as well as presence of various colours within the lesion, make border detection a challenging task [78].

Different image features such as shape, colour, texture, and brightness can be employed to perform skin lesion segmentation. Accordingly, numerous methods have been developed for automated border detection in dermoscopy images in recent decades. As suggested by Celebi *et al.* [78] segmentation methods can be categorised into the following class of techniques: (1) Histogram thresholding, which involves the determination of one or more threshold values that separate the ROI from the background, used in [26, 79, 80, 81] (2) colour-clustering methods which partition the colour space into homogeneous regions using unsupervised clustering algorithms, applied by [82, 83, 84, 85, 86, 87, 88, 89], (3) Edge-based methods which apply edge operators to determine the edges between the

background and foreground regions applied by [90], (4) Region-based methods which use region-merging and region-splitting algorithms to group the pixels into homogeneous regions [91, 92, 93, 94, 95, 96] (5) Morphological methods which start from predetermined seeds and apply the watershed transform in order to detect the contours of the object [31, 97], (6) Model-based methods which model the image as random fields and apply optimisation to determine the parameters of the model [92], (7) Active-contour methods such as snakes which use curve evolution techniques to determine the contours of the shape, as applied in [98, 99, 100], and (8) Soft-computing methods which classify pixels using soft-computing techniques such as neural networks, fuzzy logic, and evolutionary computation [101].

Some previous studies have combined the results of different border-detection methods to come up with the final ROI. For example, Pagadala [102] developed a skin segmentation method combining the segmentation results of three thresholding-based methods independently applied on red, green and blue colour channels of the dermoscopy image. We have established in this thesis that there exist a variety of colour channels from other colour spaces which can be taken into account and provide discriminatory power in separating the lesion pixels from surrounding skin pixels. Also, careful consideration needs to be taken while combining the results. Ganster *et al.* [58] also used a fusion process (logical OR operation on resultant binary segmented images) and combined three basic segmentation algorithms, to detect the border of lesions in dermoscopy images. The three basic segmentation methods involved global thresholding, dynamic thresholding, and a method based on 3-D colour clustering concept adopted from [103]. The experimental results demonstrated that [58] thresholding with the blue channels of the RGB and CIE L*a*b colour spaces along with 3D colour clustering on the X, Y, and Z channels of the CIE XYZ colour channel produced the best segmentation results. Using an image set of 40 skin lesions, the thresholding algorithms achieved a good segmentation result in about 80% of the images. The 3D-clustering approach, although it had poor performance in general (about 8%), performed reasonably well in cases where the other two methods were not successful [58]. Unfortunately the details of the three segmentation methods used are not provided in the article.

Based on the accepted assumption of colour uniformity in the skin, unsupervised colour-clustering techniques have been frequently used in applications of skin lesion segmentation. Colour clustering is used to reduce the number of colours such that they are quantised to the most representative colours. Melli *et al.* [87] combined an unsupervised clustering component with a supervised classification module to automatically extract the boundary of the skin lesion. In the clustering phase, they employed and compared four major clustering algorithms; namely, median-cut, k-means, fuzzy c-means, and mean-shift (each will be explained in the following paragraph). The clusters obtained by these clustering processes were then merged and classified into two classes of *skin* and *lesion*. It was assumed that the lesion would occupy the central part of the skin image and the image's corners would belong to the *skin* class. The corner pixels were used as the *skin* training set to train the classifier and the clusters obtained from the clustering module were merged and classified as *skin* class if they contained a particular number of pixels

whose colours have been trained as *skin* colour. After completion of classification and merging procedures, the major object within the image was identified as the *lesion* and the lesion boundary was detected. Over a set of 117 dermoscopy images, the resultant lesion boundaries were compared with the results obtained from dermatologists' manually drawn borders. The comparison, which was performed in terms of sensitivity, specificity and the average of these two parameters, revealed that the best results were obtained from the mean-shift algorithm. Furthermore, they proposed a web-based test platform wherein all four segmentation algorithms were applied to a set of 5616 skin lesions. Dermatologists voted for the best segmentation results in each four-image set, and the overall verification test indicated that the mean-shift algorithm was the best. However, for those images that are corrupted and have no identifiable corners (which is sometimes the case in practice), this technique may not be applicable.

To clarify the above-mentioned clustering process, a brief definition of each method is presented as follows: **Median-cut** [104] is a popular algorithm for colour quantisation which divides the colour space based on the distribution of colours in the original image. The main concept behind this algorithm is to have each colour in the synthesised colour map represent the same number of pixels in the original image. It repeatedly subdivides colour space into smaller and smaller boxes, such that each colour in the new image represents the average of one of these boxes. ***K*-means** [105] is one of the simplest unsupervised clustering algorithms which is designed to cluster n objects into k partitions ($k < n$) based on attributes or feature vectors. It starts with defining random k centroids, one for each cluster, and continues with inserting new objects into clusters such that an error function is minimised. The error function is defined as the sum of the Euclidian distances between the k-clusters' centroids and the objects within each cluster. **Fuzzy c-means** [106] provides a fuzzy clustering method in which each point has a degree of belonging to clusters, as in fuzzy logic, rather than totally belonging to a single cluster. **Mean-shift**, first introduced in 1975 [107], and later in 1997 was employed for colour image segmentation [108]. Mean shift is a simple non-parametric algorithm which approximates the density gradients. For each pixel of an image having a spatial location and a particular colour, the set of neighbouring pixels, within a given spatial radius and a defined colour distance, is determined. For this set of neighbour pixels, the new spatial center (spatial mean) and the new colour mean value are calculated. These calculated mean values will serve as the new center for the next iteration. The procedure is iterated until the spatial and the colour or grey-scale mean stops changing. At the end of the iteration, the final mean colour is assigned to the starting point of that iteration.

Hance *et al.* [82] compared the accuracy of six different colour segmentation techniques and investigated their effectiveness in skin border extraction; namely, adaptive thresholding, fuzzy c-means, spherical-coordinate transform/center split, principal-component transform/median cut, split and merge, and multi-resolution. For the spherical-transform segmentation method, by algorithmic definition, the number of colours for segmentation was set to four, while for other algorithms it was kept constant at three. They showed that the principal-component transform/median cut and adaptive thresholding algorithms provide the lowest average error. They also proposed a combined method of these six methods,

which resulted in further improvement in the number of correctly identified tumor borders. Schmid [83] also proposed a colour-based segmentation algorithm in L*u*v colour space. They applied a modified version of fuzzy c-means, wherein the number of clusters depends on the number of maxima in the histogram information of the three colour components.

In the dermatologist-like tumor extraction (DTEA) method [26], the histogram thresholding technique based on Otsu method [109] was applied to the blue colour channel. This was followed by labelling the components, and merging the regions smaller than a predefined size. The region that met some heuristic criteria based on size, intensity and distance to the margin of the image (defined in [26]) was taken as the initial lesion area. Lastly, the detected border was expanded by an iterative region-growing method to obtain the final border detection result.

Other recent methods include: The KPP[4] method [89] uses spatial constraints in the pigmented lesions and applies a two-stage clustering, based on k-mean++ algorithm, to merge the homogeneous regions. The JSEG method [95], which is an unsupervised approach to border detection, is a modified version of an earlier JSEG algorithm [110]. In this method a colour class map of the image is first formed during a colour quantisation step. Then, in a spatial segmentation step, similar regions are merged based on similarity in the CIE L*u*v colour space. The method was tested on a set of 100 dermoscopy images, using manually determined borders from a dermatologist as the ground truth. The results were compared with three other automated methods (discussed above) proposed by Pagadala [102], the DTEA method [26] and the KPP segmentation method [89], and were shown to be highly comparable with those. The authors acknowledge that the JSEG method may not, in certain circumstances, perform well on lesions with a lot of hair, and have suggested the use of a hair-removal method such as DullRazor [111] might improve the results for such cases. Moreover, as acknowledged by the author, using a single dermatologist as the ground-truth might not be reliable enough to validate the segmentation results.

Celebi *et al.* [78] outlined several issues to be considered when choosing a segmentation method: (i) Scalar versus vector processing, the former is preferred to avoid excessive computational time requirements and the difficulty of choosing an appropriate colour space, (ii) Automatic versus semi-automatic, the latter (*e.g.* active contour models) requires human interaction, and the former is preferred if we aim to develop a fully automated computer-based diagnosis system, (iii) Number of parameters: the more the number of parameters, the harder the model selection. Parametric methods involve a process of determination of the optimal parameter values.

In Chapter 3 of this thesis we propose a fast accurate border detection method which produces borders fairly close to those drawn by dermatologists. Our goal is to provide a practical, simple and easily re-producible solution to the problem of border detection in dermoscopy images, and avoiding complications as far as possible. This is achieved by benefiting from the basic yet useful image processing technique of histogram thresholding. The Otsu method [109] which has been successfully applied by some of the

[4]K Plus Plus ($k$mean++)

previous studies such as [99, 26] is adopted for determining the threshold in our proposed method. Despite the critics of Melli *et al.* [87] about low performance of thresholding-based methods we have demonstrated that with a proper choice of colour channel, a thresholding-based approach can produce borders highly accordant with those manually drawn by expert dermatologists, and even outperforms colour-clustering and region-based border detection methods. In Chapter 3 we show that the proposed border detection method, even though fairly straightforward, outperforms three of the above-mentioned state-of-the-art border detection methods in dermoscopy images. Another reason behind the choice of thresholding-based technique is the simplicity and low computational cost of scalar processing involved in the proposed method.

2.3.3 Feature Extraction

Feature extraction is the process of extracting certain characteristic attributes and generating a set of meaningful descriptors from an image. The purpose of the feature extraction component in a computer-aided diagnosis systems of melanoma is to extract various features from a given skin image which best characterises a given lesion as benign or malignant. The feature extraction methodology of many computerised melanoma detection systems has been generally based on the conventional clinical ABCD-rule of dermoscopy due to its effectiveness and simplicity of implementation. Its effectiveness stems from the fact that it incorporates the main features of a melanoma lesion such as asymmetry, border irregularity, colour and diameter (or differential structures), that are measurable and computable by computer. The features employed by the classification system must provide high sensitivity (high correlation of the feature with malignancy, resulting in high true positive response) and high specificity (to produce high true negative response). Although in the typical classification paradigm both factors of sensitivity and specificity are considered important (the trade-off is determined by maximising the AUC value), in the case of malignant melanoma detection or probably most cancer detection applications, obtaining a high true positive rate or alternatively suppression of false negatives is desired [12]. This section addresses some of the features employed by previous computer-based melanoma recognition studies.

Asymmetry

According to the ABCD rule of dermatoscopy, symmetry is given the highest weight among the four features [15]. Asymmetry is also evaluated in the Menzies and CASH methods. Moreover, in the Consensus Net Meeting on Dermoscopy [20], asymmetry, along with a few other criteria, was highlighted as a feature strongly associated with melanoma. As previously mentioned, in the ABCD rule of dermatoscopy [21], to quantify the asymmetry, the lesion is bilaterally segmented by two perpendicular axes positioned such that the resultant asymmetry score is minimised. The asymmetry score is calculated according to the contour (shape), colour distribution and internal structure of the lesion on either side of each axis. The asymmetry score is 2 if both axes show asymmetry. Asymmetry on only

one axis results in a score of 1 and if there is no asymmetry on either axis, the score is 0. Stolz *et al.* [21] note that in their experiment 96% of malignant melanoma cases had an asymmetry score of 2, whereas only 24.2% of benign melanocyte nevi had such a score.

Numerous approaches have been proposed to quantify the asymmetry in skin lesions. Some of the methods try to imitate the ABCD rule of dermatoscopy and investigate lesion asymmetry with respect to a symmetry-axis passing through the lesion. In these approaches, the symmetry axis is determined in a variety of ways, such as principal axis [112, 113, 114], major axis of the best-fit ellipse [115, 116], Fourier transform [117], longest or shortest diameter [118], etc. Then, the difference between the area on both sides of the axis is calculated. In another technique, symmetry determination is based on geometrical measurements on the whole lesion, such as circularity [119, 120, 118]. Some others, quantify the homogeneity of the lesion, wherein for certain patches the overall internal structure or colouration of the lesion is evaluated [114, 115, 121, 122, 123]. According to Stolz *et al.* [21], due to the fact that some equivocal lesions have symmetrical shapes, incorporating colour and structural asymmetry is very crucial to scoring the symmetry of the lesion. In the following we provide a summary of existing methods in evaluating the symmetry measure in dermoscopy images, categorised in (i) Shape Asymmetry and (ii) colour and Pattern Asymmetry.

(i) Shape Asymmetry:

In numerous studies, a *circularity index* (also known as compactness, thinness or roundness index) has been applied and claimed to be a proper indication of symmetry in skin lesion [118, 119, 120, 124] or at least introduced as a useful geometry variable in quantifying the skin lesions [122]. This index is calculated according to the following equation and returns the value of one for the perfect symmetrical shape, *i.e.*, a circle:

$$Circularity = \frac{4A\pi}{P^2} \tag{2.2}$$

In this equation A and P refer to the area and perimeter of the shape, respectively. This metric is scale-independent, which makes it a handy tool for measuring the symmetry. However, it does not show enough efficacy in lesions with fuzzy and irregular borders. In such cases, the segmented region generally has a thick border, but a small internal area. Thus, the circularity index value attains a very small value, which does not reflect the actual circularity of the lesion. Therefore, the significance of this method tightly depends on the segmentation precision.

The *symmetry distance* (SD) measure is another concept applied to skin images in order to quantify the symmetry of the lesion. Symmetry was treated as a continuous measure by Zabrodsky *et al.* [125, 126] firstly in 1992. They proposed the idea that in the real-world, analysing the symmetry feature as a discrete value (Yes for symmetrical shapes and No for asymmetrical shapes), is not sufficient to characterise symmetry phenomenon in nature, *e.g.* the face and human body as inexact symmetrical shapes. Accordingly, the *continuous symmetry measure* (CSM) was defined as a metric to calculate the amount of symmetry of the shape, using which one can compare the symmetry properties between

different shapes, as well as quantifying the degree of symmetry for a single shape. According to Zabrodsky *et al.* [125] CSM is defined as "The continuous symmetry measure quantifies the minimal distance movement that the points of an object have to undergo in order to be transformed into a shape of the desired symmetry." Later on, in 1995 they proposed [127] SD as "quantifier of the minimum effort required to transform a given shape into a symmetric shape". Formally, for any two shapes P and Q represented by a sequence of n points, the distance between these two is defined as:

$$d(P,Q) = \frac{1}{n}\sum_{i=1}^{n} \|P_i - Q_i\|^2 \tag{2.3}$$

The *symmetry transform* (ST) of the shape P is defined as the symmetrical shape $\hat{P}$ with the minimum distance $d(P,\hat{P})$ to the asymmetrical or nearly symmetrical shape P. The distance between $\hat{P}$ and P is called the symmetry distance and it is formulated as follows:

$$SD = \frac{1}{n}\sum_{i=1}^{n} \left\|P_i - \hat{P}_i\right\|^2 \tag{2.4}$$

According to the definition of SD, each shape is represented through a sequence of points. The point-selection procedure highly impacts the resultant SD value [127], and three methods were suggested to pick out a successive set of points to express a shape; (1) selection of points at equal distance in terms of curve length, which turned out to be not very efficient for noisy, fuzzy, irregular and occluded cases, and particularly for objects with extensions [118], (2) selecting points on a smoothed version of the shape boundary and projecting those selected points to the primary shape. Smoothing is done by averaging the points' neighbours on the boundary and moving the point to the centroid of the neighbourhood. The larger the size of the neighbourhood, the smoother the resultant contour, and in the ultimate case it would convert to a circle. In this case, the third method is used, (3) selecting the points at equal angles on the contour [127].

Ng *et al.* [118, 120] later applied the SD measure to investigate the symmetry property in skin lesions. They initiated their study in 1994 to collect data of three types of melanocytic lesions; benign nevi, dysplastic nevi and malignant melanoma. The purpose was to improve the performance of their system over their previous work, wherein the circularity index was applied. They used the third method (mentioned above for selecting the point to represent the contour) and established an *adaptive symmetry distance* (ASD) measure. Since the number of required representative points varies against the border's fuzziness and irregularity, it was not possible to have a predefined number of points as it is in SD. ASD calculates successive SDs by using an incremental number of boundary points, until the difference between two successive SDs hits a predefined threshold. However, due to inefficiency found in ASD when dealing with irregular and fuzzy borders, the *adaptive fuzzy symmetry distance* (AFSD) was proposed. AFSD involves a fuzzy factor f which represents the segmentation accuracy and implies that for any boundary point P_i, its actual location can be any point within a circle with P_i as centre and $\|P_i - C\| \times f$ as radius. In order to test the efficacy of the proposed symmetry measurements in [120], *i.e.*,

SD, ASD, AFSD and circularity (CIRC), these measures were applied on 84 skin images of non-hairy melanocytic lesions. Each lesion was also labelled as either symmetrical or asymmetrical by a dermatologist. They used the Student t-test[5] to compare the mean values obtained from each of the symmetry measures. Among all the methods, CIRC had a reasonably good performance, the SD method had the worst differential ability, whereas AFSD has the best performance among all. They also recognised the fact that during their tests on lesions with irregular or fuzzy borders, the circularity calculation yielded a very high value which failed to reflect the actual measurement's goals.

Introducing the fuzzy concept to the problem reflects the fact that in fuzzy SD measurements, need no exact definition for the contour and an approximate definition is sufficient to measure the SD value. Yet, since different segmented lesions yield different SDs, they tend to differentiate SD values which correspond to either symmetrical or asymmetrical shapes. Thus, the main concern in their next study [118] was to broaden the numeric range of SDs, so that one can easily label a lesion as either symmetrical or asymmetrical by looking at its SD value. To extend their previous adaptive fuzzy symmetry distance algorithm in [120], they proposed a new adaptive fuzzy symmetry distance in [118], wherein the previous fuzzy algorithm was enhanced by redefining the fuzzy factor f as the border width W. Moreover, in the new adaptive fuzzy algorithm, only unrepresentative points with W greater than the average of neighbouring points are moved, whereas in the previous algorithm all points may move inward or outward. Since the degree of fuzziness is not similar in all the points along the boundary, the authors suggested an algorithm to calculate a new parameter, called the reliability factor, to measure the fuzziness of a sample point in the border. This measure increases as the fuzziness of the border decreases.

Another difference between these two works of Ng *et al.* [118, 120] is that in the latter they applied both mirror and rotational symmetry, whereas only rotational symmetry was used in the former. Hence they [118] also propose the idea of finding the best axis of symmetry, because in mirror-symmetry ST the transformation is based on the position of the symmetry axis. The authors noted that [118] the optimal axes would yield the smallest SD value. In order to find the smallest SD value, three axes were investigated: the longest axis of the lesion, the principal axis of the lesion, and the axis with smallest length difference. They tested these three methods on an image set of 30 lesions, 15 symmetrical and 15 asymmetrical images, and calculated the *mirror symmetry distance* (MSD) value for each. The results showed the minimum value for MSD in case of using the axis with smallest length difference. In order to test the efficacy of each algorithm, they selected 120 images, 60 symmetric and 60 asymmetric, and ran all the algorithms on each image, *i.e.*, circularity index (CIRC), mirror-symmetry distance (MSD), fuzzy symmetry distance (FSD), adaptive fuzzy approach in mirror-symmetry distance (AFMSD), adaptive fuzzy approach in rotational symmetry distance (AFRSD), AFMSD with reliability factor, and AFRSD with the reliability factor. The t-test results demonstrated that both

[5]A statistical method which assesses whether the means of two groups are statistically different from each other. It is applied when the population is assumed to be normally distributed, generally with sample sizes of less than 30 cases.

adaptive fuzzy approaches, *i.e.*, AFMSD and AFRSD, are more powerful than MSD and FSD in discriminating SD values. With reliability factor applied, AFMSD works even better, whereas AFRSD failed to give good results. Moreover the authors reported that for moderate skin lesions some false positive results may arise, because SD does not differentiate symmetrical lesions from asymmetrical ones. Ng e*t al.* [118] also implemented a back-propagation neural network to enhance the classification of skin lesions as symmetric or asymmetric, according to their SD values. For this purpose, they applied the most accurate SD measures from the previous experiments, *i.e.*, AFMSD with reliability factor and AFRSD, along with the average fuzzy border width as the inputs to the neural network. Conducting several trials, they finally decided to use a 3-8-1 network architecture, which led to the overall accuracy of 80% in classification of symmetry/asymmetry skin lesions.

Bulkiness concept was applied by Claridge *et al.* [128] in order to investigate the asymmetry of the lesion in terms of closeness to the equivalent ellipse, using the following formula:

$$Bulkiness = \frac{\text{area of equivalent ellipse}}{\text{area of the original shape}} \tag{2.5}$$

The equivalent ellipse has the same center of gravity and similar moment of inertia to that of the lesion. The Bulkiness value of a regular shape is close to 1, while the value is higher for irregular lesions.

Stoecker *et al.* [112] proposed an objective definition for asymmetry. They recognised the fact that there was no agreement on what percentage of asymmetry is needed in a shape to be assumed as asymmetric. They also estimated that dermatologists dissent in 5–10% of cases when labelling pigmented lesions as asymmetric. The authors dismissed the existing algorithms for their complexity and low performance, thus suggested a new algorithm based on principal-axis theory. Taking for granted that for a symmetric shape, at least one principal axis is in accord with the symmetry axis, they assumed that for an almost symmetrical contour, the principal axis can be considered as a good approximation of the axis of symmetry. They propose the following set of equations to determine the axis of symmetry.

Angle θ_m of principal axis with respect to the x axis is given by:

$$\tan 2\theta_m = -2 \times \frac{P_{xy}}{I_x - I_y} \tag{2.6}$$

A binary shape is represented by a function $b(x, y)$ such that:

$$b(x,y) = \begin{cases} 1 & \text{if pixel } (x,y) \text{ is inside the shape} \\ 0 & \text{if pixel } (x,y) \text{ is outside the shape and belongs to the background} \end{cases} \tag{2.7}$$

The area A is given by:

$$A = \int \int_I b(x,y) dx dy \tag{2.8}$$

Moment of inertia with respect to x axis:

$$I_x = \int_I y^2 dA \tag{2.9}$$

Moment of inertia with respect to y axis:

$$I_y = \int_I x^2 dA \tag{2.10}$$

Product of inertia:

$$P_{xy} = \int_I xy dA \tag{2.11}$$

To quantify the asymmetry they propose an asymmetry index. To calculate this measure they fold the lesion along the principal axis and subtract the region on one side of the axis from the reflected region on the other side.

$$\text{Asymmetry Index} = \frac{\Delta A_{\min}}{A} \times 100 \tag{2.12}$$

To investigate the accuracy of the method, dermatologists classified 86 lesions as symmetric or asymmetric. The same lesions were tested by the proposed method and the results showed agreement in about 93% of cases. This approach has been also applied in other studies such as [113, 114, 129] to investigate the asymmetry of skin lesions.

Lastly, Clawson *et al.* [117] proposed a method based on Fourier descriptors to determine the optimal axes to quantify lesion symmetry. The authors argued that the location of the symmetry axis strongly affects the accuracy of the ABCD diagnostic algorithm. The authors used the Fourier descriptors with complex coefficients and assumed that for a symmetrical contour along the vertical axis with starting point on that axis, the real part of all Fourier coefficients equals zero. Likewise, for a horizontally symmetric contour the imaginary part of each Fourier descriptor is zero. Consequently, for a symmetric shape on both vertical and horizontal axes all even Fourier coefficients are zero.

$$a(u) = \frac{1}{N} \sum_{k=0}^{N-1} s(k) e^{-i2\pi uk/N} \;\; ; \;\; u = 0, 1, ..., N-1 \tag{2.13}$$

Taking the above properties as given, they used the closeness to zero of real parts of Fourier descriptors to find the position of optimal symmetry axes in an asymmetrical contour. Summation of the real parts of Fourier coefficients is used as the symmetry measure and they propose the following formulas for the start point and rotation normalisation:

$$a_p(u) = a(u)^{e^{-j2ku/N}} \;\; ; \;\; u, k = 0, 1, ..., N-1 \tag{2.14}$$

$$a_r(u) = a_p(u)^{e^{j\theta}} \;\; ; \;\; 0 < \theta < \pi \tag{2.15}$$

The minimum sum of absolute real values across all normalisations subsequently shows the location of the principal symmetry axis and its inclination from the vertical

axis. Due to the fact that, as minimum asymmetry corresponds to minimum value of the absolute sum of real coefficients, the proposed method is claimed to be able to rank potential axes of symmetry.

Clawson *et al.* [117] applied the proposed method, which consists of producing Fourier descriptors and finding the optimal principal axis by equalising the coefficients to zero, on a test set of 30 images. Two groups of dermatologists subjectively determined the best axis of asymmetry for the test set which results in 60% of consistency (±15 degrees) and 30% of variability, with a mean difference of 33 degrees in angle inclination. Comparing Fourier-based symmetry-axis identification with those achieved by dermatologists, they converge in 68% of cases considering all the experts as the whole, whereas 60% matching is found with one group of dermatologists and 76.6% with the other one, when investigated separately (±15 degrees). Using the algorithm as a ranking method for the potential symmetry axis of the lesion leads to a matching score of 92% (±15 degrees) between algorithm results and dermatologists' opinion. Individual investigations result in 93.3% matching with one group and 90% with the other. In 77% of cases, the best two unique axes identified by Fourier method were orthogonal to each other (±15 degrees). According to their report derived from the literature, the maximum diagnostic accuracy of automated asymmetry quantification was 86.6%, whereas they claim the matching of 92.2% of cases between their method and dermatologist scoring. However, they have not applied other symmetry measurement techniques on the same image set in order to provide a valid comparison. We also believe that other methods such as circularity index or asymmetry index using principal axis are more easily reproducible and have less complications in the implementation, since they invlove evaluation of simple mathematical expressions (Equations 2.2 and 2.6).

(ii) colour and Pattern Asymmetry:

This section provides a summary of some of the existing methods which measure the texture and colour asymmetry of the skin lesion. Change *et al.* [115] proposed a symmetry feature, called *solid pigment asymmetry index*. The pigment network beneath the epidermis is dominant in some parts of the lesion, producing a dark colour termed as solid pigment. Asymmetric solid pigment is considered as an indication for the presence of melanoma. The authors applied the histogram-based thresholding technique on the luminance image to detect solid pigment areas inside the lesion, and quantified the asymmetry of the position of solid pigmented areas within the lesion via a pigment asymmetry index. In order to calculate this index, they proposed a multi-step algorithm: the principal axis of the segmented lesion was calculated, using which the lesion was partitioned into four quadrants. Then, for each quadrant ($i = 1, ..., 4$), the area (A_i^L) and centroid location ($\bar{x}_i^L$,$\bar{y}_i^L$) are estimated. Afterwards, the area (A_i^P) and centroid location ($\bar{x}_i^P$,$\bar{y}_i^P$) of the pigment region within each quadrant were calculated. Next, they computed the distance between the centroid of the lesion in each quadrant and the centroid of the whole lesion (D_i^L), and also the distance from centroid of the pigment region in that quadrant to the centroid of the entire lesion (D_i^P). The pigment asymmetry index for each quadrant was

then calculated according to the following formula:

$$\lambda_i = \frac{A_i^P D_i^P}{A_i^L D_i^L} \quad , \quad i = 1, 2, 3, 4 \tag{2.16}$$

Finally, they defined the pigment asymmetry index of the entire lesion as the variance of pigment asymmetry indices calculated for the quadrants:

$$\alpha = \frac{\sum_{i=1}^{N} (\lambda_i - \mu_i)^2}{N} \tag{2.17}$$

where μ is the mean pigment asymmetry index over four quadrants ($N = 4$). The authors noted that this approach is computed by quadrants, which implies that the two-axis symmetry is more meaningful than merely computing the distance between pigment lesion and centroids. They stressed that this implementation is coherent with asymmetry measurements in melanoma, which is based on two-axis symmetry rather than one-axis. A possible issue of contention in this approach is the assumption that there is only one solid pigment area within each quadrant, which may not be always correct.

Clawson *et al.* [121] proposed a technique to detect colour asymmetry. The method analysed the pigment distribution along radial paths connecting the center of the lesion to each of the boundary points. The proposed method started with converting the colour lesion image to greyscale image. Then for each radial path the mean greyscale value was estimated by:

$$Av_i = \frac{\sum_{k=1}^{N_i} r_{ik}}{N_i} \tag{2.18}$$

where r_{ik} is the greyscale value of the k^{th} pixel along the radial path connecting the centroid to the i^{th} boundary point, and N_i is the number of pixels along this radial path. Next, for each lesion A_L was defined as a metric indicating the greyscale distribution over the whole lesion, given by:

$$A_L = \frac{\sum_{i=1}^{N} Av_i}{N} \tag{2.19}$$

where N is the number of lesion boundary points. Finally, for each boundary point a Normalised colour Distance (NCD) was defined as:

$$NCD_i = Av_i \times \left(\frac{100}{A_L} \right) \tag{2.20}$$

The measured NCD values were used to draw a new boundary for the lesion, *i.e.*, each NCD value showing the radial path length between centroid and the corresponding boundary point. As equation (2.20) implies, if all NCD values are equal to 100, the new contours will be a circle. The authors stressed that hyper-pigmentation areas lead to larger NCD, while regression areas result in small NCD values. In order to quantify the colour asymmetry, they defined a mismatch metric, by which the generated contour (C_L) was compared with

the circle (C_s), which is the symbol of perfect symmetry of pigment distribution. The mismatch metric was given by:

$$mismatch = \left(\frac{A_x}{A_t}\right) \times 100 \tag{2.21}$$

where A_x is estimated by applying the XOR function between C_L and C_s, and A_t is the area of C_L. They also calculate the circularity of the lesion.

Moreover, they analysed NCD values and their deviation from 100 as the reference value, to extract some parameters: namely, number of peaks (values greater than 100), number of troughs (values smaller than 100), largest peak area (LPA) showing the area of hyper-pigmentation within the lesion, largest trough area (LTA) presenting the area of regression inside the lesion, area/width of the largest peak (LPA/LPW) quantifying the darkness of the largest area of hyper-pigmentation, area/width of the largest trough (LTA/LTW) quantifying the lightness of the largest area of regression, the largest value of area/width across all peaks (LPS) corresponding to the darkest region within the lesion; and the largest value of area/width across all troughs (LTS) corresponding to the lightest region within the lesion. An image set of 30 lesions with different degrees of colour symmetry were segmented manually; for each, 64 boundary points were selected. The above nine metrics were measured on each image. Moreover, the Classification and Regression algorithm and decision tree (C5) and also a neural network were used to investigate the classification rules, according to which, each lesion could be classified as either symmetrical or asymmetrical. Across all classification methods, the most significant parameter was LTA/LTW, quantifying the largest area inside the lesion where colour is lighter than average, in relation to the arc length of this region. When applying the neural network, two other important parameters were identified which were also related to possible areas of regression within the lesion: namely, LTS and LTA.

Seidenari *et al.* [130] proposed a method to objectively evaluate pigment distribution in three types of skin lesions; namely, early melanoma (MM), atypical nevi (AN) and benign nevi (BN). The aim was to investigate the efficacy of measured pigment distribution parameters in differentiating malignant lesions from normal ones. The authors noted that a big difference among the values of the distribution parameters indicates non-homogeneity within the lesion. The pigment distribution assessment consists of three steps: First, the lesion is simplified into colour blocks by removing the colour details. To attain this, they select a desired block size, *i.e.*, level of detail, and subdivided the image with a grid of the selected block size. In the next step, the average colour within each block is calculated. In the last step, the mathematical colour difference parameters between blocks, *i.e.*, mean (PIXMED), variance (PIXVAR), and maximum colour differences (PIXMAX) are computed, using Euclidean distance in RGB colour space. For each lesion, a set of three parameters, in two levels of detail are obtained, *e.g.* PIXVAR1, PIXVAR2, etc. Thus, they obtained six measurements per lesion, *i.e.*, three parameters for each detail level of 1% and 2%.

The method was applied on a set of 459 images (76 AN, 288 BN and 95 MM). SPSS statistical software was used to statistically analyse the results, in which mean and

standard deviation of PIXMED, PIXVAR and PIXMAX values were computed for MM, AN and BN lesions. The results showed that distribution values for MM were higher than those of both AN and BN, signifying complex and non-homogeneous structure and colour variation. Furthermore, the obtained values for AN were higher than those of BN. This ordering of values for MM, AN and BN reflects the corresponding histology of these lesions. In order to test the discrimination power of proposed statistical parameters and identify the more distinctive variables, discriminant analysis was conducted on a training set of 229 lesions, which is half of the initial image set in each lesion category. Moreover, ROC analysis was performed to calculate the sensitivity and specificity and also the diagnostic accuracy of the classification of the lesions in the test set, which are the remaining 230 images. The authors found that discriminant analysis method indicates that PIXMAX1, PIXMED2, PIXVAR2 and PIXMAX2 variables can be used as differentiators of MMs from others in the training set. On the test set the values for accuracy, sensitivity and specificity are reported as 86.6%, 87.5% and 85.7%, respectively. By this approach, the authors tried to resolve the problem of symmetry axis identification, since each block is compared with every other block in the image, rather than only with its symmetrical pair. However, this may require a high number of comparisons, *i.e.*, $n(n-1)/2$ for n-block lesion, which can be computationally expensive. A benefit of this algorithm is that applying different levels of detail, *i.e.*, different degrees of subdividing, ameliorates the image-resizing issue.

In a similar study, Seidenari *et al.* [114] performed a numerical assessment of asymmetry based on comparison of CIE L*a*b[6] colour components within the lesion colour blocks. After finding the boundary, centroid and principal axes of the lesion, the lesion was rotated along the principal axes, so that the major axis was aligned with horizontal axis and the minor axis with the vertical one. Then, the image was subdivided in a similar way to that explained above. They provide four block sizes corresponding to resizing of the image to 1%, 2%, 5% and 8%. They presumed that colours were distributed as a multivariate distribution in each colour block, so that they can be described via mean CIE L*a*b* vector and its corresponding covariance matrix. In the last step, they investigated the colour asymmetry by finding the distance between each block and its symmetric pair with respect to the major and minor axes. This distance is measured by the Bhattacharyya equation given by:

$$B(a,b) = \frac{1}{8} \times (m_a - m_b)^T \left[\frac{\Sigma_a + \Sigma_b}{2}\right]^{-1} (m_a - m_b) + \frac{1}{2} \log \frac{\left|\frac{\Sigma_a + \Sigma_b}{2}\right|}{\sqrt{|\Sigma_a|\,|\Sigma_b|}} \tag{2.22}$$

where m_a and m_b are mean vectors, and Σ_a and Σ_b are the covariance matrices of the distribution of two colour blocks. They average the calculated distance over all blocks to obtain the total distance. They measure a total number of eight parameters, *i.e.*,, a set of two parameters for each of the axes, for each block size.

Manousaki *et al.* [122] suggested the fractal dimension of lesion surface (FDMB) to calculate the pigment distribution irregularity on the surface of the lesion. They applied

[6]CIE L*a*b* is the most complete colour space specified by the International Commission on Illumination. It describes all the colours visible to the human eye and was created to serve as a device-independent model to be used as a reference.

Minkowski-Boulingand algorithm, also known as box-counting, to estimate FDMB. They also proposed a colour texture feature of heterogeneity, which refers to uncompleted space filling within the lesion. The parameter, called grey-scale lacunarity of lesion (Lac grey), shows the number and size of colour voids inside the lesion. Holmstrom [123] stated that certain hue[7] values indicate malignancy. The areas containing these specific hue values were identified and their centroids calculated. The distance between these centroids and the center of gravity of the whole lesion was also estimated. Comparison between the acquired distance values was performed to reveal the colour symmetry/asymmetry property of the lesion.

Border

Holmstrom [123] breaks down the border-feature problem into two separated problems of (i) border blurriness and (ii) border irregularity. In the following we summarise some of the existing methods in quantifying each.

(i) Border Blurriness:

Day [131] investigated the blurriness of the lesion boundary by measuring the lightness gradient of the lesion along its border. Similarly to the clinical ABCD technique for measuring B feature, the lesion was divided into eight segments, and in each segment five equally-distant boundary points are chosen. The author defined the sharp pigment cut-off at each segment, as a major difference between lightness values of skin and lesion in that segment. The L value from CIE L*a*b colour space was used as the lightness measure. To evaluate the lightness rate at each boundary point P_i, the lightness value of 30 pixels inside the border and 30 pixels outside the border were measured. These points were located along the line connecting P_i to the centroid of the lesion. Thus, for each of the five boundary points at each of the eight segments, 60 lightness values are recorded. Then, at each boundary point P_i, the slope of the lightness values is measured by using the method of least squares. For every segment, five lightness rates are obtained, the maximum of which was selected as the B score for that segment. These scores were compared with a given threshold to decide whether each segment has a sharp border ($Bscore = 1$) or a blurry one ($Bscore = 0$). The results were correlated with dermatologists' ratings, upon which a threshold value could be tuned to acquire a better scoring. The author acknowledged the shortcoming of the proposed method that in comparison between the scores obtained from the algorithm and those found by the dermatologists, only the total B score was applied, meaning that the algorithm and dermatologists might have marked different segments as sharp cut-off. Manousaki *et al.* [122] measured the sharpness of the border by calculating the standard deviation (Std) of grey intensity on the border of lesion. Moreover, they computed the coefficient of variation of Std of grey intensity. The authors noted that these values indicate the degree of distinction of the lesion form the

[7]Hue is is the dominant wavelength of a colour and is considered as one of the main properties of a colour described with names such as red or blue. Hue is also one of the three dimensions in some colour spaces such as HSV or HSL along with saturation and value/lightness.

surrounding normal skin and that higher values correspond to sharper borders.

(ii) Border Irregularity:

Different techniques have been introduced to measure the border irregularity of the skin lesions, some of which are discussed as follows. Claridge *et al.* [128] identified two types of irregularities on the border of the lesion and called them *textural* irregularity and *structural* irregularity. The former refers to small variations along the border of the lesion, which are sensitive to segmentation noise. The latter denotes the global indentations and protrusions on the whole shape, which show high correlations with malignancy of the lesion. Lee *et al.* [132, 133, 134] introduced a new shape descriptor which measures and locates structural irregularities along the border of the skin lesion. To analyse the indentations and protrusions, the lesion border was parameterised as a planar curve L_0, defined as follows:

$$L_0 = (x(t), y(t)) \quad ; \quad 0 \leq t \leq 1 \tag{2.23}$$

where t is the path length variable along the planar curve, and $L_0\left(x(0), y(0)\right) = L_0\left(x(1), y(1)\right)$. Curvature was used to measure the irregularities. For a curve $L_0\left(x(t), y(t)\right)$, the curvature function $K(t)$ is given by [135]:

$$K(t) = \frac{\frac{dx}{dt}\frac{d^2y}{dt^2} - \frac{dy}{dt}\frac{d^2x}{dt^2}}{\left(\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2\right)^{2/3}} \tag{2.24}$$

Convex areas are associated with positive curvature values ($K > 0$), whereas negative values show concavity. Local extrema of curvature show the peaks of concave and convex segments, which can be found by zero-crossings of the first deviation of $K(t)$. Accordingly, the lesion border was divided into curve segments composed of three consecutive local curvature extrema $[t_1, t_2, t_3]$, where t_2 shows the segment peak and its type. Terms t_1 and t_3 have the same sign, yet opposite from t_2. Indentations and protrusions were defined as follows:

$$\begin{aligned} \text{indentation} &= (K(t_1) > 0, K(t_2) < 0, K(t_3) > 0) \\ \text{protrusion} &= (K(t_1) < 0, K(t_2) > 0, K(t_3) < 0) \end{aligned} \tag{2.25}$$

Or, similarly:

$$\begin{aligned} \text{indentation} &= (\text{convex,concave,convex}) \\ \text{protrusion} &= (\text{concave,convex,concave}) \end{aligned} \tag{2.26}$$

The border's curvature function was then scanned to find the triples of consecutive curvature extrema with alternating signs, so that the lesion border can be described as a set of indentation/protrusion segments.

To find the location of convex and concave curvature extrema, and consequently pinpoint indentations and protrusions, Lee *et al.* [132, 133, 134] proposed an algorithm called *extended curvature scale-space filtering*, which modified the classic curvature scale-space

filtering. The classic scale-space filtering technique, which is useful in object-recognition systems, represents the shape contour by locating the zero curvature locations. This process is done by convolving the original border function $L_0\left(x(t), y(t)\right)$ with a Gaussian kernel $g(t, \sigma)$ to extract the global features of the shape.

$$\begin{aligned} L(t,\sigma) &= L_0\left(x(t), y(t)\right) \otimes g(t,\sigma) \\ &= \left(X(t,\sigma), Y(t,\sigma)\right) \end{aligned} \tag{2.27}$$

where $L(t, \sigma)$ is the smoothed border, $\otimes$ refers to the convolution operator and the Gaussian function is given by:

$$g(t,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-t^2/2\sigma^2} \tag{2.28}$$

The Gaussian standard deviation σ shows the degree of smoothing and the lesion contour is iteratively convolved by the Gaussian kernel with an increasing σ. This smoothing process is continued until all border cavities are eliminated, and at each smoothing level, zero curvature locations are recorded in a binary 2D scale-space image. This approach was extended [134] so that it can be used to investigate the structural irregularities, *i.e.*, analysing the indentation and protrusion segments and locating their positions along the lesion border. In order to do this, at each smoothing scale σ, the curvature function of the smoothed curve was calculated and the zero-crossings of the first derivative of the curvature function were determined. Moreover, their location and concavity/convexity attributes were recorded. Herein, the binary image of the classic approach was required to be extended to a three-valued image so that it could represent the new requirements, *i.e.*, locations of the curvature extrema along with their concavity/convexity attribute. Lee *et al.* [132] introduced a new descriptor to indicate the overall ruggedness of the border. The descriptor, called *sigma-ratio* (SR), measures the structural irregularities along the lesion border. It was defined as the normalised minimum σ required to remove all the indentations and protrusions on the border. The normalisation is done by dividing the final σ value by the length of the lesion, so that the measure is scale-invariant and independent of the size of the lesion.

Lee *et al.* [132, 133] acknowledged that the sigma-ratio descriptor is sensitive to long and narrow indentations, *e.g.* presence of a dark hair on the lesion image. Such a case would lead to a very high value for the sigma-ratio, which is a false positive. Hence, preprocessing of the image and removing the hairs before applying sigma-ratio operator is necessary. Also, due to the non-linearity of the Gaussian filter, the sigma-ratio is a non-linear smoothing process. However, SR was demonstrated to be more sensitive to structural irregularities than other measures such as compactness index (CI), fractal dimension (FD) and structural fractal dimension (SFD). They extended [133, 134] the SR measure and proposed new area-based measurements, called *irregularity index* (II). During the process, indentation or protrusion areas under the smoothing operation are either filled or removed. The size of the filled/removed area, also known as irregularity area, was used to specify the irregularity index. The index needed to be normalised to be compared with the II of other lesions. Thus, the irregularity area was divided by a smoothed border at

the corresponding smoothing level σ. The II of an indentation/protrusion segment U, referred to as I_U, was defined by:

$$I_U = \frac{\Delta_U}{R_U} \times 100\% \tag{2.29}$$

where Δ_U denotes the irregularity area of the segment U, and R_U is the area of the smoothed shape at σ level. By scanning the entire border, a set of irregularity indices $(I_1, I_2, ..., I_n)$ was obtained, each of which could be used as a border descriptor measuring the structural irregularities along the border of the lesion. Lee *et al.* [133, 134] suggested two of these descriptors: namely, the most significant irregularity index (MSII) and the overall irregularity index (OII), defined by the following equations:

$$MSII = \max\{I_1, I_2, ..., I_n\} \tag{2.30}$$

$$OII = \sum_{j=1}^{n} I_j \tag{2.31}$$

To test the proposed measures, the performance of MSII and OII with that of CI, FD and SFD was compared. The experiment was carried out on an image set of 40 lesion borders, and the results were correlated with the clinical results obtained by 14 dermatologists. The authors demonstrated that the new area-based measures presented the highest correlations with dermatologists' evaluations, and the suggested indices properly reflected the structural irregularities along the borders of the lesion. Lee *et al.* [136] conducted another study to investigate the predictive power of the irregularity index on malignant melanoma by employing statistical analysis. They suggested the four following parameters to characterise the lesion of the border:

- Indentation Irregularity Index (III): the sum of Irregularity Indices for all structural indentation segments.
- Protrusion Irregularity Index (PII): the sum of Irregularity Indices for all structural protrusion segments.
- Maximum Indentation Irregularity Index (MIII): the largest III's.
- Maximum Protrusion Irregularity Index (MPII): the largest PII's.

To analyse the predictive power of the proposed indices, three statistical methods were applied: namely, Student's t-test, linear discriminant analysis, and ROC analysis. In Student's t-test the mean and standard deviation of the four parameters derived from benign and malignant image sets were compared. For discriminant analysis, the following linear model was proposed:

$$D = B_0 + B_1 \times III + B_2 \times PII + B_3 \times MIII + B_4 \times MPII \tag{2.32}$$

The goal was to find the best coefficients B_i which provide the best separation between malignant and benign lesions. A thresholding was applied on D to classify the lesions;

values higher than D show irregular borders and malignancy, while values smaller than D indicate smooth border and benignancy. The rate of successful classification is determined by:

$$\text{Classification Rate} = \frac{\text{number of true positives} + \text{number of true negatives}}{\text{number of all cases}} \tag{2.33}$$

The ROC analysis was performed both on irregularity indices and discriminant score D to evaluate the sensitivity and specificity of all possible cut-off points. The area under ROC curve was also measured to investigate the predictive power of the classification methods. Moreover, the odds ratio (OR) was calculated to approximate the likelihood of a melanoma with irregular border over the probability of a benign lesion with irregular border. OR is given by:

$$OR = \frac{\text{melanomas with irregular border} \times \text{benign lesions with smooth border}}{\text{melanomas with smooth border} \times \text{benign lesions with irregular border}} \tag{2.34}$$

where $OR > 1$ implies the chance of associating melanoma with irregular border. The results of Student's t-test demonstrated that all four irregularity indices could be used to predict malignancy of the lesion. Discriminant analysis shows the coefficients for III has the largest magnitude. Moreover, ROC analysis indicated that the III have the largest area under ROC curve. These results implied that the prediction power of the indentations is greater. The authors justified these findings by the fact that the regression, which is a major index of malignancy, generally causes indentations along the border of the lesion.

Besides the asymmetry measurement described in 2.3.3, circularity or *compactness index* (CI) has been also used to investigate the border irregularity of lesions [129, 113, 116, 137]. Holmstrom [123] recognises two disadvantages of using this method: First, this index is too sensitive to noise and any error in calculating the perimeter of the shape will be squared (according to the compactness formula defined in equation 4.14). Second, while the method effectively responds to textural irregularities, it fails to recognise structural irregularities, which have higher correlation with melanoma [133, 123]. Lee [138] also notices that the same compactness index can be obtained from lesions with different boundary shapes. Maglogiannis [139] terms this index *thinness ratio*. The author also introduces two other border irregularity features as follows:

$$Irregularity\ A = \frac{Perimeter}{Area} \tag{2.35}$$

$$Irregularity\ B = \frac{Perimeter}{Greatest\ Diameter} \tag{2.36}$$

Bono *et al.* [124] used *smoothness* to measure the border irregularity of the skin lesion. They defined this measure as the ratio of the perimeter of the lesion's convex hull to the perimeter of the lesion.

Ng *et al.* [140] applied *fractal dimension* (FD) to measure border irregularity along the border of the lesion. To this end, they use a box-counting algorithm which counts the

number of boxes of size r required to cover the lesion border. The relationship between number of boxes $N(r)$ and fractal dimension D is given by [128, 140, 134]:

$$N(r) = \lambda r^{-D} \tag{2.37}$$

Therefore,

$$\log \frac{1}{N(r)} = D \times \log r - \log \lambda \tag{2.38}$$

where λ is a constant related to the scale. Thus, D is the slope of equation (2.38) and $\log D$ is the intercept of the line. To calculate the fractal dimension D a line-fitting algorithm, *e.g.* least squares method, can be applied on the plot of $\log \frac{1}{N(r)}$ versus $\log r$. For some lesions, the resultant plot is made up of two line segments, having a breakpoint in the line's slope. Herein, Claridge *et al.* [128] consider two types of fractal dimensions: the *textural fractal dimension* delineating the global irregularities on the contour, and the *structural fractal dimension* (SFD) defining the fine irregularities along the border of the lesion. The above-mentioned breakpoint is where these two fractals (textural and structural) fall apart. Figure 2.6 [128] shows an example of plots and fractal dimensions for malignant and benign lesions. The vertical dashed line shows the breakpoint between structural and textural fractal dimension.

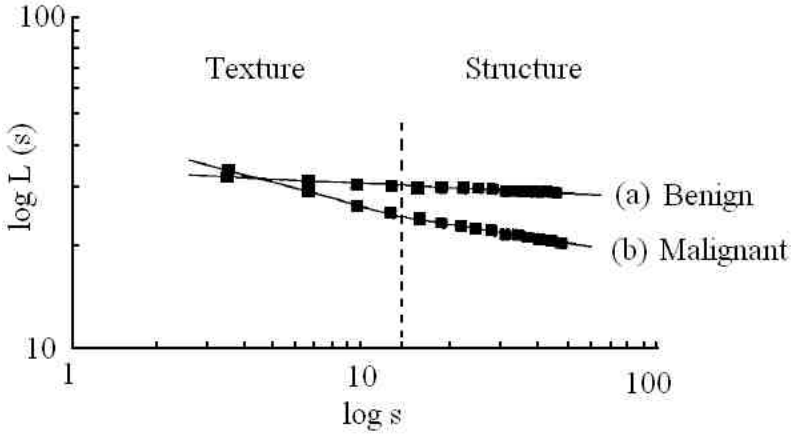

Figure 2.6: Structural and textural fractal dimension [128]

Change *et al.* [115] used Best-fit Ellipse and the following equation to quantify the border irregularity along the border of the skin lesion:

$$I = \frac{ab}{2\pi(a^2 + b^2)} \times \frac{P^2}{A} \tag{2.39}$$

where a and b are the lengths of the major and minor axes of the best-fit ellipse, and P and A are the perimeter and area of the lesion, respectively.

Aribisala *et al.* [141] proposed a measure of border irregularity based on conditional *entropy*. The authors defined entropy as the level of uncertainty or unpredictability of

a sequence or system. They have associated this concept to border irregularity, and concluded that the more irregular the border of the lesion, the more unpredictable the sequence of contour points. Conditional entropy is given by:

$$H(Y|X) = -\sum_{x\in X}\sum_{y\in Y} P_r(x,y)\log_2 P_r(y|x) \tag{2.40}$$

where X and Y are sequences of random variables of finite dimensions and $P_r(x,y)$ is the joint probability of X and Y for particular values of x and y, respectively. $P_r(y|x)$ is the conditional probability of Y, given X. Aribisala *et al.* [141] recognised two weaknesses of entropy and consequently any entropy-based measures (such as conditional entropy); First, it does not depend on the relative position of the elements in a sequence. Second, it rigorously depends on the bin size. To solve these problems, conditional entropy was extended into a new measure called *element-based conditional entropy* (EBCE). EBCE, denoted $H_e(X)$, which is defined as "the overall conditional entropy of the successive elements of X based on the preceding elements" given by:

$$H_e(X) = -\sum_{t=m}^{t=n} P_r(x_t,(x_{t-1},...,x_{t-m+1}))\log P_r(x_t|(x_{t-1},x_{t-2},...,x_{t-m+1})) \tag{2.41}$$

where $X = x_1, x_2, ..., x_n$ is a random sequence of length n and m is the length of different subsequences of X. Due to the importance of having proper values for m and useful definition for the conditionally variable q_t, the following notions were carefully chosen:

$$q_t = P_r(x_t|(x_{t-1},x_{t-2},...,x_{t-m+1})) = (x_t - x_{t-m+1}) \tag{2.42}$$

Considering a dynamic range for m, *i.e.*, $2 \leq m \leq n$, S, defined as the set of all subsequences of X each of length m, is given by:

$$S = \coprod_{i=1}^{i=n-m+1} \{x_i, x_{i+1}, ..., x_{i-m+1}\} \tag{2.43}$$

To apply the method on the lesion, the sequence of points on the lesion border is presented in polar coordinates, centred at the centroid of the lesion. Two groups of data were analysed; simulated and real data. For the simulated data, a Sine function in polar coordinates, equivalent to an ellipse corresponding to a benign lesion, was considered. Then, different amounts of Gaussian noise were introduced to the original signal, to make different noisy signals representing lesions with different degree of irregularities. The authors demonstrated that for both simulated and real data, EBCE increases exponentially with the increment of m.

Colour

Colour is another indicative feature in the classification of skin lesions, which is carefully considered in clinical practices such as the ABCD rule of dermatoscopy [15], Menzies

method [16] and the CASH algorithm [18]. We observed three different trends in assessing the colour feature in computer-aided diagnosis systems of melanoma: (i) developing a colour palette which comprises certain defined colour groups and associating different areas within each skin image with these colour groups, (ii) evaluating the relative colour histogram to define melanoma colours, and (iii) extracting statistical parameters from different colour channels. We present a more detailed review of each of these approaches in the following.

(i) colour Palette:

Seidenari *et al.* [142] established an approach to imitate the human perception of colours in skin lesions. In an interactive process incorporating dermatologists' opinions, they created a colour palette constituting of six colour groups (*i.e.*, black, dark brown, light brown, red, white and blue-grey). The palette involved different numbers of colour patches for each colour group, *e.g.* the light brown group comprises 28 different light brown patches, all being perceived as light brown by human observers. To test the method and detect the colour regions within the lesions, each pixel of the lesion was attributed to its corresponding colour patch. This correspondence was obtained by finding the colour patch with minimum Euclidean distance of RGB value to the pixel. After assigning all the pixels to the most relevant colour patch, those pertaining to the same group were connected to form the unified segment of that particular colour. Thus, for each lesion the number and type of the colours were investigated and the results compared with clinical data by using statistical analysis. They also applied discriminant analysis to investigate the accuracy of the proposed method in diagnosis of melanoma. The results revealed a high correlation between computer results and dermatologists' opinions, *i.e.*, both methods found that the black, white and blue-grey colours occurred more frequently in melanoma cases.

(ii) Relative colour Histogram Analysis:

Faziloglu *et al.* [143] presented a three-dimensional relative colour histogram analysis method to specify the colour characteristics pertaining to melanoma, termed as *melanoma colours*. Using relative colours in melanoma image analysis was first suggested by Umbaugh *et al.* [144] and has been widely used afterwards [115, 137, 143, 145, 146, 147]. The relative colour, in which the surrounding skin colour is subtracted from the lesion colour, aims to handle the colour distortion in the image acquisition. colour distortion occurs for several reasons; namely, variations in film type, camera settings and lighting condition, different colour skin and pigmentations, and also the Tyndall effect[8] for different patients. Faziloglu's method [143] was comprised of seven steps: first, a training set of melanoma and benign lesion images was obtained and the lesion boundary of each image was extracted. Then, they determined a representative colour value for the surrounding skin. This representative colour value was obtained by averaging the pixel values over a defined circular region surrounding the lesion, centred at the lesion's centroid, and defined by the following equation:

$$A_C = A_L + A_S \tag{2.44}$$

[8]The Tyndall effect is the visible scattering of light along the path of a beam of light as it passes through a system containing discontinuities.

where A_C is the circular area including lesion area A_L and the surrounding normal skin area A_S, given by:

$$A_S = \begin{cases} 4A_L & if \;\; A_L < 4700 \\ 3A_L & if \;\; 4700 \leq A_L < 10000 \\ 20000 & if \;\; 10000 \leq A_L \end{cases} \tag{2.45}$$

The radius of the circular region was determined by:

$$r_c = \sqrt{\frac{A_C}{\pi}} \tag{2.46}$$

The next step in Faziloglu's method is building the relative-colour histogram and analysing it to determine the melanoma colours. The relative colour was obtained by:

$$\vec{R} = \begin{bmatrix} r_{rel} \\ g_{rel} \\ b_{rel} \end{bmatrix} = \vec{C}_{les} - \vec{C}_{skin} = \begin{bmatrix} r_{les} - r_{skin} \\ g_{les} - g_{skin} \\ b_{les} - b_{skin} \end{bmatrix} \tag{2.47}$$

where $\vec{C}_{les}$ is the lesion colour value, *i.e.*, $\vec{C}_{les} = Col.(r_{les}, g_{les}, b_{les})$, $\vec{C}_{skin}$ is the surrounding skin colour and $\vec{C}_{rel}$ denotes the relative colour, bounded by $-255 \leq r_{rel} \leq 255$, $-255 \leq g_{rel} \leq 255$ and $-255 \leq b_{rel} \leq 255$. The resulting relative colour values were translated to be in the range of 0–510 by subtracting the minimum possible relative value of -255 from each pixel value. To produce the histogram the following bin value was selected:

$$bin = \text{truncate}(\frac{r_{rel} + 255}{4}) \tag{2.48}$$

This division by four for the three components combines 64 bins of the original relative colour histogram into a single bin, yielding a $128 \times 128 \times 128$ histogram. This quantisation was done to avoid having a sparsely populated histogram. At this stage, the histogram was populated using the training image set. For a bin to be populated, they considered the criterion of $K = 0.00125A_L$; showing the minimum number of pixel voting into that particular bin. By applying this strategy, all populated bins for melanoma images are found. Afterward, for each histogram bin the total number of images populating that bin was counted. This counting process was done for benign lesions and melanoma images of the training set, separately and then combined for labelling the bins as follows. Assume $\vec{B}$ being the set of all relative histogram bins (*i.e.* 128^3), each bin $\vec{B}_k \in \vec{B}$ maintains a benign colour bin probability $P(\vec{B}_k|benign)$ and a melanoma colour bin probability $P(\vec{B}_k|mel)$, which is defined as the number of benign/melanoma images populating $\vec{B}_k$. Accordingly, the following mapping was defined:

$$C(\vec{B}_k) = \begin{cases} \text{unpopulated} & for \;\; P(\vec{B}_k|mel) = P(\vec{B}_k|benign) = 0 \\ \text{melanoma} & for \;\; P(\vec{B}_k|mel) > P(\vec{B}_k|benign) \\ \text{uncertain} & for \;\; P(\vec{B}_k|mel) = P(\vec{B}_k|benign) \neq 0 \\ \text{bening} & for \;\; P(\vec{B}_k|mel) < P(\vec{B}_k|benign) \end{cases} \tag{2.49}$$

According to the melanoma colour mapping in the above equation, depending on melanoma and benign probabilities, each bin was labeled as melanoma, benign, uncertain or unpopulated. The uncertain and unpopulated bins underwent a relabelling process using an iterative region growing technique. The colour feature called *percent melanoma* feature was then defined as the percentage of melanoma-coloured pixels, given by:

$$C = \frac{100M}{A_L} \tag{2.50}$$

where M is the total number of pixels within the lesion falling into a melanoma-labelled bin in the relative colour histogram, and A_L denotes the lesion area. The last step of the algorithm was to calculate this feature for all the training images and an automated classification schema performed, out of which the classification threshold C_T was determined. This threshold was selected such that; true positives = true negatives. The proposed technique was applied on an image set of 256 lesion images, resulting in an average rate of 84.6% correct melanoma diagnosis and 83% correct diagnosis of benign lesion. The authors [143] acknowledged that in spite of acquiring reasonably good classification results, colour features have to be combined with other features to decrease the false negative and false positive rates.

Chen *et al.* [145] extended the relative-colour-histogram analysis to evaluate lesion classification based on colour feature computation in different regions of the skin lesion. To this end, beside *percent melanoma colour* proposed in the previous related study [143], they introduced the *colour clustering ratio*, which provides a measure of what percentage of the lesion pixels are melanoma-coloured. Similar to Faziloglu's method [143], the training set was obtained, the lesion boundary was determined, and the relative-colour histogram was generated, using which the colour characteristics of melanoma were identified. Two features; *percent melanoma colour* and *colour clustering ratio* were calculated over different regions of the skin lesion. They defined two regions within the lesion; *boundary-area percentage* (the lesion region from the boundary towards its centroid which holds a certain percentage of the lesion area) and *offset boundary-area percentage* (the lesion region from the boundary towards outside which holds a certain percentage of the lesion area). The authors calculated the percent melanoma colour feature via Equation 2.50 noting that this feature does not consider the spatial distribution of the colour, whereas the colour-clustering ratio provides information about the grouping of melanoma colours within the lesion. Experiments were performed on an image set comprising 129 malignant and 129 benign cases. *Percent melanoma colour* and *colour-ratio clustering* features were computed for different values of *boundary-area percentage* and *offset boundary-area percentage*. The results showed that regions close to the boundary of the lesion contain the most useful colour feature information. Moreover, the *percent melanoma colour* feature produced higher classification results than the *colour-clustering ratio feature*, across different regions of the skin lesion.

Chang *et al.* [115] also applied the *percent melanoma colour*, *percent non-melanoma colour* and *colour-clustering ratio* features in their melanoma diagnosis system. The relative-colour histogram analysis done by [115, 143, 145, 146] were all applied on clinical

images, whereas Stanley *et al.* [147] applied the relative-colour histogram to dermoscopy images. A similar process to that performed in [115] was applied in [147]. However, differing results were obtained: *i.e.*, the *colour clustering ratio* (structured colour information) revealed a higher discriminant power than that yielded from the *percent melanoma colour feature* (unstructured colour information). The study also showed that the interior regions of the dermoscopy images contained important colour information, while peripheral regions were identified as the crucial areas of the clinical images.

(iii) Statistical Parameters:

In order to quantify the colour features of the skin lesion, various statistical parameters of different colour spaces have been used. Celebi *et al.* [60] measured mean and standard deviation statistical parameters over the three channels of six different colour spaces; namely, RGB, rgb (normalised RGB), HSV, I1/I2/I3 (Ohta space), L1/2/3 and $CIEL^*u^*v^*$. The *mean* parameter represents the average colour and the *standard deviation* shows the colour variegation over each channel. A total number of 108 colour features were calculated: (6 colour spaces) $\times$ (3 channels in each colour space) $\times$ (2 parameters: mean and std) $\times$ (3 regions: lesion, inner periphery, outer periphery). Furthermore, the differences and ratios of the two parameters (mean and std) over the three defined regions (lesion, inner periphery, outer periphery) were computed: (outer/inner), (outer/lesion), (inner/lesion), (outer/inner), (outer/lesion), and (inner-lesion). These relative parameters revealed significant information: *e.g.* a big difference between inner and outer periphery might indicate malignancy of the lesion.

Sboner *et al.* [81] analysed the colour features of the image in RGB and HSV colour space and calculated the *mean* and *standard deviation* of hue and saturation components. They also computed the fractal dimension of blue, green, hue and saturation channels. Moreover, they measured the distances between the centroids of red, green, blue, hue, saturation and value channels and the lesion centroid, average distance from the centroid of the RGB and HSV channels from the lesion centroid, and also the difference, mean and standard deviation of hue and saturation channels over the lesion border. The automated diagnosis was made based on a voting schema integrating the outputs of three classifiers; namely, discriminant analysis, k-nearest neighbourhood and decision tree. The system was trained and validated on a set of 152 skin images and the results obtained from the system were compared with the diagnosis obtained from a group of eight expert dermatologists. Experimental result showed that the proposed system performs as well as the eight dermatologists with respect to sensitivity. Regarding the specificity, the system performed better than at least one of the physicians. Furthermore, the systems never exhibited poor performance with respect to both sensitivity and specificity, compared to any of the dermatologists.

In another study [148] the *average* and *variance* of pixels in three colour spaces; RGB, HIS (Hue, Intensity, Saturation) and spherical coordinates CIE L*a*b were measured. The two latter colour spaces are obtained from RGB colour space by the following equations:

$$
\begin{aligned}
I &= \frac{R+G+B}{3} \\
S &= 1-\frac{3}{R+G+B}[\min(R,G,B)] \\
W &= Cos^{-1}(\frac{R-\frac{1}{2}(G+B)}{[(R-G)^2+(R-G)(G-B)]^{1/2}}) \\
\text{where} & \\
H &= \begin{cases} W & if \;\; G>B \\ 2\pi - W & if \;\; G<B \\ 0 & if \;\; G=B \end{cases}
\end{aligned} \tag{2.51}
$$

and

$$
\begin{aligned}
L &= \sqrt{R^2+G^2+B^2} \\
AngleA &= Cos^{-1}[\frac{B}{L}] \\
AngleB &= Cos^{-1}[\frac{R}{L \times Sin(AngleA)}]
\end{aligned} \tag{2.52}
$$

She *et al.* [116] defined the colour variegation as the normalised standard deviation of red, green and blue colour channels by the following equations:

$$
\begin{aligned}
C_r &= \frac{\sigma_r}{M_r} \\
C_g &= \frac{\sigma_g}{M_g} \\
C_b &= \frac{\sigma_b}{M_b}
\end{aligned} \tag{2.53}
$$

where σ_i and M_i are the standard deviation and maximum value of the i colour channel (i: red, green, blue), respectively. The authors put great emphasis on the normalisation process due to fact that in skins with higher pigmentation of surrounding normal skin, the lesion also exhibits higher pigmentation values. The above-mentioned colour-based features were combined with features of asymmetry, border irregularity, diameter, as well as measuring the skin line direction and intensity, resulted in area under ROC curve of approximately 0.94.

Furthermore, several colour parameters were suggested by Manousaki *et al.* [122]; namely, minimum, maximum, mean and standard deviation values of grey, red, green and blue intensities. They also measured the skewness from Gaussian curve over each colour channel, *i.e.*, the deviation of the histograms of the grey, red, green and blue colour channels from the normal (Gaussian) distribution curve. These colour features were combined with geometry, border and texture measures resulting in a total of 43 variables, from which five was selected through a feature-selection process based on logistic regression. The system resulted in 60.9% sensitivity and 95.4% specificity.

Cheng *et al.* [137] also applied a pattern recognition software toolkit to classify clinical skin lesions using relative colour features. They extracted different histogram statistical features; namely, *mean* (for image brightness in red, blue and green colour channels), *standard deviation* (for image contrast in red, blue and green colour channels), *skewness* (indicating the asymmetry about the mean in intensity level distribution in red, blue and green colour channels), *energy* (indicating the distribution of the intensity level in red, blue and green colour channels) and *entropy* (indicating the number of bits required to code image data in red, blue and green colour channels). The best results were obtained using a multi-layer perceptron neural network model, resulting in 79% accuracy.

Differential Structure

In this section we briefly review the existing methods on analysing differential structures in dermoscopy images.

Grana *et al.* [149] proposed a method based on Gaussian-derivative kernels and Fisher linear-discriminant analysis to detect and localise network patterns. An experiment was conducted on a set of 60 skin images which had been categorised into three classes of: no network pattern, partial network pattern and complete network pattern by dermatologists. The results revealed no false positives, five false negatives on the partial network class and two on the complete network class, with overall accuracy of 88.3% for network pattern recognition. Fischer *et al.* [150] proposed a multi-step approach consisting of transforming RGB colour space by the Karhunen-Loeve transform [151] and applying colour segmentation on the new colour components, with histogram equalisation and greyscale morphology to enhance and filter pigment network patterns. Another algorithm was suggested by Fleming *et al.* [152], which aimed to analyse pigment networks and also measure the network lines' width and network holes' size. Murali *et al.* [153] developed a texture-based algorithm to detect blotchy areas. They also aimed to determine the degree of importance of this feature in differentiating between malignant and benign lesions. The proposed technique involved calculating the neighbouring grey-level dependence matrix (NGLDM) and was tested on 37 dermatoscopy images. Celebi *et al.* [154] provided a machine-learning approach to detect blue-white areas in dermoscopy images. The approach was comprised of several steps; preprocessing (determining the background skin colour and selecting the training and test sets), colour-feature extraction (absolute and relative colour features), pixel classification (decision-tree induction), rule application, and postprocessing. The algorithm was tested on 224 images and the results visually assessed by dermatologists as satisfactory. Stoecker *et al.* [155] analysed areas of granularity (or multiple blue-grey dots defined as accumulation of tiny, bluegrey granules in dermoscopy images) to discriminate granularity in melanoma from visually similar areas in non-melanoma skin lesions. The granular areas in dermoscopy images of 88 malignant (including 14 melanomas in-situ and 74 invasive melanoma) were manually selected. For 200 non-melanoma dermoscopy images, those areas which most closely matched granularity, in terms of size, colour and texture were similarly selected. The texture measures, including the average and range of energy, inertia, correlation, inverse

difference, and entropy (based on co-occurrence matrix), and colour measures, including absolute and relative RGB averages, absolute and relative RGB chromaticity averages, absolute and relative G/B averages, CIE X, Y, Z, X/Y, X/Z and Y/Z averages, R variance, and luminance were measured for each granular area of the melanomas and similar areas in the non-melanoma images. ROC analysis revealed that the best classification result is obtained by using a combination of six texture and five relative-colour measures, resulting in an accuracy of 96.4%.

In chapter 4 of this thesis we propose a novel, comprehensive and highly effective feature extraction method which combines different types of features; novel ones and a few adopted from the above-mentioned studies. In the proposed method texture and border (including geometry and time-series) features are extracted and optimally combined.

2.3.4 Feature Selection

Feature selection is an important component in data preprocessing/analysis in various applications such as computer vision, data mining, image mining, etc. It reduces the number of features by removing irrelevant, redundant, or noisy data, and has an immediate effect on application by accelerating the classification, clustering or data mining algorithm, as well as improving performance [3]. A typical feature selection process consists of four basic steps [156] (shown in Figure 2.7), namely, subset generation, subset evaluation, stopping criterion, and result validation. Subset generation is a search procedure which produces candidate feature subsets based on a certain search strategy [157, 158]. Each candidate subset is then evaluated and, based on a certain evaluation criterion, is compared with the previous best subset. If the new subset defeats the previous one, the old subset is replaced by the new one. The process of producing and evaluation the subsets is repeated until a given stopping criterion is satisfied. Ultimately, the selected best subset will be validated and testified by prior knowledge or different tests using simulated or real-world data [3].

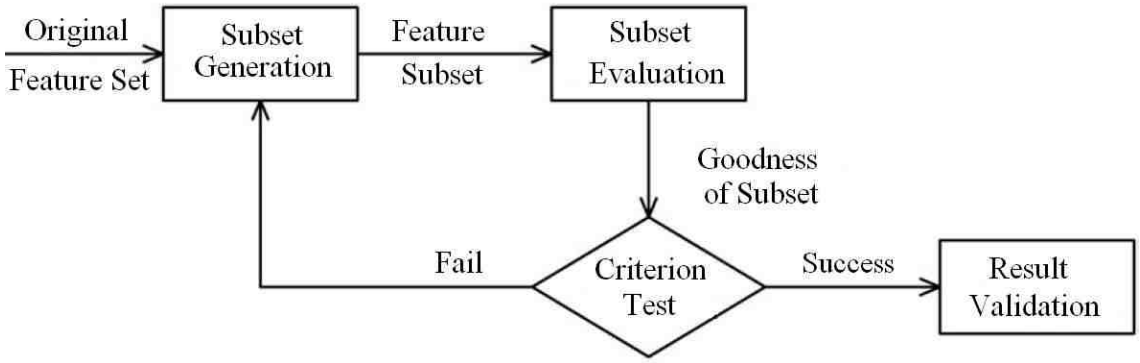

Figure 2.7: Basic steps of a typical feature selection process [3].

Feature selection algorithms can be categorised into three classes [3]: (i) the filter model [159], which uses the general characteristics of the data to evaluate the features and selects the optimal feature subset(s) with no classification or learning algorithm, such as methods based on correlation [160, 161], entropy, mutual information, etc. (ii) the wrapper model [162], which uses a learning algorithm and searches for features which improves the learning performance, such as methods based on greedy or genetic algorithms [163, 164, 165], (iii) the hybrid model which combines the two approaches [166, 167, 168]. Filter model feature selectors are less computationally expensive than wrapper-based approaches.

Some existing computer-based melanoma diagnosis systems have applied feature selection in their study; for example, Principal component analysis was used by Walvick *et al.* [169] to obtain an optimal set of four features from an initial set of eleven features. Roβ *et al* [170] performed feature selection by applying the sequential forward selection (SFS) method, reducing the feature set from 87 to five features. Nimunkar *et al.* [171] applied a statistical feature selection approach (unpaired t-test for normally distributed features and Wilcoxon rank-sum test for non-normally distributed features), to cut down the feature vector from 34 to five optimal features. In [80] a neural network was applied and node pruning was used to reduce the set size from 21 features to six. Ganster *et al.* [58] started with 122 features and used three statistical approaches of Sequential Floating Forward Selection (SFFS), Sequential Floating Backward Selection (SFBS), and Leave One Out (LOO) to reduce the number of the selected features to 21. Zagrouba and Barhoumi [94] also applied SFFS and SFBS obtaining an optimal set of five features (from an initial 14), with which they obtained an accuracy of 77.7% with half of the computational time compared to using all 14 features. Iyatomi *et al.* [172] developed an automated method to detect the dermoscopic criteria for diagnosing melanomas, based on two diagnostic algorithms of the ABCD rule and the 7-point checklist. A total of 356 features were extracted from a set of 105 dermoscopy images and their dermoscopic findings determined by four dermatologists were used as the gold standard. They applied correlation-based feature selection to reduce the feature space to 163 features and built multiple-regression models for the 15 clinical findings defined by the two diagnostic algorithms. The model reportedly provides results comparable to dermatologists in recognising almost all of the dermoscopic criteria.

In Chapter 4 of this thesis we discuss the application of a feature selection scheme from the filter-based family. The rationale behind this choice is the high-dimensional feature space and the necessity of applying a low-computational feature selection method.

2.3.5 Classification

Classification is the process of distributing items into classes or categories of similar type. In a computer vision system, similarity is defined in terms of the features which have been extracted in the feature extraction module. This section provides a brief review of the most popular classification methods which have been applied to computer-based melanoma diagnostic systems.

Discriminant Analysis

Discriminant analysis has been employed to classify pigmented skin lesions in [171, 81, 113, 139, 173, 174, 175]. The main aim of discriminant analysis [176] is to classify a set of observations into predefined classes, based on the values of certain measurements, known as predictors. The discriminant analysis discovers the combination of these predictor variables which best characterises the interclass differences. The coefficients are calculated and the obtained discriminant functions can be used to predict the class of a new observation with unknown class. For a k-class problem k discriminant functions are constructed. Given a new observation, all k discriminant functions are evaluated and the observation is assigned to class i if the $i-th$ (i=1,2,...,k) discriminant function has the highest value. Another application of this technique is to find out the most discriminating predictor, which is either done by successively eliminating those predictor variables that do not contribute significantly to the discrimination between classes, or by successively identifying the predictor variables that have a major contribution. The Likelihood function and Fisher's Linear discriminant function are two common discriminant rules which are employed in this technique.

Artificial Neural Network

The Artificial Neural Network (ANN) approach has been applied to classify skin-lesion images in [169, 59, 94, 113, 139, 170, 177]. Neural networks are non-linear statistical data modeling tools which can be used to model complex relationships between inputs and outputs or to find patterns in data. The methodology of neural networks involves mapping a large number of inputs into a small number of outputs; therefore it is frequently applied to classification problems [178]. The artificial neural network or neural network (NN) [179] is a mathematical or computational model which calculates posterior class membership probabilities by minimising a cross-entropy error function. Based on biological neural networks, ANNs consist of highly interconnected processing units called artificial neurons. Information is propagated between neurons positioned in different layers and stored as connection weights between neurons. A minimisation process is implemented as a rule which updates the weights in the network. An ANN is generally an adaptive system that changes its structure based on external or internal information (feedback) that flows through the network during the learning phase.

K-Nearest Neighbourhood

The k-nearest-neighbourhood (K-NN) algorithm has been used to classify pigmented skin lesions in [58, 59, 81, 170, 175]. The k-NN algorithm [180] is a popular density-estimation algorithm for numerical data. It classifies objects based on closest training examples in the feature space. This algorithm uses the elements of the training set to estimate the density distribution of the data by applying a distance measure such as Euclidean distance. For a given distance measure, the only parameter of the algorithm is k, the

number of neighbours, which determines the smoothness of the density estimation; larger values consider more neighbours, and therefore smooth over local characteristics. Smaller values consider only limited neighbourhoods.

Support Vector Machine

The support vector machine has been applied to skin lesion classification in [12, 59, 60, 113, 148, 181, 182, 65]. The support vector machine (SVM) [183] is a popular classification algorithm based on statistical learning theory. Considering input data as two sets of vectors (support vectors) in an n-dimensional space, SVM calculates a separating hyperplane in that space, which maximises the distance (margin) between the two data sets. Celebi *et al.* [60] recognised several advantages of SVM over other classification algorithms such as neural networks. The advantage arises from the logic that since the training phase of SVM involves optimisation of a convex cost function, there is no risk of local minima to complicate the learning process as in the case of backpropagation neural networks.

Decision Trees

Decision trees have been used to classify skin lesion images in [12, 59, 81, 184, 185]. The decision tree as described in [186, 187] provides a classification schema by dividing the data set into smaller and more uniform groups, based on a measure of disparity (usually entropy). To achieve this separation, a variable and a corresponding threshold in the domain of this variable are identified, which can be used to divide the data set into two groups. The best choice of variable and threshold is that which minimises the disparity measures in the resulting groups. In these tree structures, leaves represent class labels and branches represent conjunctions of features that lead to those classifications.

Comparison of Different Machine Learning Methods

Dreiseitl *et al.* [59] utilised five machine learning (classification) techniques; namely, k-nearest neighbourhood, logistic regression, artificial neural network, decision tree, and support vector machine, to classifying pigmented skin lesions as common nevi, dysplastic nevi, or melanoma. Regarding the experimental results, the authors recognised that the decision tree paradigm was not a good option to apply on skin lesion classification, since most of the variables (features) basically have continuous values in such applications. However, k-NN performed well, and logistic regression, ANN and SVM claimed to exhibit very good performance over the given image set. Another comparison between different classification methods was performed by Burroni *et al.* [175], where the performance of two statistical classifiers, linear discriminant analysis and k-nearest neighbourhood, was evaluated. The experiments were conducted on an image set containing 391 melanoma and 449 melanocytic nevi images. Using independent test sets of lesions, the results achieved by the linear classifier and the K-nearest neighbourhood classifier revealed a mean sensitivity

of 95% and 98% and a mean specificity of 78% and of 79%, respectively.

Sboner *et al.* [81] used a voting schema integrating the outputs of three different classifiers of discriminant analysis, k-nearest neighbour and decision tree, applied to a set of 152 skin images. The authors claimed that the voting scheme increases the diagnostic sensitivity and specificity over applying each individual classifier.

Maglogiannis and Doukas [12] presented a survey on existing computer vision systems for characterisation of skin lesions in 2007 (published in 2009). They also utilised a fairly large set of 3707 images, extracted 31 features of colour, border and texture from those, and applied three feature selection methods based on correlation, principal component analysis and GSFS (generalised sequential feature selection), aiming to (i) detect melanoma from nevus, (ii) classify dysplastic (nevus) versus nondysplastic skin lesions, (iii) classify the three classes of melanoma, dysplastic, and nondysplastic. A group of classifiers were applied, *e.g.* Bayes networks, SVM, CART (classification and regression trees), multinomial logistic regression (MLR), locally weighted learning (LWL), neural network (multilayer perceptron), etc. The result showed that regarding the first aim, MLR had the best performance followed by SVM, LWL, and CART. For the second aim, all classifiers achieve lower accuracy, indicating the difficulty of differentiation between dysplastic and nondysplastic nevi. Among all the three classifiers of multilayer perceptron, SVM and Bayes networks had the best performance. In the last experiment the majority of the examined classifiers performed satisfactorily with accuracy ranges from 68.70% (Bayes networks) to 77.06% (SVM). The authors emphasised the overall superior performance of SVM, followed by regression and Bayes networks.

Table 2.2 shows the classification results of some of the existing computer-based melanoma diagnosis systems. Obviously the results obtained from those studies with a reasonable number of images (*e.g.* at least 100), with balanced distribution between classes of benign and malignant, and with separated image sets for training and testing the classifier, provide more validity and reliability.

In Chapter 4 of this thesis we explain the classification methods applied in our proposed diagnostic system, where we have applied four different classes of classification techniques; namely, Random Forest (RF) [188], Logistic Model Tree (LMT) [189], Hidden Naive Bayes (HNB) [1], and Support Vector Machine (SVM). To the best of our knowledge, the first three classifiers (RF, HNB and LMT) have not been utilised in melanoma diagnosis applications before.

2.4 Summary

In this chapter we provided an overview of the problem in hand; skin structure and its cellular distortion which leads to melanoma. Moreover, we looked into existing solutions for early diagnosis of this skin cancer, including imaging technology and medical diagnostic algorithms. This was followed by surveying the existing computer-based melanoma diagnostic systems which have been previously developed in order to assist dermatologists

Table 2.2: Classifier performance results from existing systems.

Classifier	Performance Results	Nom. of Images	Reference
Discriminant Analysis	TP 93% , FP 0–21%	28	[171]
	SNS 65% , SPC 83%	152	[81]
	SNS 93% , SPC 100%	34	[113]
	ACC 71%	70	[173]
	SNS 100%, SPC 86%	29	[174]
	SNS 95% , SPC 78%	821	[175]
Artifical Neural Network	TP 100% , FP 23.5%	60	[169]
	SNS 91% , SPC 94%	1619	[59]
	ACC 77.7%	-	[94]
	SNS 93% , SPC 100%	34	[113]
	ACC 87%	-	[170]
	ACC 85–89%	250	[177]
K-Nearest Neighbourhood	SNS 87% , SPC 92%	5393	[58]
	SNS 85% , SPC 90%	1619	[59]
	SNS 35–68% , SPC 90–97%	152	[81]
	ACC 93%	-	[170]
	SNS 98% , SPC 79%	821	[175]
Support Vector Machine	SNS 92% , SPC 95%	1619	[59]
	SNS 93% , SPC 92%	564	[60]
	SNS 86% , SPC 100%	34	[113]
	ACC 82.5%	4277	[181]
	SNS 100% , SPC 63.5%	977	[182]
	SNS 92& , SPC 93%	60	[65]
Decision Trees	SNS 80% , SPC 90%	1619	[59]
	SNS 64% , SPC 84%	152	[81]
	ACC 70%	251	[184]

and clinicians and facilitate the early accurate diagnosis of the disease. We also reviewed some of the methods applied in those studies in order to develop an automated diagnostic system. In the next two chapters we propose and detail various components of our computer-aided diagnostic system for melanoma.

Chapter 3

Segmentation or Border Detection

Segmentation or border detection is a fundamental step towards development of a computer-aided diagnosis of melanoma. It involves separating the lesion from the background skin. The accuracy of the detected border is essential for accurate implementation of the subsequent parts of the system, *i.e.*, feature extraction and classification, since diagnostic features such as asymmetry and border irregularity are highly dependent on the border detection result. In addition, the lesion inside the detected border reveals information about homogeneity, dermoscopic patterns and lesion colour, and excluding any portion of the lesion may result in significant information overlooked and possible false diagnosis.

Detecting the boundary of the lesion in dermoscopy images is challenging [78] due to (1) the low contrast between the lesion and the background skin, (2) presence of unwanted artifacts within the image such as hairs, skin lines, blood vessels, air bubbles, black frames, camera scale and blue or purple surgical markings, (3) fuzziness of the border and (4) the colour variegation within the lesion.

Various image features such as shape, colour, texture, and luminance have been employed to perform skin lesion segmentation. Numerous border detection methods have been proposed in the literature [78] including histogram thresholding [190, 80], colour clustering [87, 82, 83], JSEG algorithm based on colour quantisation and spatial segmentation [95], statistical region merging [96], two-stage k-means++ clustering followed by region merging [89], etc. Melli *et al.* [87] criticised adaptive thresholding methods for inaccuracy due to the lack of sharp bimodal luminance distribution between the surrounding skin and the lesion. However, in this study we demonstrate that with a proper choice of colour channel, a thresholding-based approach can produce borders highly accordant with those manually drawn by expert dermatologists. The rationale behind this choice is the simplicity and low computational cost of scalar processing as described in the following sections.

In Section 3.1 of this chapter, a novel border detection method based on colour-space analysis and clustering-based histogram thresholding is proposed. The method first determines the most effective and discriminative colour channels in dermoscopy applica-

tion, then applies global thresholding followed by a set of pixel-based computations and morphological operations to detect the boundary of the lesion.

In Section 3.2, the proposed border detection method is improved by proposing the Hybrid border detection method which integrates local information around the initial borders detected by the global thresholding method to expand the initial boundary towards the background skin. Both global and local histogram thresholding are applied to the optimised colour channels investigated in first part of the chapter.

The Section 3.3 involves a comparative study between the Global and Hybrid approaches proposed in the first two sections of the chapter. The purpose is to evaluate the necessity of integrating local pixel information into the segmentation process and demonstrate the superiority of the Hybrid thresholding approach over the Global method through a newly introduced evaluation metric, *Performance Index*.

The next two sections (Sections 3.4 and 3.5), present a comprehensive objective evaluation of border-detection results. Two quantitative metrics of *Weighted Performance Index* and *Optimised Weighted Performance Index* are proposed as useful means for assessing performance of the border detection methods as well as comparing between the different border detection methods. Section 3.6 provides a summary and conclusion of the chapter.

3.1 Global Thresholding on Optimised Colour Channels

This section proposes a novel automatic segmentation algorithm based on colour-space analysis and clustering-based histogram thresholding. It determines the most effective and discriminative colour channel for melanoma application. Different colour channels from various colour spaces have been incorporated to maximise the discrimination between two clusters of pixels within the image, lesion versus normal skin pixels, to obtain a more accurate histogram analysis. Each colour channel undergoes preprocessing such as intensity adjustment and noise removal by the use of a pill-box low-pass filter. Clustering-based thresholding is applied, and a set of pixel-based computations and morphological operators is utilised to obtain the segmented lesion. Producing 24 different channel images from each initial RGB skin image, we have determined the colour space and the corresponding channel that outperforms others in detecting the lesion borders. The proposed border detection method consists of several steps, as shown in Figure 3.1, described in the following.

3.1.1 Hair Removal

Lesions occluded with dark thick hairs can cause problems in the segmentation process. In such cases, our proposed method starts with a hair-removal procedure, consisting of the following steps [111]: (1) localising dark hairs using a morphological closing operation in the vertical, horizontal and diagonal directions, (2) interpolating the removed hair pixels

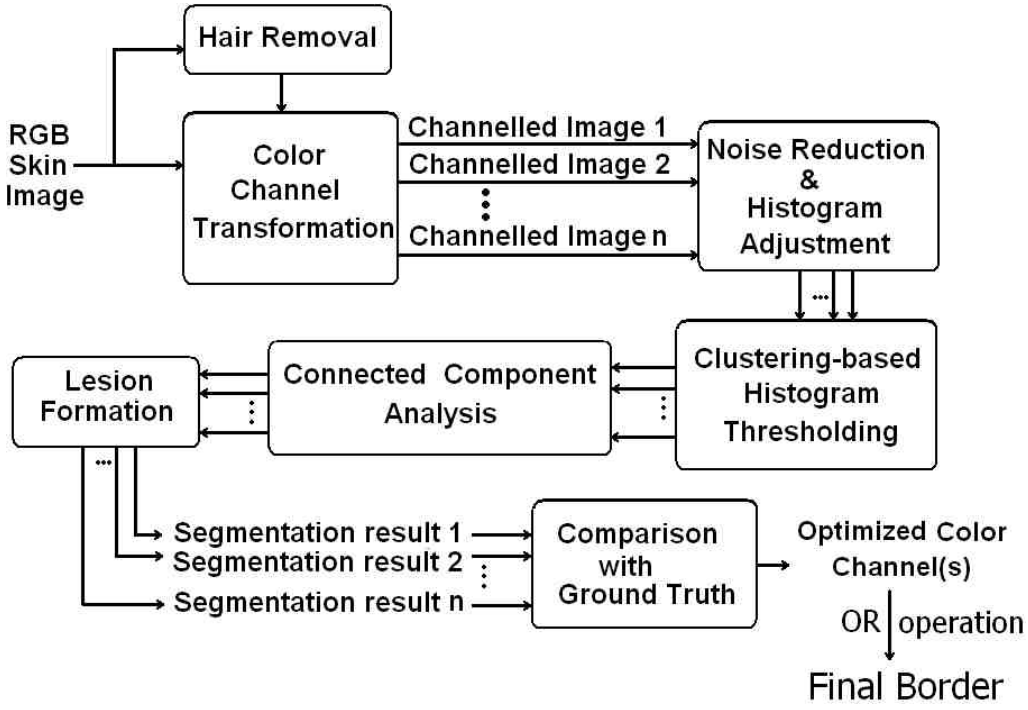

Figure 3.1: Global thresholding method for border detection in dermoscopy images encompassing colour optimisation process.

using nearby non-hair pixels, and (3) smoothing the final result using a median filter to eliminate remaining thin lines.

3.1.2 Colour-Space Transformation

Colour information plays a significant role in dermoscopy image analysis. This step incorporates colour information in a dermoscopy image into the segmentation process, where the original RGB image is transformed into various colour spaces, and the corresponding colour channels are extracted. Despite the existence of numerous colour spaces in the literature, the choice of optimal colour space is application-dependent [191]. In this study we investigated various colour spaces [192]; namely, RGB, HSV, HSI, CIE-XYZ and CIE-LAB. The original RGB dermoscopy image is transformed into a set of 24 colour channel images from the above-mentioned colour spaces. The selection of these 24 colour channels was made after analysing visually all possible combinations. As shown in Table 3.1, these images include single colour channel images such as R from the RGB colour space and X from the XYZ colour space, as well as combinations of them such as XoYoR, where "o" stands for logical OR, which combines the results from X and Y colour channels from the XYZ colour space with those from the R colour channel from the RGB colour space. For instance, to build the RoGoB segmentation result the R, G and B colour channels are

extracted and each undergoes the segmentation process illustrated in Figure 3.1. Finally, the three binary segmentation results are combined through the OR operation. The RGB colour channel in Table 3.1 is calculated by forming a weighted sum of the R, G, and B components (expressed in Equation 3.1, adopted from Matlab) and converting RGB values to greyscale values. Figure 3.2 shows a dermoscopy image before and after colour-space transformation. In the following a brief explanation about colour spaces and the corresponding colour channels is provided.

$$RGB = 0.2989 \times R + 0.5870 \times G + 0.1140 \times B \tag{3.1}$$

Table 3.1: Colour channels used in colour-space transformation.

Channel No.	**Colour Channel(s)**	**Colour Space**
1-12	R, G, B, RGB, RoB, GoB, RoG, RoGoB, RGBoR, RGBoG, RGBoB, RGBoRoGoB	RGB
13	I	HSI
14	V	HSV
15	L	LAB
16-22	X, Y, Z, XoY, XoZ, YoZ, XoYoZ	XYZ
23-24	XoYoR, XoYoZoR	XYZ and RGB

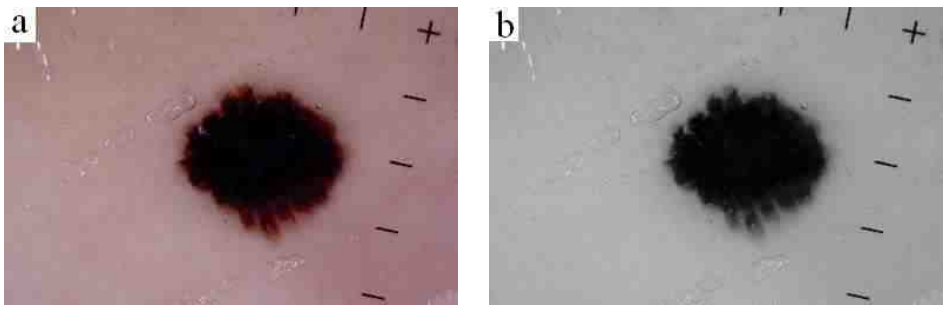

Figure 3.2: Colour-space transformation for a sample dermoscopy image: (a) original RGB image (b) channeled image.

Colour Spaces and Colour Channels:

The most frequently used colour presentation in image processing is the RGB colour space where colours are represented by their red, green, and blue spectral wavelength responses in an orthogonal Cartesian space. Consequently, each pixel has a 3D vector with component values ranging from 0 to 255. To overcome the limitation of the RGB colour space in high level processing, other colour spaces have been developed based on mathematical

transformation of the original RGB colour channels. HSV and HSI colour spaces are common cylindrical-coordinate representations of points in an RGB colour model, which rearrange the geometry of RGB in an attempt to be more perceptually relevant than the Cartesian representation, mimicking the human visual perception of colour in terms of *hue*, *saturation* and *intensity* (in HSI) and *hue*, *saturation* and *value* (in HSV). The hue component is proportional to the average wavelength of the colour, saturation indicates the amount of white in the colour and intensity represents the brightness or the amount of energy in the colour.

The I colour channel from the HSI colour space, and V from the HSV colour space, which are used in this study (13 and 14th colour channels in Table 3.1) are define as follow:

$$I = \frac{R+G+B}{3} \tag{3.2}$$

$$V = max\{R, G, B\} \tag{3.3}$$

However, the colour spaces outlined above do not provide perceptual uniformity; *i.e.*, colour vectors with a particular distance in colour space are placed in the same distance when perceived by human visual system. To meet the uniformity requirement the CIE-LAB colour space was proposed. CIE-XYZ (Equation (3.4) and (3.5)) [192] colour space, which is one of the first mathematically defined colour spaces created by the International Commission on Illumination (CIE) [193] in 1931, is another colour space investigated in this study. The transformation from RGB colour space to XYZ colour space can be achieved through the following relationship:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = T \times \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{3.4}$$

where T is a 3×3 constant matrix, the parameters of which may differ from one application to another according to the reference white value. In our study we have chosen the following parameters, for the illuminant D65:

$$T = \begin{bmatrix} 0.4125 & 0.3576 & 0.1804 \\ 0.2127 & 0.7152 & 0.0722 \\ 0.0193 & 0.1192 & 0.9502 \end{bmatrix} \tag{3.5}$$

The L colour channel from the LAB colour space, used in this study (15th colour channel in Table 3.1) is given by:

$$L = 100\sqrt{\frac{Y}{Y_n}} \tag{3.6}$$

where $Y_n = 100.00$ and Y is the CIE value 3.5.

3.1.3 Noise Filtering

To enhance the accuracy of segmentation and save computational time, it is useful to eliminate the artifacts that might be present in the image. In dermoscopy images, external

artifacts include skin lines, air bubbles or other random noise caused by the imaging process. In this study, the dermoscopy images are smoothed using a circular averaging low-pass filter with a radius of 5, using the pill-box point spread function shown in Figure 3.3. The result of applying this filter on the dermoscopy image of Figure 3.2 is shown in Figure 3.4.

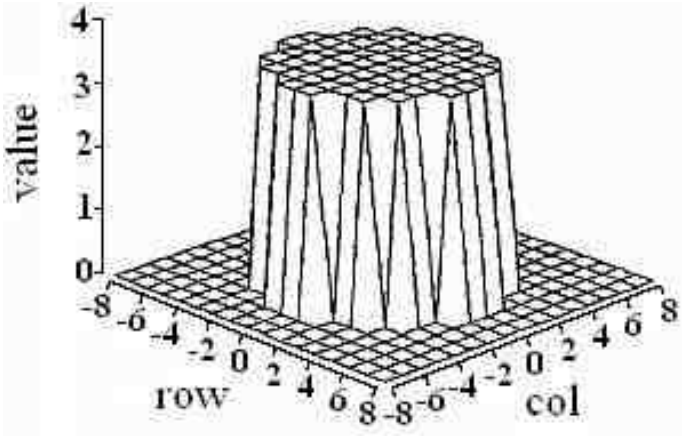

Figure 3.3: Pill-box point spread function

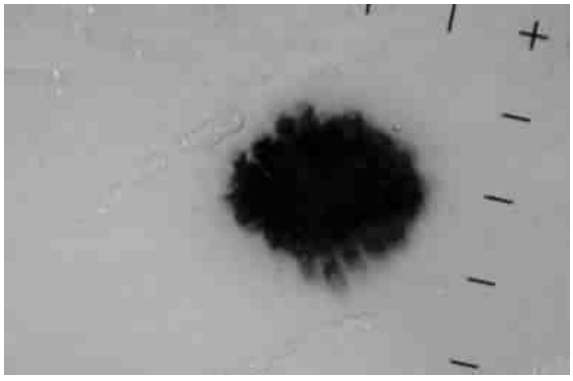

Figure 3.4: Result of applying noise filtering step on the dermoscopy image of Figure 3.2.

3.1.4 Intensity Adjustment

This step is essentially an enhancement process in which the dynamic range of pixel values in the image is mapped into a new range. The purpose is to smooth and stretch the image histogram and increase the contrast of the image in order to determine a more accurate threshold value in the thresholding step. Intensity adjustment works by rescaling the intensity values in the original image to cover the entire dynamic range. Figure 3.5

illustrates the result of applying this intensity adjustment step on the dermoscopy image of Figure 3.4.

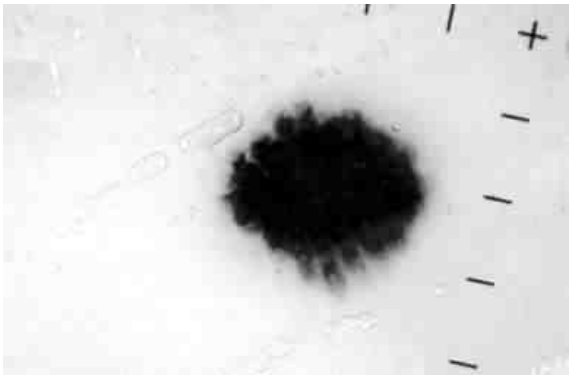

Figure 3.5: Result of applying intensity adjustment step on the dermoscopy image of Figure 3.4.

3.1.5 Clustering-based Histogram Thresholding

Thresholding is the process of classifying the pixels of a greyscale image into two classes, so that the image can be converted to binary by assigning each pixel either a 0 or 1, depending on its grey level. The thresholding procedure used in this paper is based on the well-known Otsu's thresholding method [109]. The basic premise of this method is the assumption that an image contains two clusters of pixels, *e.g.* foreground and background, which, in our case, correspond to the lesion and its surrounding skin, respectively. To identify these two clusters accurately, an algorithm is used to search for an optimal threshold level using discriminant analysis, where zero-th and first-order cumulative moments of the histogram are calculated and used to define a measure of separability between the two clusters. An optimal threshold level separating the two clusters is obtained by minimising the within-cluster variance (σ_ω^2), which is defined as a weighted sum of variances of the two clusters:

$$\sigma_\omega^2(t) = \omega_1(t)\sigma_1^2(t) + \omega_2(t)\sigma_2^2(t) \tag{3.7}$$

where weights ω_i are the probabilities of the two clusters separated by a threshold t and σ_i^2 denotes the variances of these clusters. It can be shown that minimising the within-cluster variance (σ_ω^2) is equivalent to maximising between-cluster variance. The between-cluster variance (σ_b^2) is recalculated in the Otsu method [109] as the following:

$$\sigma_b^2(t) = \sigma^2 - \sigma_\omega^2(t) = \omega_1(t)\omega_2(t)[\mu_1(t) - \mu_2(t)]^2 \tag{3.8}$$

where μ_i are the mean values of the two clusters. Starting from an initial threshold value of $t = 1$, ω_i and σ_i^2 are updated iteratively and in each iteration $\sigma_b^2(t)$ is calculated. The optimal threshold corresponds to the maximum value of $\sigma_b^2(t)$. The output binary image has values of 1 for all pixels in the input image with luminance greater than the threshold

level and 0 for the remaining pixels. The result of applying the thresholding step on the dermoscopy image of Figure 3.5 is given in Figure 3.6.

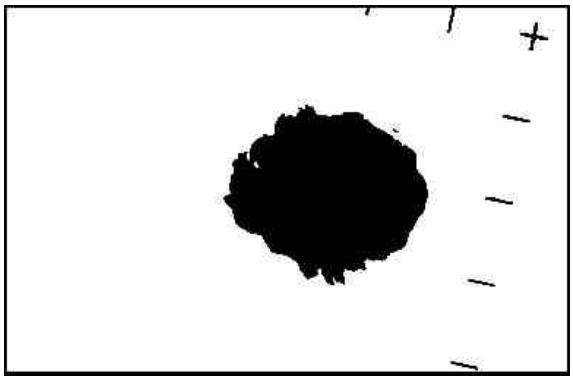

Figure 3.6: Result of applying thresholding step on the dermoscopy image of Figure 3.5.

3.1.6 Connected Component Analysis

In some of the dermoscopy images, extra objects appear in the surrounding skin area, such as blue marks made by dermatologists when examining the patient's skin. These objects, which have not been eliminated in the noise removal step, appear with intensity values similar to that of the lesion and may be misclassified as lesion. The purpose of this step is to exclude these objects from the segmentation output. To this end, the number of connected objects within the image is determined using the run-length encoding technique [194] and the connected objects are labeled. Finally, the two largest areas (*i.e.*, lesion and surrounding skin) are kept and all other components are discarded. Figure 3.7 shows the result of applying this step on the dermoscopy image of Figure 3.6.

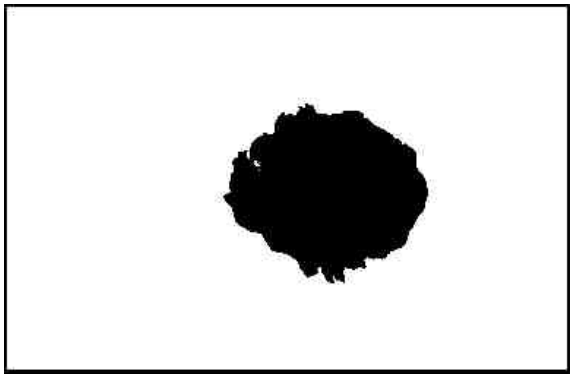

Figure 3.7: Result of connected component analysis on the dermoscopy image of Figure 3.6.

3.1.7 Lesion Formation

In order to obtain the final lesion object, the holes inside the boundary are filled by performing morphological filling on the binary image. Figure 3.8 shows the final segmentation result. Figure 3.10 shows the segmentation sequence on another dermoscopy image. Figure 3.9 shows more samples of the detected border by the proposed algorithm.

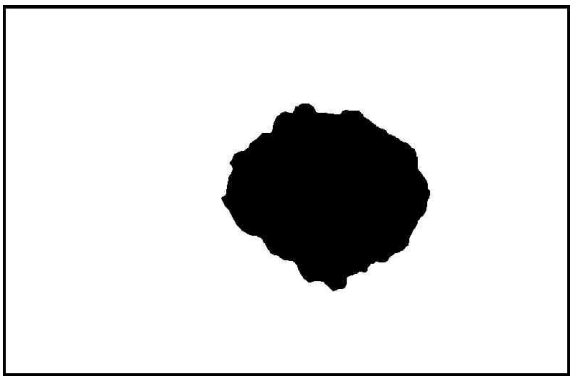

Figure 3.8: Applying morphological operator on the dermoscopy image of Figure 3.7 to form the final segmentation result.

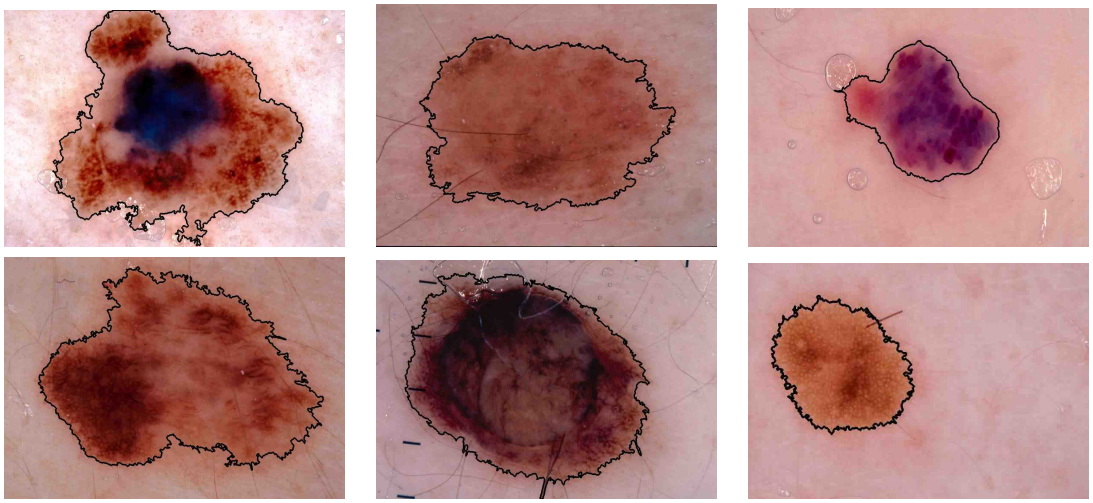

Figure 3.9: Sample border detection results using Global thresholding method

3.1.8 Experimental Results: Determining The Optimal Colour Channel

In order to determine the optimal colour channels for border detection, we conducted a pilot study on a set of 30 high-resolution dermoscopy images provided by the Royal

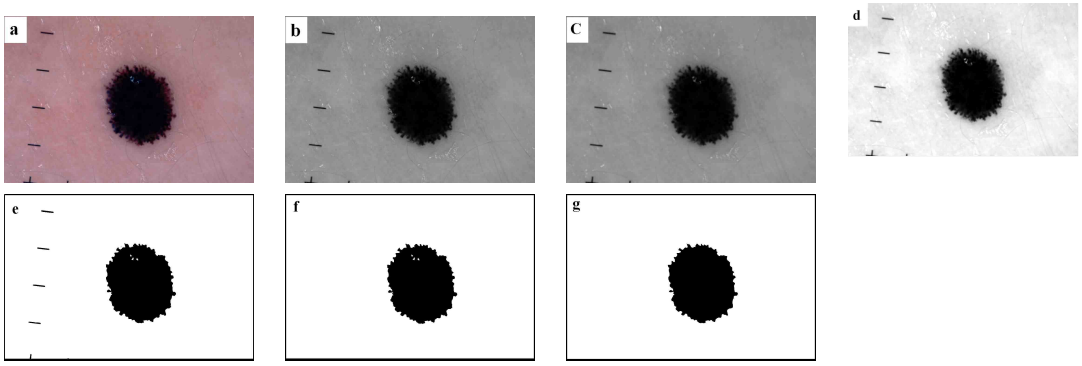

Figure 3.10: Global thresholding method: (a) Original image, (b) Colour space transformation, (c) Noise removal, (d) Intensity adjustment, (e) Thresholding, (f) Connected component analysis, (g) Morphological filling.

Melbourne Hospital, Australia. As shown in Figure 3.9, the image set contains a variety of dermoscopy images in terms of colour, texture and shape, which provides a strong basis for the experiment to make the algorithm generic and applicable to real-world applications.

As ground truth, for each dermoscopy image, four manual borders were independently drawn by four experts (two experienced dermatologists and two dermatology registrars). Figure 3.11 shows the borders drawn by dermatologists for a sample image. To perform a comparison between the automatic borders (generated by the proposed border detection method) and the manual borders (drawn by the dermatologists), in order to evaluate the goodness of the detected borders, various metrics have been used in the existing dermoscopy applications [78]. Here three of those metrics are used; namely, sensitivity (3.9), specificity (3.10), accuracy (3.11). Sensitivity shows the percentage of actual lesion that has been truly detected. Specificity shows the percentage of actual background skin that has been truly detected. Accuracy shows the degree of closeness of the detected border to the actual border, which takes both background skin and lesion pixels into account.

$$\text{Sensitivity} = \frac{TP}{TP + FN} \times 100\% \tag{3.9}$$

$$\text{Specificity} = \frac{TN}{TN + FP} \times 100\% \tag{3.10}$$

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN} \times 100\% \tag{3.11}$$

where TP, TN, FP, and FN refer to true positive, true negative, false positive and false negative, respectively. TP and TN represent the number of pixels which are classified correctly as part of the lesion and background skin in both the manual and automatic borders, respectively. FP represents the number of pixels which are classified as part of the lesion in the automatic border, but are labeled as part of the background skin in the manual border. Finally, FN represents the number of pixels which are classified as part of

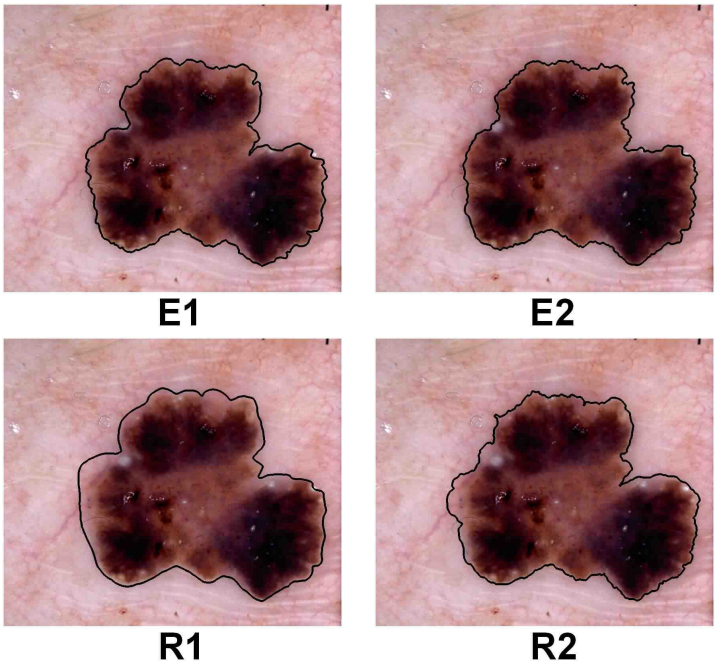

Figure 3.11: Manual borders of the same lesion drawn by four dermatologists; E1 and E2: experienced dermatologists, R1 and R2: dermatology registrars.

the background skin in the automatic border, but are labeled as part of the lesion in the manual border.

In the above three metrics, the manual borders are taken as the ground truth and the calculated measures are used to quantify the discrepancy between the automatic borders and the manual ones. However, another perspective in the analysis is to evaluate the inter-observer variabilities among the four dermatologists, as demonstrated in Figure 3.11. To achieve this, a fourth metric, *i.e.*, (Sorensen) similarity index [195], is used to quantify the degree of similarity between any two borders, without taking either of them as the ground truth. This metric (Equations 3.12 or 3.13) has been used in a variety of domains, yet it has not been used in dermoscopy.

$$\text{Similarity} = \frac{2A_{1 \cap 2}}{A_1 + A_2} \tag{3.12}$$

where A_i is the areas of any two segmented images and $A_{1 \cap 2}$ is the number of pixels the two images share (intersection of the two images). The similarity metric can also be expressed in terms of TP, TN, FP, and FN, given by:

$$\text{Similarity} = \frac{2TP}{2TP + FN + FP} \times 100\% \tag{3.13}$$

Table 3.2 shows the mean and standard deviation values of similarity among the borders drawn by the four dermatologists. The similarity values indicate that there are

higher similarities between the two experienced dermatologists and two dermatology registrars, although differences are all less than the standard deviations. The overall similarity index is high which indicates the reliability of the ground truth. We have also used the similarity metric to investigate which automatic border is closest to the gold standard (manual border).

Table 3.2: Mean ± standard deviation for similarity (%) between dermatologists; E1 and E2: experienced dermatologists, R1 and R2: dermatology registrars.

	E1	**E2**	**R1**	**R2**
E1	-	95.92 ±1.76	95.09 ±1.75	94.78 ±2.25
E2	-	-	94.68 ±1.73	94.78 ±2.22
R1	-	-	-	95.86 ±1.25

As mentioned earlier, the main purpose of this analysis is to determine the colour channels that lead to the most accurate borders. To achieve this aim, 24 different colour channels listed in Table 3.1 are extracted from the 30 images, resulting in 24 different borders per image. Each border (out of 24×30), is then separately compared to each of the four manual borders using the above-mentioned four metrics. The 30 values of each metric are then averaged, resulting in 24 average metrics of sensitivity, specificity, accuracy and similarity. The maximum values of each of these metrics and their corresponding colour channels are then identified, resulting in four colour channels for each metric, as shown in Table 3.3. As shown in Table 3.3, using the average values, the initial 24 colour channels are narrowed down to the four colour channels of X, XoYoR, XoYoZoR and R. The R colour channel produces short borders (small segmented lesions), thus small FP and large TN values, and accordingly results in the highest specificity values. Borders obtained by using XoYoRoZ colour channel, which is the result of integrating the X, Y, Z, and R colour channels via OR operation, produce a larger border, thus small FN value. Consequently, they result in high sensitivity.

Table 3.3: Colour channels with the largest mean for four ground truths; E1 and E2: experienced dermatologists, R1 and R2: dermatology registrars.

Expertise	**Accuracy**	**Similarity**	**Sensitivity**	**Specificity**
E1	XoYoR (96.03)	X (91.45)	XoYoZoR (93.52)	R (99.96)
E2	XoYoR (96.01)	X (91.55)	XoYoZoR (94.33)	R (99.92)
R1	XoYoR (95.14)	XoYoZoR (90.61)	XoYoZoR (90.60)	R (99.99)
R2	XoYoR (94.89)	XoYoZoR (90.56)	XoYoZoR (90.37)	R (99.99)

The results presented in Table 3.3 are calculated using each manual border as the ground truth, hence there are four sets of results. Another useful analysis is obtained when a single ground truth for each image is used. This ground truth is obtained by finding the intersection of the areas within the four manual borders (in the ensuing, we will refer to these as *common manual borders*). Then the common manual border is used

as the ground truth and compared to the four colour channels shown in Table 3.3 for the image set. The result of the four metrics are shown in Table 3.4. Here, we report means and 95% confidence intervals. An observed sample mean provides the best estimate of the true population value; the confidence interval describes a range of values for the true population mean that are consistent with the data that was observed. We report these results as $\bar{x} \pm me$, where $\bar{x}$ is the sample mean and me is the margin of error; a 95% confidence interval ranges from $\bar{x} - me$ to $\bar{x} + me$ [196].

To precisely analyse the sensitivity and specificity metrics, the AUC (Area Under ROC Curve[1]) is obtained by drawing the ROC graph and calculating the corresponding AUC value for R, X, XoYoZ and XoYoRoZ colour channels. Table 3.4 suggests that in terms of the accuracy metric and AUC values, colour channels X and XoYoR provide the highest scores. With respect to similarity, X and XoYoR colour channels give the highest average. Figure 3.12 shows the ROC curves for the four colour channel X, R, XoYoY and XoYoZoR. Table 3.5 shows the performance ranking of the colour channels according to different evaluation metrics, which shows that colour channels X and XoYoR provide the best overall results. Figures 3.13 and 3.14 show the accuracy and similarity percentage of these two colour channels over the image set of 30 dermoscopy images, taking the common manual border as the ground truth.

It is well known that XYZ is a standard colour space defined by CIE. However, in the domain of skin lesion segmentation, we have shown the appropriateness of XYZ colour space and more specifically the X and Y colour channels when combined with the R channel.

Table 3.4: Segmentation results (mean ±margin of error) for optimal colour channels, the common manual borders as ground truth.

Colour Channel	**Accuracy**	**Sensitivity**	**Specificity**	**Similarity**
X	96.80 ±0.01	90.49 ±0.03	99.02 ±0.00	93.18 ±1.82
XoYoR	96.94 ±0.01	92.86 ±0.02	98.51 ±0.00	92.89 ±3.15
XoYoZoR	95.15 ±0.03	96.80 ±0.01	94.47 ±0.05	91.63 ±5.15
R	92.42 ±0.03	73.40 ±0.07	99.90 ±0.00	82.91 ±5.45

Table 3.5: Performance ranking of the colour channels.

Rank	**Accuracy**	**AUC**	**Similarity**
1	XoYoR	X, XoYoR	X
2	X	XoYoZoR	XoYoR
3	XoYoZoR	R	XoYoZoR
4	R		R

[1]Receiver operating characteristic (or ROC) curve is a graphical plot of the sensitivity versus (1–specificity).

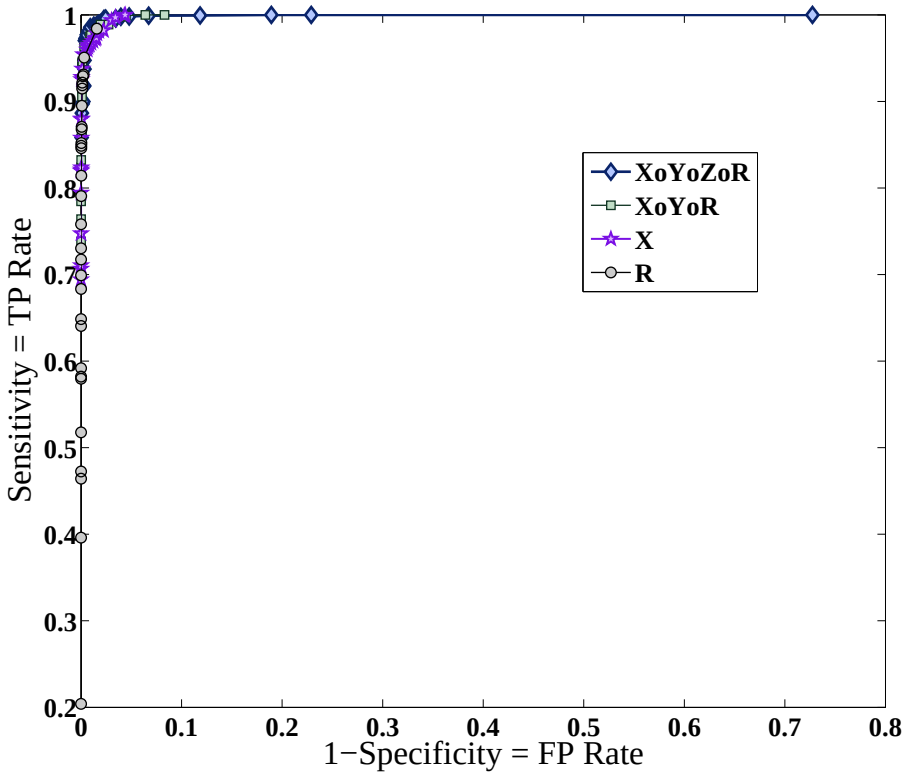

Figure 3.12: ROC curves for four colour channels X, R, XoYoR and XoYoZoR.

3.1.9 Comparison with Other Automated Methods

The proposed method is compared with two state-of-the-art skin lesion segmentation methods, namely; statistical region merging [96] and the JSEG method [95]. Table 3.6 shows the segmentation evaluation result obtained by SRM (statistical region merging) and JSEG, and our proposed Global thresholding method when applied to the X and XoYoR colour channels. The results demonstrate that the proposed thresholding-based method, in spite of its simplicity, with a proper choice of suggested colour channels is highly competitive with other well-known skin lesion segmentation methods. X channel outperforms other automated methods with respect to specificity and along with the JSEG method, it obtains the best performance in terms of similarity, while XoYoR follows them closely. The proposed method in both colour channels gains the highest AUC value, while JSEG follows them closely, and SRM comes last. XoYoR colour channel gain the highest sensitivity and accuracy on average, followed closely by others methods. In addition, our method is potentially faster since it mainly involves scalar processing as opposed to the SRM and JSEG methods, which are based on vector processing.

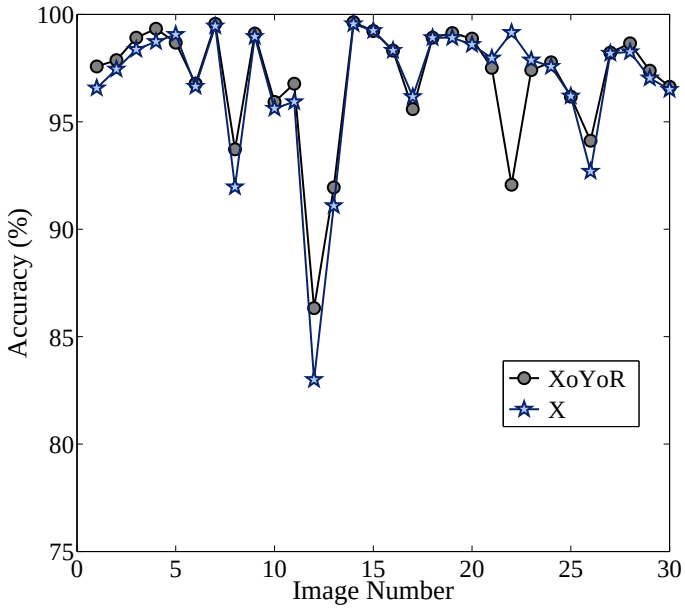

Figure 3.13: Accuracy of the optimal colour channels, X and XoYoR, over the image set.

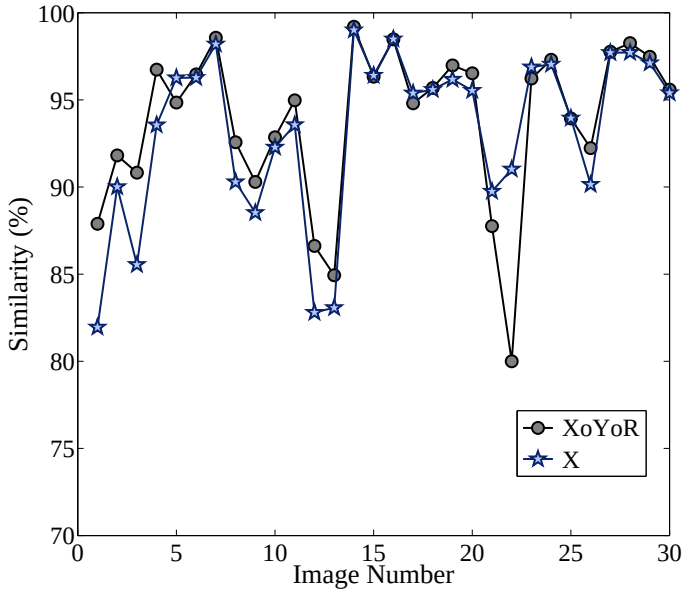

Figure 3.14: Similarity of the optimal colour channels, X and XoYoR, over the image set.

3.1.10 Conclusion

This section presented an accurate automatic segmentation method based on colour space analysis and clustering-based histogram thresholding which determines the most effective

Table 3.6: Segmentation evaluation results (mean ±margin of error): comparative study.

Rank	**Accuracy**	**Similarity**	**Sensitivity**	**Specificity**	**AUC**
SRM	96.27 ±0.01	91.55 ±2.32	92.45 ±0.03	97.43 ±0.01	0.990
JSEG	96.70 ±0.01	93.23 ±3.14	92.05 ±0.04	98.35 ±0.01	0.997
X	96.80 ±0.01	93.18 ±1.82	90.49 ±0.03	99.02 ±0.00	0.998
XoYoR	96.94 ±0.01	92.89 ±3.15	92.86 ±0.02	98.51 ±0.00	0.998

and discriminative colour channel for detecting the borders in dermoscopy images. Various colour channels from different colour spaces were incorporated to maximise the discrimination between two clusters of pixels within the image, lesion and surrounding skin pixels, to obtain more accurate histogram analysis, and consequently to detect more accurate borders. Each colour channel underwent preprocessing steps, clustering-based histogram thresholding and a set of pixel-based computations and morphological operations to eventually identify the border of the lesion.

Segmentation results were quantitatively evaluated by comparing automated results to manual borders independently drawn by four dermatologists. The comparison was done with respect to four different metrics of accuracy, sensitivity, specificity and similarity, incorporating ROC analysis. Experimental results indicated that X and XoYoR colour channels obtain the highest overall performance with an accuracy of approximately 97%. The results were also compared with two state-of-the-art automated methods, which demonstrate that the proposed thresholding-based method, in spite of its simplicity, with a proper choice of suggested colour channels is highly competitive with the well-known skin lesion segmentation methods, and outperforms them with respect to accuracy, specificity, and AUC metrics. Furthermore, the proposed method is potentially faster since it mainly involves scalar processing as opposed to vector processing performed in those methods.

3.2 Hybrid Thresholding on Optimised Colour Channels

Our experience with dermatologists has consistently shown that manual borders tend to be larger than the corresponding automatic borders. This fact, which has also been reported in several studies [78], is illustrated in the dermoscopy images shown in Figure 3.15, where three areas are identified; namely, core-lesion, edge-lesion and background skin. The width of the edge-lesion area can be quite variable depending on the skin colour, lesion colour, sharpness or fuzziness of the border, and the lesion pattern. Automated border detection methods can easily identify the core-lesion area by finding the sharpest pigment change. However, they often fail to precisely detect the edge-lesion area and thus exclude it from the segmentation result. In contrast, experienced dermatologists prefer to choose the outmost detectable pigment to minimise the risk of incorrect diagnosis. In several studies the discrepancy between the manual and automatic borders has been reduced by expanding the automatic borders using various methods including morphological filtering [96], Euclidean distance transform [96], iterative region growing [26], and gradient information [197]. In

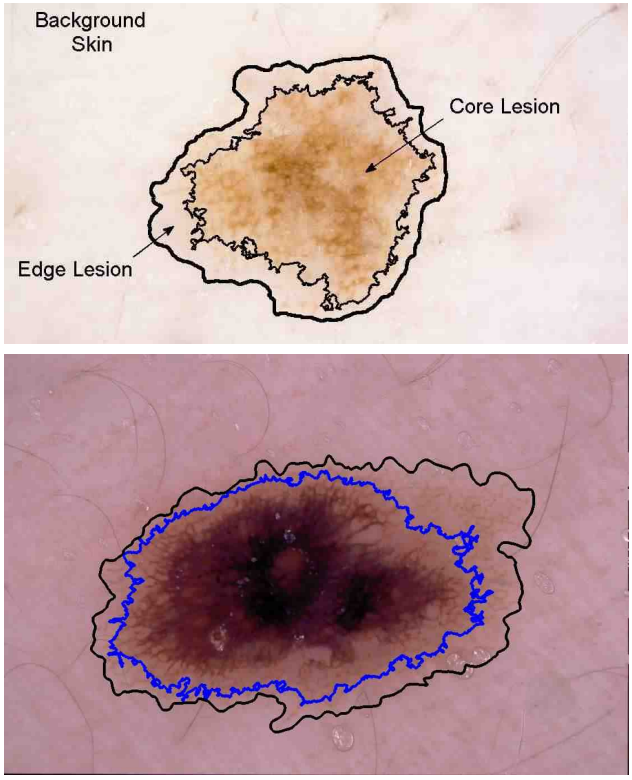

Figure 3.15: Three areas appears in dermoscopy images: core-lesion, edge-lesion and background skin.

this study, we introduce a novel border detection method to decrease such discrepancy and improve the results obtained from our first method, reported in the previous section.

The newly proposed method, called Hybrid thresholding, is comprised of two stages; in the first stage, we identify the core-lesion using our Global method (explained in Section 3.1). In the second stage, the edge-lesion area is identified by integrating local colour information around the edge-lesion's boundary, and applying an adaptive histogram thresholding method on the optimal colour channels determined in the previous section. These two stages are detailed next.

3.2.1 Global Thresholding: Forming the Core-Lesion

The core-lesion area is detected by applying the Global method proposed in Section 3.1; applying the global clustering-based histogram thresholding method on the optimal colour channels of XoYoR obtained in the colour optimisation procedure (Figure 3.1). As described previously, this step includes the pre-processing operations of hair removal, noise filtering and intensity adjustment. This is followed by application of Otsu's histogram thresholding algorithm to the XoYoR colour channel, and performing connected com-

ponent analysis and morphological operations to obtain the initial border and form the so-called core-lesion area. The histogram analysis procedure is quite fast, since it performs one-dimensional clustering and the results of each X and Y and R colour channel are integrated using the logical OR operation.

In this process we have made a new choice for the noise removal filter. In the proposed Global method, the averaging pillbox filter was used, whereas we have observed that using a Gaussian low-pass filter [198] (Equation 3.14) results in less blurred images and retains more of the important details that can improve the border detection accuracy. The Gaussian filter utilised a 11×11 kernel with a standard deviation (σ) value of 0.5.

$$G(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{3.14}$$

3.2.2 Local Thresholding: Forming the Edge-Lesion

To expand or shrink the core-lesion boundary to the edge-lesion boundary, we have applied a local clustering-based thresholding technique based on Otsu's method explained in Section 3.1.5. After performing noise removal (using a Gaussian filter with similar parameter settings as above) and intensity adjustment on the image, the histogram thresholding is applied to the X colour channel which was determined as optimal in the colour optimisation procedure, as part of the Global method described in the previous section , Figure 3.1.

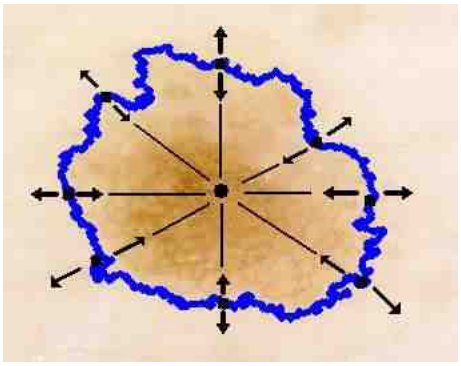

Figure 3.16: Shrinkage and expansion is done radially with respect to the centroid of the lesion.

Starting from an arbitrary point on the initial core-lesion boundary, the local threshold is calculated over a window (using the Otsu method on the X colour channel). Let us define Ws to be a window of size $s \times s$, which is a square window centered on a core-lesion border pixel. If the local threshold value is less than a predefined threshold called T_{expand} the boundary is expanded by one pixel. If it is greater than a predefined threshold called T_{shrink} the boundary is shrunk. Otherwise, we are in the state of *No Change*. This implies that based on the previous moves, we should make a decision to either laterally move to

the adjacent pixel, or to expand or shrink the boundary. As shown in Figure 3.16 the inward and outward moves (shrinkage and expansion) are done radially (along the line connecting the centroid of the lesion to the pixel on the core-lesion boundary). The lesion centroid is calculated by using Equation (3.15).

$$(x_c, y_c) = \left(\frac{\sum_{i=1}^{n} x_i}{n}, \frac{\sum_{i=1}^{n} y_i}{n} \right) \tag{3.15}$$

where n is the number of pixels inside the lesion, and (x_i, y_i) are the coordinates of the i-th lesion pixel.

To define the threshold values for shrinkage and expansion, a bandwidth is calculated based on background skin value and core-lesion value. We have applied the same thresholding method to the background skin area and core-lesion area to obtain estimates of these values. The bandwidth is calculated by Equation (3.16). Both core-lesion and background skin are taken into account when calculating the bandwidth. In this way the Bandwidth is normalised, so that variations in the colour of different lesions on the same person can be accounted for. Without this normalisation process the lesion border will only expand regardless of the difference between the background skin colour and the core-lesion colour. The threshold values for expansion and shrinkage are given by Equations (3.17 and 3.18), respectively.

$$Bandwidth = \%B \times (T_{skin} - T_{core-lesion}) \tag{3.16}$$

where B, the bandwidth factor, is the percentage we wish to expand from the core-lesion towards the background skin. T_{skin} and $T_{core-lesion}$ refer to estimates of the background skin and core-lesion area, respectively.

$$T_{expand} = T_{skin} - Bandwidth \tag{3.17}$$

$$T_{shrink} = T_{skin} + Bandwidth \tag{3.18}$$

$$\begin{cases} Expand\ (move\ outward) & if\ T_{local} \leq T_{expand} \\ Shrink\ (move\ inward) & if\ T_{local} \geq T_{shrink} \\ No\ Change & if\ T_{expand} < T_{local} < T_{shrink} \end{cases} \tag{3.19}$$

The local threshold is calculated for each and every window centered on successive pixels along the core-lesion boundary and the process is stopped when a pixel is revisited. Figure 3.17 shows samples of the detected borders by the proposed method.

3.2.3 Experimental Results: Optimising B and W Parameters

The proposed border detection method outlined above has been tested on a set of 85 high-resolution dermoscopy images obtained from Melbourne Royal Hospital. The images are taken by professional photographers using a Canon EOS 450D camera under unified zooming and lighting conditions. They are 24-bit RGB colour images with dimensions of

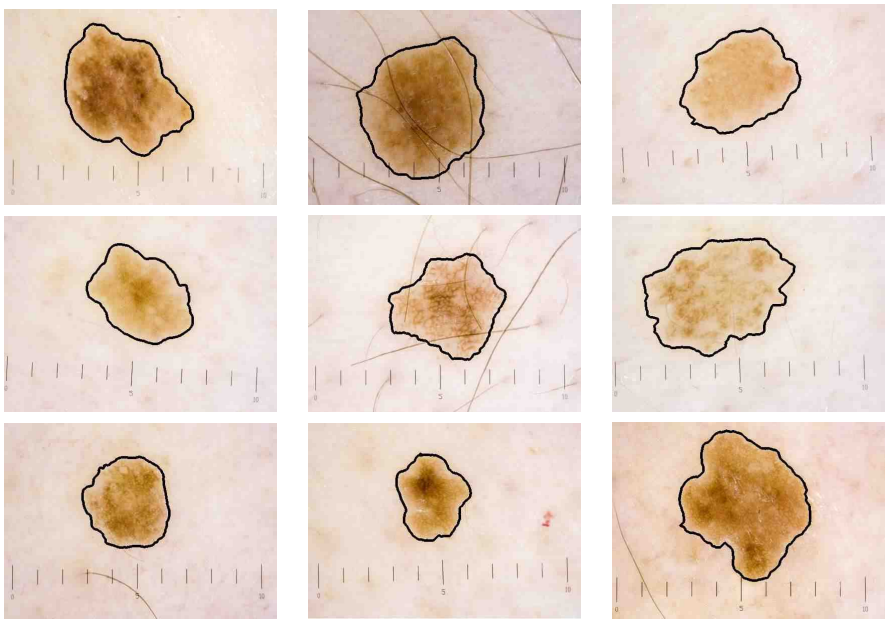

Figure 3.17: Sample border detection results using hybrid thresholding on optimised colour channels.

2000×1334 pixels. Manual borders are drawn by an experienced dermatologist using a Wacom Intuos A4 size Tablet PC. These borders are then used as ground truth for the evaluation of the automatic borders. We have also compared the automated borders with a second ground truth, manually drawn by a less-experienced dermatologist (a first-year dermatology registrar).

The purpose is to (i) determine the optimal values for W (window size) and B (bandwidth factor) parameters[2], discussed in Section 3.2.2, via two different approaches; automated borders versus experienced dermatologist, and automated borders versus less-experienced dermatologist, and investigate their consistency, (ii) determine which of the two borders; the automated borders drawn by our method, and the less experienced dermatologist' borders, is closer to the borders drawn by the experienced dermatologist.

A comprehensive experiment is performed on the set of 85 dermoscopy images, with W ranging from 30 to 70 and B ranging from 10% to 90% (steps of 10). Initially, windows of sizes from 10×10 to 90×90 are investigated and the result showed that those below 30×30 and above 70×70 are insignificant, therefore not included in the analysis. As shown in Figure 3.18 different borders can be obtained by using different values of W and B. Consequently, 45 borders are obtained for each dermoscopy image.

In order to quantitatively compare the automatic borders to the manual borders, the three evaluation metrics of sensitivity, specificity, and accuracy used in the previous

[2]W is the window size for calculating the local threshold, and B is the extent to which the core-lesion area is expanded towards the background skin depends.

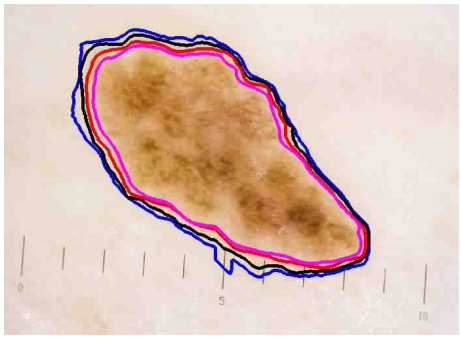

Figure 3.18: Different borders produced by different values of W and B parameters.

section (Equations 3.9, 3.10 and 3.11), are applied, as well as the two metrics of precision and border error. Precision or positive predictive value (Equation 3.20) is basically an index of reproducibility which shows the degree to which further experiments will show the same results. In this application, the precision metric shows what percentage of the detected border is the true lesion. Border error or XOR metric (Equation 3.21 or 3.22) measures the discrepancy between automatic and manual borders.

$$\text{Precision} = \frac{TP}{TP + FP} \times 100\% \tag{3.20}$$

$$\text{Border Error} = \frac{Area(A \otimes M)}{Area(M)} \times 100\% \tag{3.21}$$

where M is the manually segmented image, $A \otimes M$ represents the differential segmentation obtained by the manual and automatic borders using XOR operation, and $Area(.)$ denotes the number of pixels in the binary images of M and $A \otimes M$. The border error metric can also be expressed in terms of TP, TN, FP, and FN, given by:

$$\text{Border Error} = \frac{FP + FN}{TP + FN} \times 100\% \tag{3.22}$$

3.2.4 Comparison with Experienced Dermatologist (First Ground Truth)

Figure 3.19 shows the mean and standard deviation values for accuracy and precision metrics for different B and W parameters over the image set, using the experienced dermatologist's borders as the ground truth. As illustrated in the graphs, mean accuracy and precision increase with decreasing window size. The variability around mean (measured by standard deviation) is smaller for smaller W (W=30) in most of the cases. With respect to the accuracy metric, both mean and std graphs suggest that the Bandwidth factor (B) is optimal in the medium range of [40..60]%. Precision shows an incremental trend in mean value, along with a decreasing trend of std as B grows, which suggests a value of larger than 50% for B may be optimal.

Figure 3.20 shows the mean value for sensitivity and specificity metrics, along with the corresponding AUC value, and also the proportion of images with border error smaller than 20%, for different B and W parameters over the image set, taking the experienced dermatologist's borders as the ground truth. The graphs indicate that the highest AUC value is obtained by a smaller W and larger B which is in agreement with the results obtained from the accuracy and precision metrics. Overall, the value of 30 for W and 60% for B seems to be a reasonable trade-off considering all the metrics. Table 3.6 gives the mean and margin of error for each metric for the proposed method when using the optimised parameters (W30B60). High mean values (over 93%) are observed for all metrics, except for sensitivity which is close to 90%.

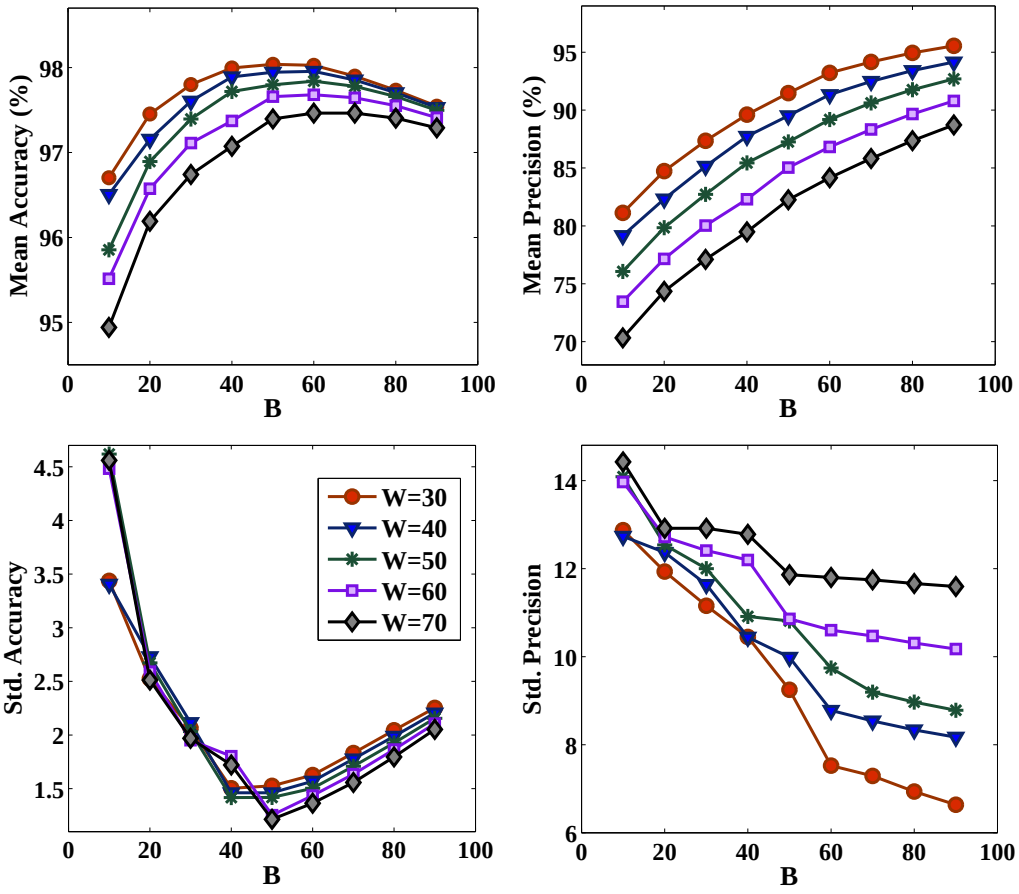

Figure 3.19: Mean and standard deviation values for accuracy and precision metrics for different B and W over the image set compared to experienced dermatologist.

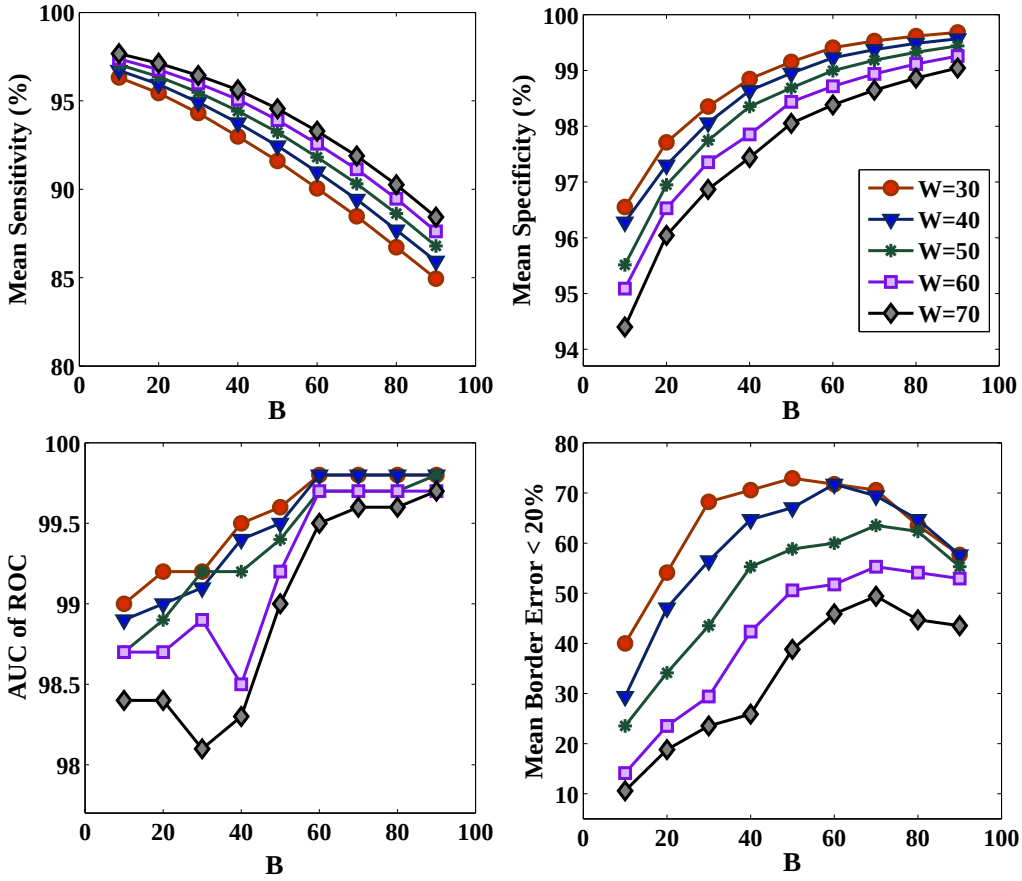

Figure 3.20: Mean sensitivity and specificity, the corresponding AUC, the proportion of images with border error $\leq 20\%$, for different B and W over the image set, compared to experienced dermatologist.

3.2.5 Comparison with Dermatology Registrar (Second Ground Truth)

Figure 3.21 shows the mean value for accuracy and precision metrics, the AUC calculated from sensitivity and specificity metrics and also the proportion of images with border error smaller than 20%, for different B and W parameters over the image set, using the dermatology registrar's borders as the ground truth. The trend is similar to that observed in Figures 3.19 and 3.20 with the experienced dermatologist as the ground truth. Large W values are ruled out, and similarly, W=30 appears to be optimal across the range examined. However, for accuracy and border error smaller B values tend to obtain better results than large B values. Considering precision and AUC it can be seen that an intermediate value for B in the range of [30..50]% is more likely to produce results that are consistent with the dermatology registrar.

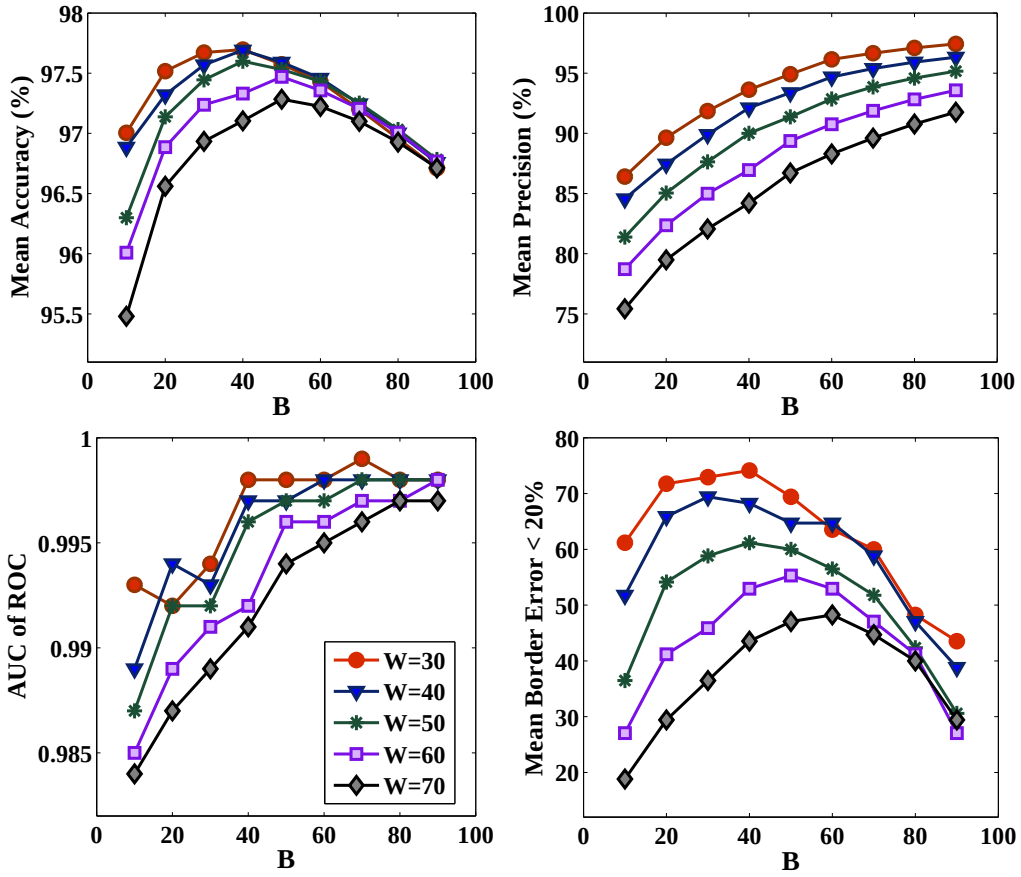

Figure 3.21: Mean accuracy and precision, AUC, proportion of images with border error $\leq 20\%$, for different B and W over the image set, compared to dermatology registrar.

3.2.6 Comparison between Automated Borders and Less-experienced Dermatologist's Borders

As mentioned previously, we have taken the borders drawn by the experienced dermatologist as the ground truth and compared the results of our method (W30B60) to those of the dermatology registrar's. Figure 3.22 shows a comparison between our results and dermatology registrar's with respect to different evaluation metrics. From the figure it can be seen that our method yields either equivalent or closer results to those of the experienced dermatologist's than the dermatology registrar's to the experienced dermatologist's. For example, taking the experienced dermatologist as ground truth, the mean precision for our algorithm is 5.3% higher than for the dermatology registrar. The margin of error was 1.9, giving a 95% confidence interval of 3.4% to 7.2%. We can be reasonably confident that our method is, on average, between 3.4% and 7.2% better than the dermatology registrar, which implies that our method is more effective than the dermatology registrar.

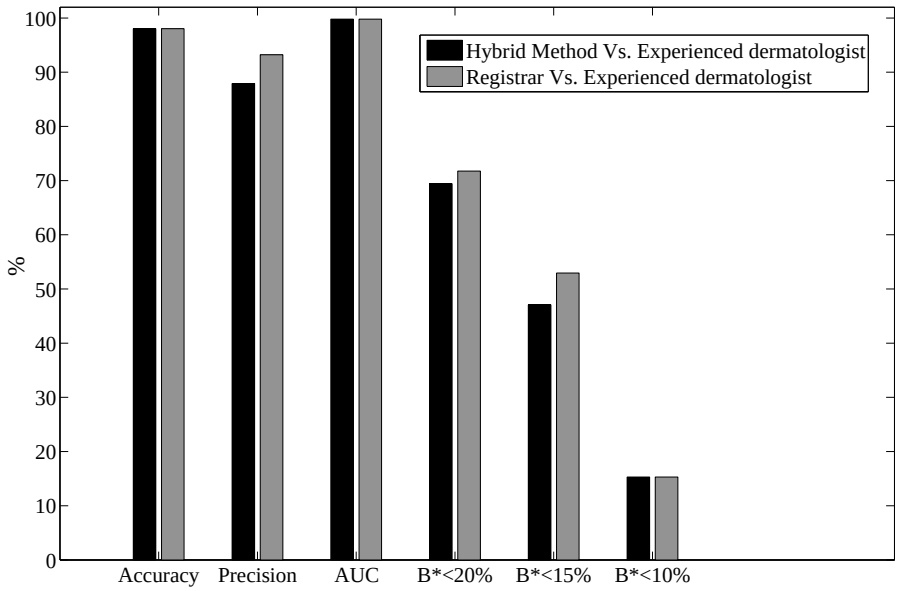

Figure 3.22: Comparison between the border detected by our method to manual border drawn by dermatology registrar, using the experienced dermatologist's manual borders as the ground truth (B^* = Border Error).

3.2.7 Comparison with Other Automated Methods

The proposed method with optimised parameters (W30B60) is compared with three state-of-the-art lesion border detection methods; namely, JSEG [95], DTEA [26] and KPP [89]. Tables 3.7 and 3.8 show the segmentation evaluation results obtained by JSEG, DTEA, KPP and the proposed thresholding method using XoYoR colour channel, in forming core-lesion boundary, and X colour channel, in forming edge-lesion boundary, with window size of 30 and bandwidth factor of 60, considering the borders drawn by the experienced dermatologist as the ground truth.

Table 3.7 shows that the proposed method performs better on average in terms of sensitivity and AUC metrics, whereas it is highly competitive with the JSEG method with respect to the accuracy and specificity. With respect to the accuracy and sensitivity, DTEA comes third, followed by KPP and these two also yield fairly competitive results in terms of specificity and AUC. However, KPP exhibits the highest precision, followed by DTEA, JSEG and W30B60, respectively. Table 3.8 shows the number of images with percentage border error less than 10%, 15%, 20%, 30% and 40% for different automated methods, which reflects the same trend as the similarity metric. Table 3.8 indicates that the proposed method achieves the best scores among the automated methods, while JSEG follows it closely, and DTEA and KPP come next.

Table 3.9 provides estimates and margin of error for the mean differences between W30B60 and the alternative methods. The mean differences between W30B60 and JSEG

Table 3.7: Comparing automated methods in terms of accuracy, precision, sensitivity, specificity (mean ±margin of error) and AUC, using the experienced dermatologist's manual borders as the ground truth.

	Accuracy	**Precision**	**Sensitivity**	**Specificity**	**AUC**
W30B60	98.01 ±0.35	93.91 ±1.43	89.64 ±1.90	99.43 ±0.13	99.86
JSEG	98.11 ±0.35	95.39 ±1.36	88.57 ±2.46	99.40 ±0.20	99.78
DTEA	97.63 ±0.49	96.92 ±0.77	83.84 ±2.68	99.70 ±0.07	99.80
KPP	97.52 ±0.38	97.10 ±1.32	80.95 ±2.97	99.63 ±0.16	99.68

Table 3.8: Comparing different border detection methods: Distribution of border error percentage, using the experienced dermatologist's manual borders as the ground truth.

Border Error	≤ 10	≤ 15	≤ 20	≤ 30	≤ 40
W30B60	16.25	55.00	72.50	92.50	98.75
JSEG	27.50	51.25	73.75	90.00	95.00
DTEA	7.50	47.50	67.50	88.75	93.75
KPP	13.75	28.75	50.00	78.75	88.75

are generally small and are unlikely to be important. The mean differences between (W30B60, DTEA) and (W30B60, KPP) for sensitivity and border error are relatively large, supporting our contention that W30B60 performs better. Moreover, the proposed method is potentially faster since it mainly involves scalar processing as opposed to vector processing.

3.2.8 Conclusion

This section presented a novel automatic border detection method based on colour-space analysis and clustering-based histogram thresholding. The method first applied the Global method on the optimal colour channel of XoYoR to detect an initial boundary, then applied the Hybrid method to expand the boundary and form the final segmentation result. The hybrid border detection method is tested on a set of 85 high-resolution dermoscopy images and the automatic borders are compared to borders drawn by two dermatologists using various evaluation metrics including accuracy, precision, sensitivity, specificity, and border error.

The proposed method, which is comprised of two stages, is designed to increase specificity in the first stage and sensitivity in the second stage. The proposed hybrid method is also compared to three state-of-the-art border detection methods and found to perform as well or better in terms of accuracy, sensitivity, specificity, AUC, and border error. In addition, our method is potentially faster since it mainly involves scalar processing as opposed to vector processing performed in the other methods. Taking the borders drawn by the experienced dermatologist as ground truth, and comparing the automatic borders with those drawn by the dermatology registrar, we have also shown that our method functions

Table 3.9: Mean ±margin of error for comparisons of automated methods, using the experienced dermatologist's manual borders as the ground truth.

	W30B60–JSEG	**W30B60–DTEA**	**W30B60–KPP**
Accuracy	-0.09 ±0.23	0.37 ±0.18	0.49 ±0.24
Precision	-1.47 ±1.37	-3.00 ±1.17	-3.18 ±1.63
Sensitivity	1.07 ±1.95	5.80 ±1.36	8.69 ±2.78
Specificity	0.03 ±0.17	-0.26 ±0.09	-0.19 ±0.16
Border Error	-0.50 ±1.90	2.10 ±1.70	5.30 ±2.30

closer to the experienced dermatologist than the dermatology registrar. This implies that the proposed Hybrid method is more effective than the dermatology registrar.

3.3 Global versus Hybrid Thresholding

In this section we demonstrate the superiority of the automated Hybrid thresholding approach to border detection in dermoscopy images, discussed in Section 3.1, over the Global thresholding method, explained in Section 3.2, through a newly introduced evaluation metric: Performance Index.

In the previous section, we proposed the Hybrid thresholding method, which combines global and adaptive local thresholding and uses a colour optimisation process to detect the lesion borders, and demonstrated its superiority over other state-of-the-art skin lesion segmentation algorithms. Here, we provide a new analytical and experimental framework to justify the necessity of integrating the local pixel information into the segmentation process in the border detection process. Moreover, the statistical analysis and optimisation procedure are shown to be convergent in determining the optimal parameters for the local thresholding procedure in order to obtain the most accurate borders.

3.3.1 Image Set and Gold Standard

The proposed method is tested on a set of 55 high-resolution dermoscopy images taken by a Canon EOS 450D camera in Royal Melbourne Hospital (24-bit RGB colour images, 2000×1334, TIFF).

To validate the borders produced by our methods, manual borders were independently drawn by three expert dermatologists using a Wacom Intuos A4 size Tablet PC. We are aware that different approaches can be used to determine a single ground truth, such as voting, averaging, etc. However, after discussion with dermatologists and considering the practical nature of the melanoma diagnosis, which calls for extreme caution when excluding portions of the image, and taking into account the inter-observer and intra-observer variability in borders drawn by dermatologists, we calculated the union of the four manually drawn borders for each image and considered that as the final ground

truth.

Figure 3.23 shows the various dermatologist-drawn borders and the corresponding obtained union border for two sample dermoscopy images. As shown in Figure 3.23 borders are close enough to each other, therefore there would be no adverse effect for the way the ground truth is calculated.

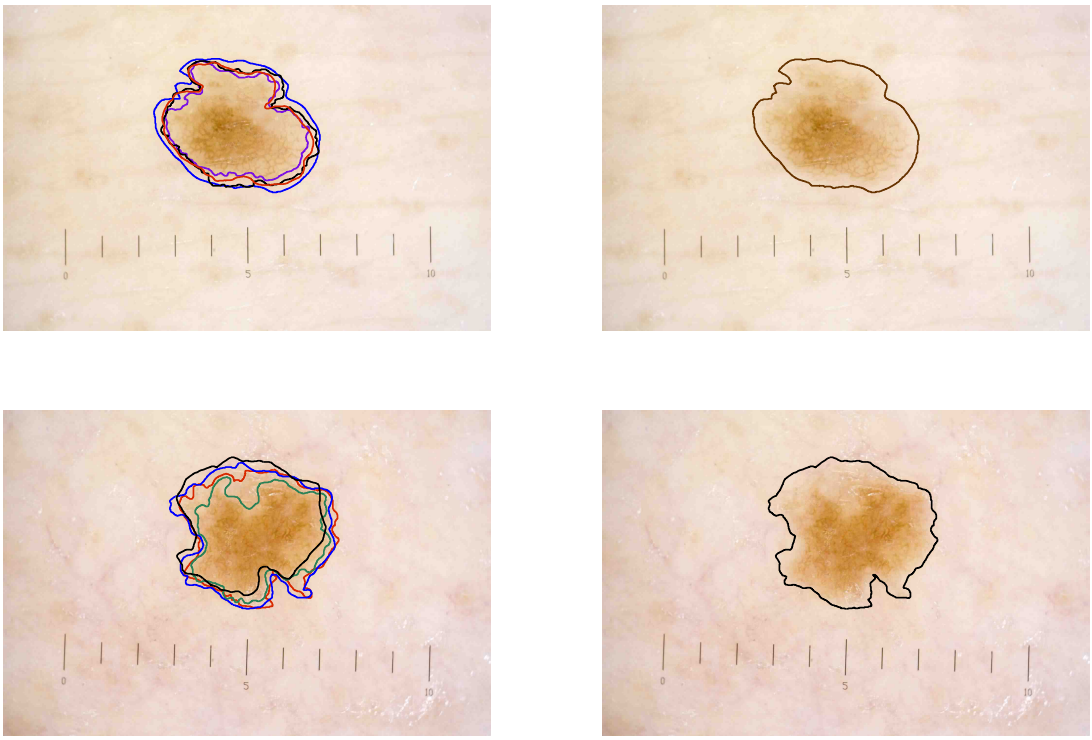

Figure 3.23: Gold standard for two sample dermoscopy images; different dermatologists-drawn borders and the corresponding union borders.

3.3.2 Standardisation of the Images

The lesion inside a dermoscopy image generally appears in different sizes and locations. Two metrics of accuracy and specificity include the TN factor which refers to the number of pixels of the surrounding background skin which are truly detected by the automated method. However, the TN factor strongly depends on the size of the lesion and its ratio to the whole image, thus the value of accuracy and specificity metrics can be biased in images with small lesions. To the best of our knowledge, this issue has not been addressed in previously published studies. To balance the effect of large TN and normalise the accuracy and specificity metrics, we set a frame of background skin around the lesion,

such that the area of the rectangular image frame is twice as big in area as that of an imaginary rectangular encompassing the lesion. This has the effect that the number of background pixel and lesion pixels is roughly the same. Figure 3.24 shows the segmentation result of a dermoscopy image before and after standardisation.

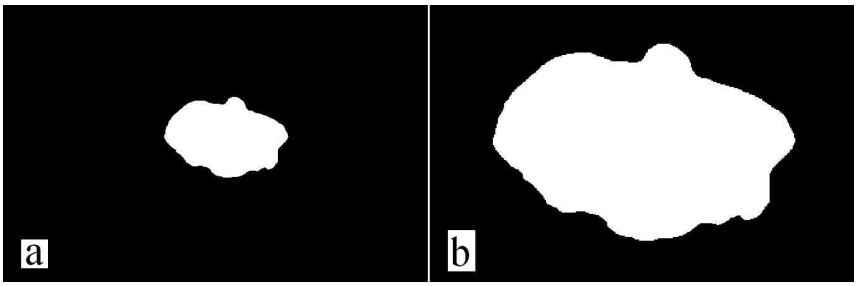

Figure 3.24: Segmentation result (a) before and (b) after standardisation.

3.3.3 Optimal Choice of Window and Bandwidth

As previously mentioned, the two parameters involved in the proposed Hybrid method are window size (W) and the bandwidth factor (B). Here we provide a new analytical framework in order to determine the optimal value for these two parameters. A comprehensive set of experiments is performed on the set of 55 dermoscopy images, with W ranging from 30 to 70 and B ranging from 10% to 90% (steps of 10). Different borders can be obtained by using different values of W and B. Consequently, 45 borders are obtained for each dermoscopy image. To evaluate the results, each border is compared with the ground truth in terms of the five previously defined evaluation metrics of sensitivity (Equation 3.9), specificity(Equation 3.10), accuracy(Equation 3.11), similarity(Equation 3.13) and border error (Equation 3.22).

Statistical Analysis:

Figures 3.25–3.27 show the mean value and 95% confidence interval (CI) for metrics of sensitivity, specificity, accuracy, similarity and border error for different values of W and B parameters, respectively. We set levels of acceptability to the lower bound of the confidence interval. Due to the importance of sensitivity, we start the analysis from this metric. As shown in Figure 3.25, given the level of acceptability of 87% for sensitivity, 22 sets of W and B, out of 45, are selected. These are marked by filled circles in the graph. With the level of acceptability of 93% for specificity, nine sets from the previous 22 sets are nominated as illustrated in Figure 3.25. Having 93.5% level of acceptability for accuracy, as shown in Figure 3.26, the set is narrowed down to six sets of W and B, *i.e.*, (30,10), (30,20), (40,20), (30,30), (40,30) and (40,40). Having these six sets, two metrics of similarity and border error are studied to make the final decision about the optimal value for B and W.

As shown in Figure 3.27 according to both of these metrics the pair of (30,30) gains the best results.

We acknowledge the difference between the optimal setting for W and B found in our previous experiment (Section 3.2.3) and what is determined here. The difference is attributed to our new choice of ground truth. After extensive discussions with experienced dermatologists we decided to use in this experiment the union of the four manual borders drawn by the dermatologists. The implication here is that the outmost detectable border is used as the ground truth and therefore as the reference for optimising our method. This rational behind this choice is to minimise the risk of excluding any part of the lesion from the final segmentation result.

Performance Index:

It is often the case that different methods provide different results when calculating the five evaluation metrics. The problem is which method provides the best result when all of the five metrics are considered simultaneously. To answer this, we introduce the following procedure. First, we define a metric called Performance Index (PI):

$$PI = \frac{Sns + Spc + Acc + Sml + BErr}{5} \tag{3.23}$$

where Sns, Spc, Acc, Sml and BErr refer to sensitivity, specificity, accuracy, similarity and 100%-borderError, respectively. Second, the index is calculated for different sizes of W and B over the image set of 55 dermoscopy images, using the mean value of the metrics. Third, for each W, a family of PIs versus B is plotted, as shown in Figure 3.28. Fourth, we choose the maximum PI point which reveals the optimal combination of W and B.

As shown in Figure 3.28, the maximum value of PI is obtained for (W=30, B=30). This result is in accordance with the results we obtained from the statistical analysis above. We have chosen (W30, B30) to proceed with. However, there are other combinations for which PI is very similar, *e.g.* (W30, B20) or (W30, B10). As these combinations provide better sensitivity, they would also be reasonable options to be considered.

Cross Validation:

In order to provide a stronger proof for the optimal B and W identified through the statistical analysis and the proposed performance index, we also perform a 11-fold cross-validation process [199], wherein the data set of 55 images is iteratively partitioned into a 50-image train subset and a 5-image test subset, resulting in 11 sets with unique combinations of test and train data. For each of the test sets, a family of PI curves for different W and B is plotted, as illustrated in Figure 3.29, which shows that all train sets converge to the value of (30,30) for B and W, except for set 6, where the optimal PI occurs when W=30 and B=20, which is close to the PI value for (W=30, B=30). Table 3.10 shows the resultant PI for images and the corresponding mean and standard deviation for each test

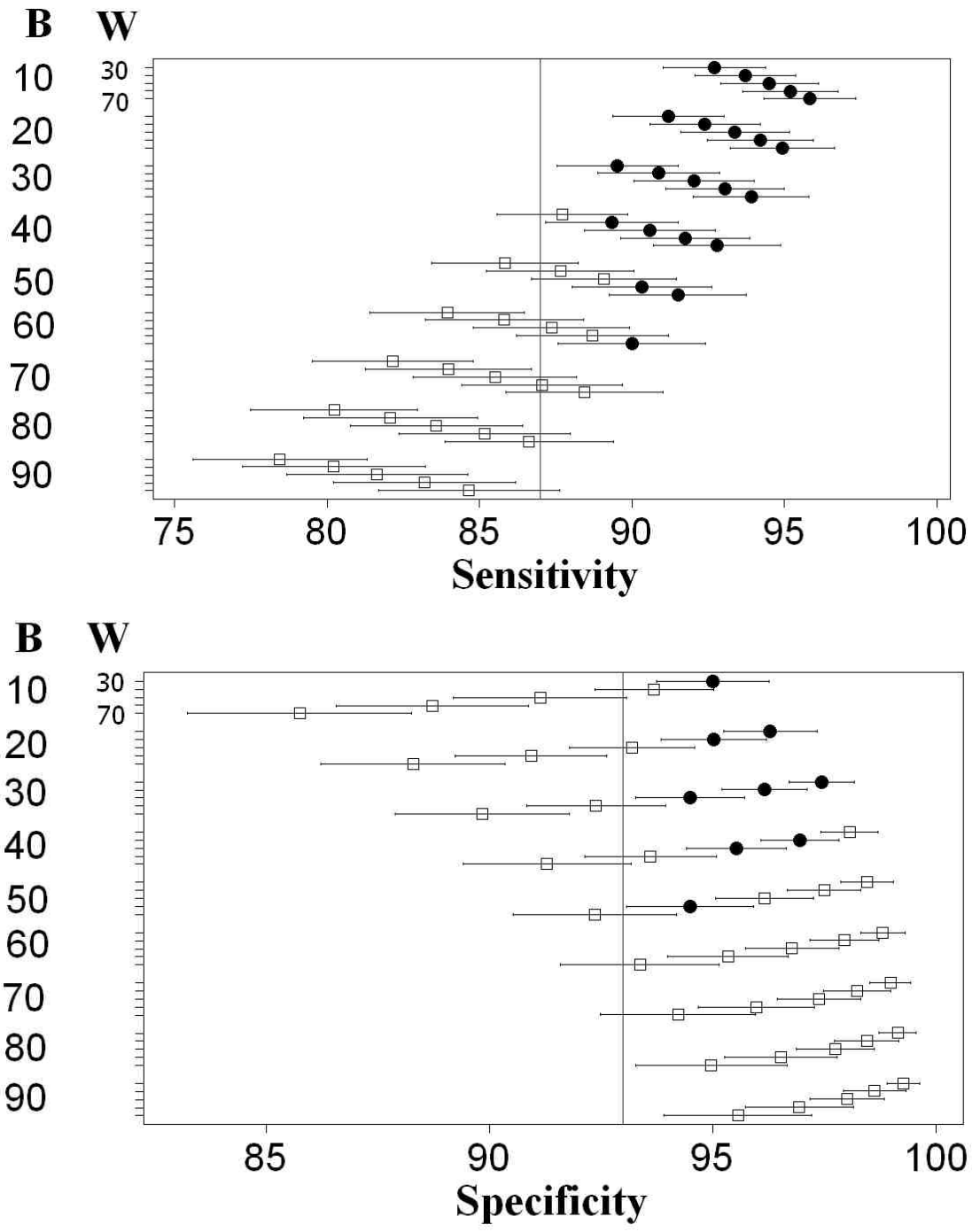

Figure 3.25: Mean and 95% confidence interval of sensitivity and specificity metrics for various W and B values over the image set.

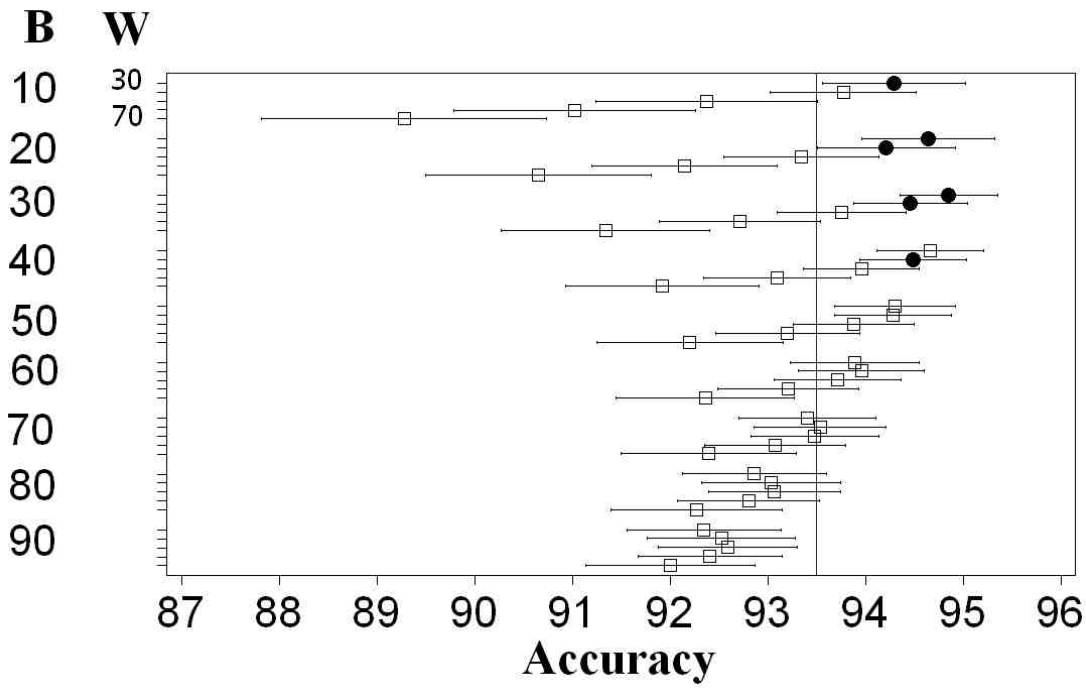

Figure 3.26: Mean and 95% confidence interval of accuracy metrics for different values of B and W over the image set.

set using the optimal setting of (W=30, B=30), which demonstrates the acceptability of the identified setting.

3.3.4 Comparison between Global and Hybrid

Figure 3.30 shows the automatic borders obtained by the global and the hybrid approach and also the corresponding ground-truth border drawn by dermatologists (the union of the three manual borders) for a sample lesion. It shows that the hybrid approach produces a border much closer to the ground truth, compared to the border produced by the global thresholding method. In order to objectively show the extent of improvement by evaluation metrics, when the lesion border is expanded from the core-lesion area (the result of the Global thresholding method) to the edge-lesion area (the result of the Hybrid thresholding method with the optimal values of B=30 and W=30), the mean difference between these two methods is calculated.

Table 3.11 gives the 95% confidence intervals for the mean for each metric, for global and hybrid (W30, B30) methods, and also the results of mean difference between these two approaches. In the Hybrid column, it is shown that the average values for each metric are clearly in an acceptable range. The mean specificity is 97.4; the 95% confidence interval is 96.9 to 98.1, and we can be confident that sensitivity will be high when the parameters

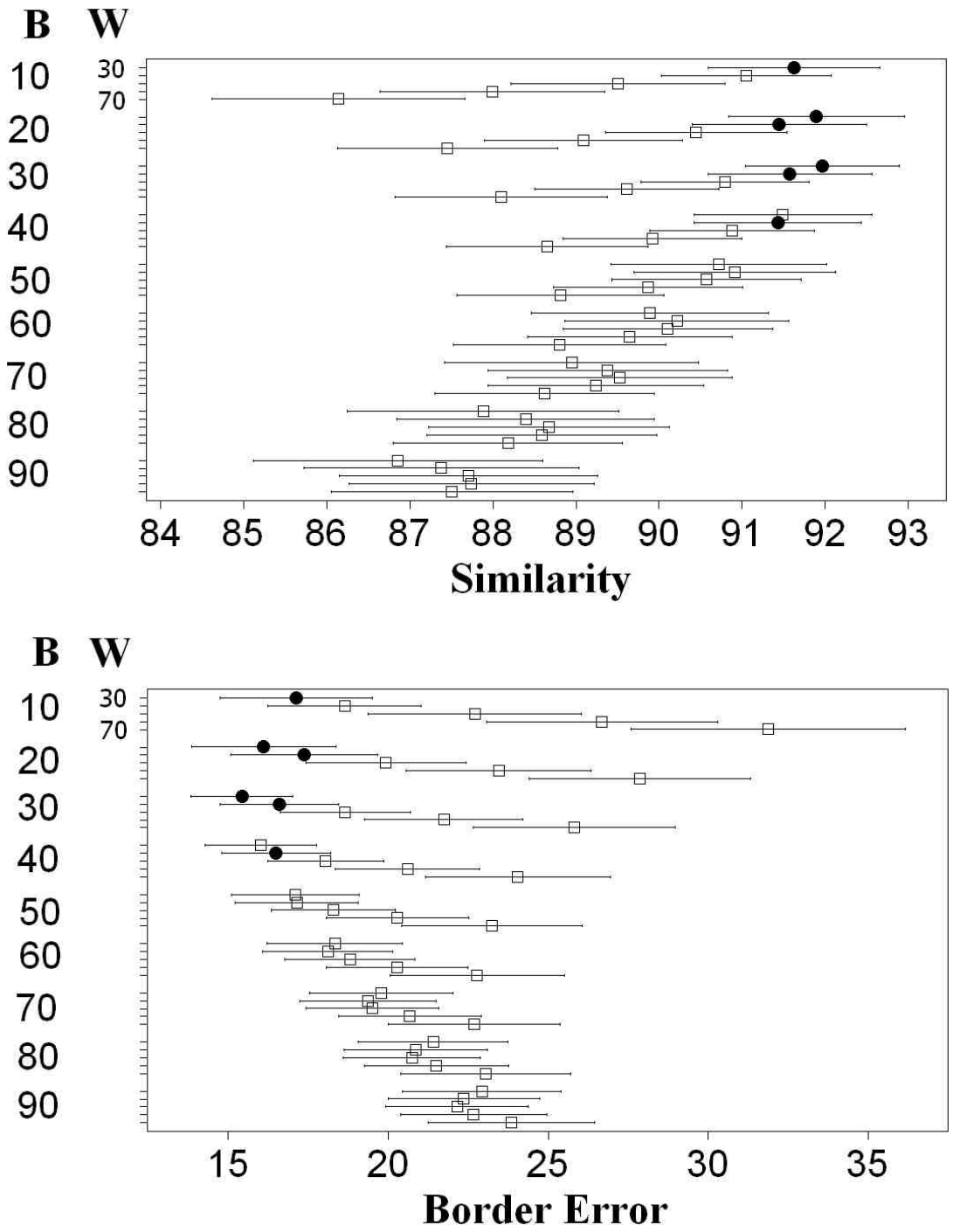

Figure 3.27: Mean and 95% confidence interval of similarity and border error metrics for various W and B values over the image set.

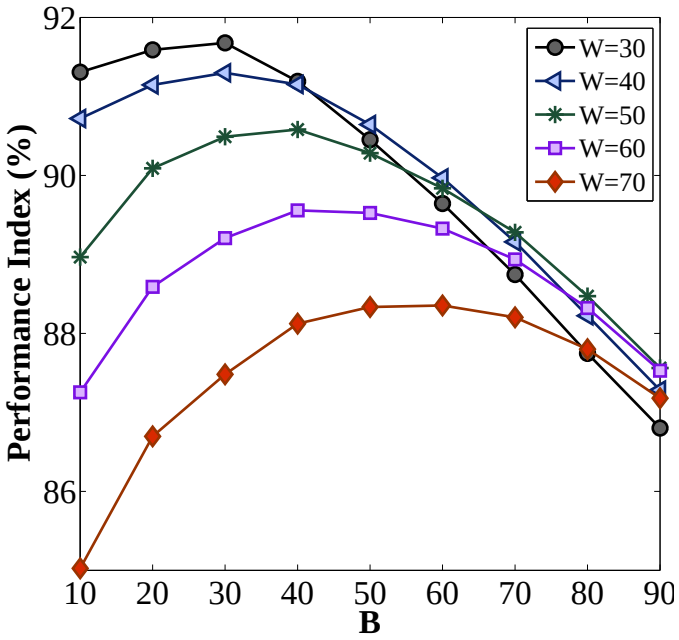

Figure 3.28: Performance Index for different values of B and W over the image set.

(W=30, B=30) are used.

Table 3.11 also demonstrates that with respect to all metrics there is a significant improvement, over the Global method, when the Hybrid thresholding method is applied, though specificity is almost the same. It is worth pointing out that specificity is of a less importance than accuracy and sensitivity in the opinion of experienced dermatologists. That is mainly due to the fact that they are more concerned not to leave out any part of the lesion, than erroneously including some part of the surrounding skin in the segmentation result. Moreover, the mean level of specificity is very high in both approaches (Global and Hybrid). As shown in Table 3.11 sensitivity, similarity and accuracy are increased by 16.17%, 8% and 3.95%, respectively and border error is reduced by 11.84% which demonstrate the significant superiority of the hybrid approach. The Performance Index is also elevated by 5.6% which proves the better performance of the hybrid thresholding method.

3.3.5 Conclusion

In this section we provided a new analytical framework, including statistical analysis, optimisation and cross validation, for determining the optimal parameters for the proposed Hybrid border detection method. We also compared the two proposed border detection methods; Global and Hybrid, and demonstrated the superiority of the latter. The comparison was performed through a newly introduced evaluation metric: Performance Index, which is composed of standard metrics of sensitivity, specificity, accuracy, similarity, and

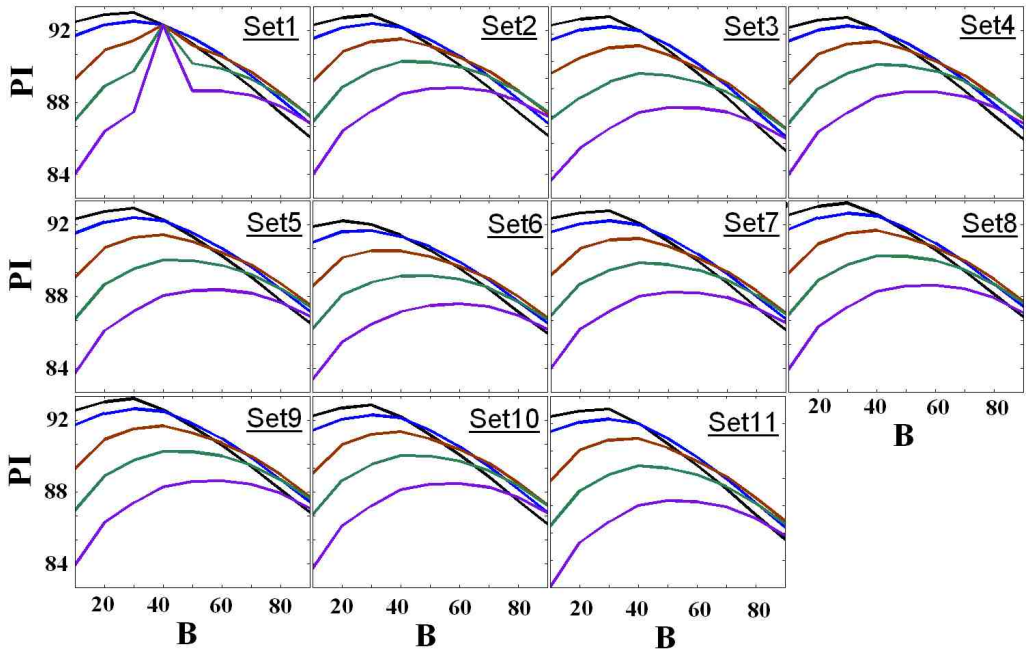

Figure 3.29: 11-fold cross validation using Performance Index evaluation metric.

border error. The experiment was performed on 55 high-resolution dermoscopy images which clearly showed the significant advantages of the proposed Hybrid method over the Global thresholding approach.

3.4 Objective Evaluation: Weighted Performance Index

Objective evaluation of border detection methods has not been explored in depth [78]. Existing evaluation methods are either through visual assessment of the detected borders by dermatologists, which suffers from subjectivity, or through an objective evaluation, where the closeness of an automatic border produced by the border detection method is compared with that manually drawn by dermatologists. With respect to objective evaluation methods, different metrics have been used; namely, sensitivity, specificity, accuracy, border error, similarity and precision, which have been applied in the previous sections of this chapter, as well as pixel misclassification probability [200] and normalised probabilistic rand index [201].

There exist two problems in the evaluation of border detection methods: first, the above-mentioned metrics are generic and have been widely applied in different domains. However, in application to border detection in dermoscopy images of melanoma lesions, it is crucial that dermatologists' perspectives are taken into account, which raises the need to customise the standard metrics to reflect their respective practical importance in the evaluation process. Second, it is often the case that a border detection method provides

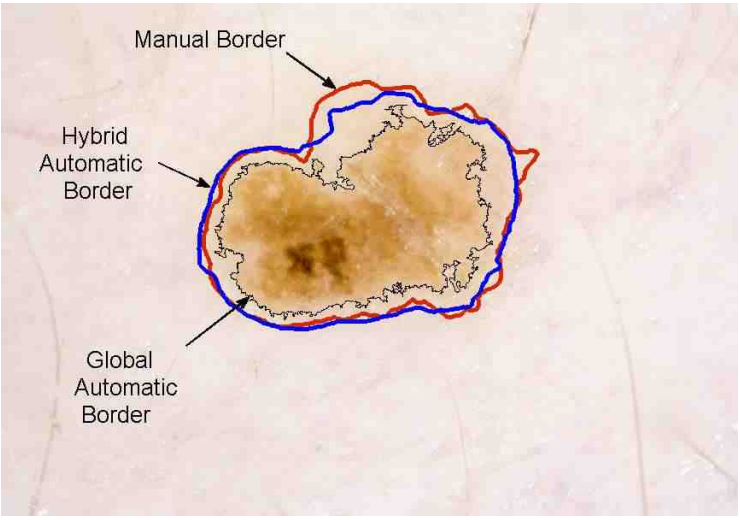

Figure 3.30: The Global and Hybrid borders produced by the automated methods and the manual border drawn by dermatologists.

a superior result according to one evaluation metric, yet is defeated by other methods with respect to another metric(s). In other words, there is no comprehensive metric for comparing different automated methods. Thus, comparison between border detection methods is not easy.

In the previous section, we proposed a Performance Index (PI) integrating the five standard metrics of sensitivity, specificity, accuracy, similarity, and border error. In this section we revise the index to incorporate the precision metric. Moreover, in order to provide evaluation metrics that are meaningful in the context of melanoma application, we introduce specific weightings into the six standard metrics and propose a new comprehensive metric called Weighted Performance Index for objective evaluation of border detection results in dermoscopy images.

3.4.1 Revised Performance Index

The PI metric previously defined is revised here to incorporate the precision metric as follows:

$$PI = \frac{Sns + Spc + Acc + Sml + BErr + Prc}{6} \tag{3.24}$$

where Sns, Spc, Acc, Sml, Prc and BErr refer to sensitivity, specificity, accuracy, similarity, 100%-borderError and precision, respectively.

Table 3.10: Performance Index for different images in each test set.

PI	**Image 1**	**Image 2**	**Image 3**	**Image 4**	**Image 5**	**Mean**	**Std.**
Set 1	90.59	89.62	92.01	91.41	91.63	91.05	0.95
Set 2	91.47	93.80	91.43	92.58	91.40	92.14	1.05
Set 3	95.27	94.41	92.26	87.99	91.17	92.22	2.87
Set 4	94.91	92.58	91.36	95.28	91.12	93.05	1.94
Set 5	89.53	91.47	91.81	89.37	96.28	91.69	2.79
Set 6	92.81	90.90	80.83	85.82	92.22	88.51	5.10
Set 7	92.43	90.53	94.14	85.06	96.67	91.77	4.37
Set 8	92.96	95.35	92.59	94.19	89.28	92.87	2.28
Set 9	90.08	96.48	83.89	94.00	83.24	89.54	5.91
Set 10	93.50	89.52	92.57	90.91	94.95	92.29	2.13
Set 11	94.99	94.44	86.89	94.35	95.87	93.31	3.63

Table 3.11: Evaluation metrics (mean ±margin of error) for Global and Hybrid thresholding methods and the difference between them.

	Global	Hybrid	Improvement
Accuracy	90.90 ±0.7	94.85 ±0.5	3.95 ±0.7
Sensitivity	73.35 ±2.6	89.52 ±1.9	16.17 ±1.5
Specificity	99.68 ±0.1	97.44 ±0.7	-2.23 ±0.6
Similarity	83.94 ±1.8	91.97 ±0.9	8.02 ±1.4
Border Error	27.24 ± 2.4	15.4 ±1.5	11.84 ±2.0
Performance Index	84.12 ±1.5	91.67 ±0.9	5.6 ±1.0

3.4.2 Weighted Evaluation Metrics

Objective evaluation of melanoma lesion border detection methods requires metrics reflecting the particular characteristics of the melanoma diagnosis application. Our experience with dermatologists has consistently shown TP to have highest importance compared with the other three parameters of TN, FP and FN, since dermatologists want the entire lesion to be included in the automatic border. On the other hand, FP (the areas included by the automated method, yet excluded by the gold standard) has a minor degree of importance compared with FN (the areas excluded by the automated method, yet included by the gold standard). Accordingly, after extensive consultation with dermatologists, we attach a weighting of 1.5 to TP to reflect its overall importance. Furthermore, to emphasise the importance of FN over FP, in those metrics which include both FN and FP, we assign a factor of 0.5 to FP. As a result, the new set of equations is given by:

$$WSensitivity = \frac{1.5TP}{1.5TP + FN} \times 100\% \tag{3.25}$$

$$WAccuracy = \frac{1.5TP + TN}{1.5TP + 0.5FP + FN + TN} \times 100\% \tag{3.26}$$

$$WSimilarity = \frac{3TP}{3TP + FN + 0.5FP} \times 100\% \tag{3.27}$$

$$WBorderError = \frac{0.5FP + FN}{1.5TP + FN} \times 100\% \tag{3.28}$$

$$WPrecision = \frac{1.5TP}{1.5TP + FP} \times 100\% \tag{3.29}$$

where WSensitivity, for instance, stands for weighted sensitivity. The specificity metric does not contain TP factor. Also, FP and FN do not appear simultaneously in this metric. Accordingly it remains unchanged as in Equation 3.10.

3.4.3 Weighted Performance Index

The performance of existing border detection methods have been commonly evaluated by using either a single metric (*e.g.* border error) or a pair of metrics (*e.g.* sensitivity and specificity). As each of the metrics has its specific meanings and implications, it becomes difficult to interpret the results objectively. This problem is compounded when a particular border detection method yields superior results with respect to one evaluation metric but is defeated by other methods according to other metrics. Thus, the question of which method provides the best possible result when all metrics are considered simultaneously has not been previously resolved. To address this a comprehensive metric, Weighted Performance Index (WPI), is defined to objectively evaluate border detection methods for dermoscopy images. The index takes into account the weighted metrics defined in Equations 3.10 and 3.25–3.29. WPI is in fact a weighted average, where the weights are imbedded within each metric.

$$WPI = \frac{WSns + Spc + WAcc + WSml + WBErr + WPrc}{6} \tag{3.30}$$

where WSns, Spc, WAcc, WSml, WBErr and WPrc refer to WSensitivity, specificity, WAccuracy, WSimilarity, 100%-WBorderError and WPrecision, respectively.

3.4.4 Image Set and Gold Standard

To assess the effectiveness of the proposed evaluation indices, they are tested on the same set of 55 high-resolution dermoscopy images used in the previous section. In this experiment, however, we had an additional experienced dermatologist draw the manual borders for us. Thus, to validate the borders produced by existing methods, manual borders are independently drawn by four expert dermatologists using Wacom Intuos A4 size Tablet PC, and the union of those borders is taken as the ground truth. The choice of union for integrating the manual borders, is in line with the emphasis we put on the TP factor in the proposed evaluation metrics.

Furthermore, we have applied the image standardisation procedure discussed in Section 3.3.2 to balance the effect of large TN parameter and normalise the accuracy and specificity metrics.

3.4.5 Optimisation of the Parameters

The proposed metrics of PI and WPI can be used to optimise automated border detection methods by tuning their parameters. For instance, in the Hybrid thresholding method (Section 3.2) the two parameters of window size (W) and bandwidth factor (B) are determined.

To optimise settings of (W, B) for the method, we perform a set of experiments on the image set of 55 high-resolution dermoscopy images, with W varying from 20 to 70 and B ranging from 10% to 90% (steps of 10), which results in 54 borders for each image. Each resultant border is then compared with the ground truth and the standard and weighted metrics of sensitivity, specificity, accuracy, similarity, border error and precision, and the corresponding performance index and weighted performance index are calculated. In the ensuing the optimal pairs of B and W are obtained.

Performance Index and Weighted Performance Index:

Performance index and weighted performance index are calculated for various W and B values over the image set using the value of the standard and weighted metrics of sensitivity, specificity, accuracy, similarity, border error and precision according to Equations 3.24–3.30 and are averaged over 55 images.

For each W, the family of mean PIs and mean WPIs versus B is plotted, as shown in Figures 3.31 and 3.32. The two graphs of PI and WPI metrics are coherent and both reveal the optimal setting of (30, 30) for B and W, as shown in Figures 3.31 and 3.32.

However, there is a distinction between the two analyses, *i.e.*, standard vs. weighted metrics. The performance index, which is based on standard metrics, yields a lower result than weighted performance index, which is based on proposed weighted metrics. For example for W30B30, the mean value of PI is 92.30, whereas the mean value of WPI is 94.90. Thus, the weighted metrics, which are defined to reflect the dermatologists' perspectives, show a higher degree of agreement between automatic and manual borders, compared with standard metrics.

Statistical Analysis:

Figures 3.33–3.35 show the mean value and 95% confidence interval (CI) for metrics of weighted sensitivity, specificity, weighted accuracy, weighted similarity, weighted border error and the weighted precision for various values of W and B parameters, respectively. We set levels of acceptability for the lower bound of the confidence interval. The levels are arbitrary but reasonable, and are a helpful guide for identifying acceptable parameter values across a range of metrics.

Due to the importance of sensitivity, we start the analysis from this metric. As shown in Figure 3.33, given the level of acceptability of 90% for weighted sensitivity, 26

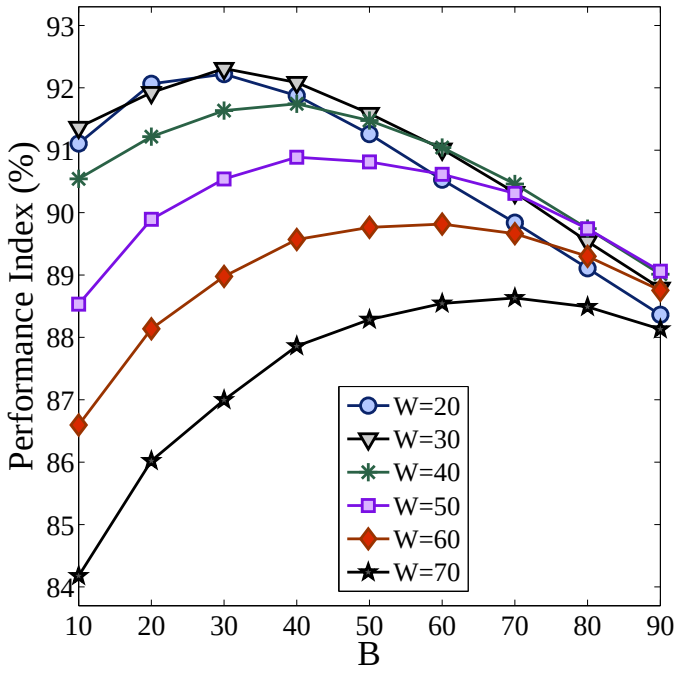

Figure 3.31: Performance Index for various W and B values over the image set

sets of W and B, out of 54, are selected which are marked by filled circles in the graph. Having the level of acceptability of 94% for specificity, 9 sets from the previous 26 sets are nominated as illustrated in Figure 3.33.

For the level of acceptability of 95.5% for weighted accuracy, as shown in Figure 3.34, seven pairs of W and B meet the criteria. As illustrated in Figures 3.34 with the level of acceptability of 94.5% for weighted similarity, these seven pairs are further narrowed down to 5 sets of (20,20), (30,20), (40,20), (30,30), (40,30).

Having these five sets, the weighted border error metric is studied and as shown in Figure 3.35 the above-mentioned 5 pairs are highly competitive where their respective 95% confidence of mean are very similar. To make the final decision about the optimal value for B and W, weighted precision metric is investigated. As shown in Figures 3.35 the pair of (30,30) gains the best result, followed by (20,20). This result is in accordance with the results obtained from the performance index and weighted performance index in the previous section.

Cross Validation:

In order to provide a stronger proof for the optimal B and W identified through the statistical analysis and the proposed Weighted Performance Index, we also perform a 11-

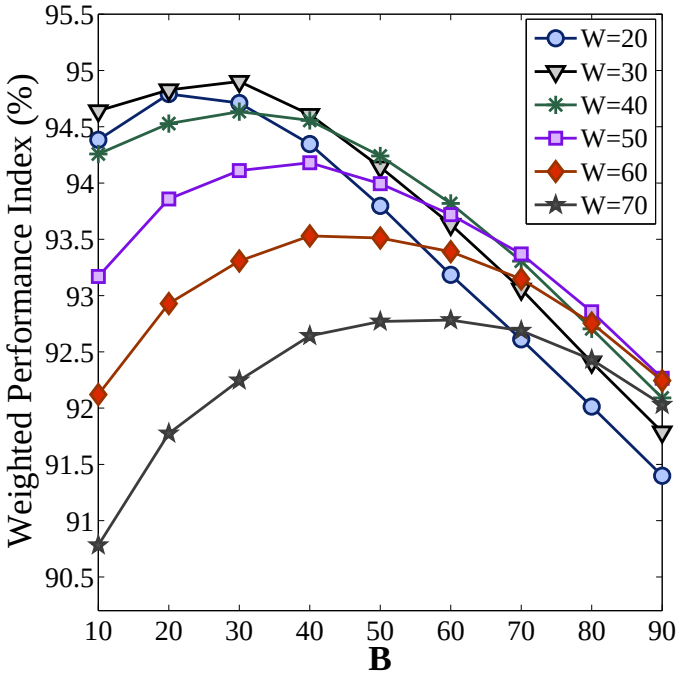

Figure 3.32: Weighted Performance Index for various W and B values over the image set

fold cross-validation process, wherein the data set of 55 images is iteratively partitioned into a 50-image train subset and a 5-image test subset, resulting in 11 sets with unique combinations of test and train data. For each of the test sets, a family of WPI curves for different W and B is plotted, as illustrated in Figure 3.36, which shows that all training sets converge to the value of (30,30) for B and W, except for set 6. According to set 6, the optimal WPI is (W=30, B=20), yet its resulting WPI is almost similar to the WPI value for (W=30, B=30). Table 3.12 shows the resultant WPI for images and the corresponding mean and standard deviation for each test set using the optimal setting of (W=30, B=30), which demonstrates the acceptability of the identified setting.

3.4.6 Comparison between Automated Methods

The proposed WPI metric has been used to evaluate five recent border detection methods applied to the image set of 55 dermoscopy images. These methods are dermatologist-like tumor extraction algorithm (DTEA) [26], JSEG [95], KPP [89], the Global method on optimised colour channel of XoYoR (Section 3.1), and the Hybrid method (Section 3.2) with optimal parameters of (W=30,B=30).

Table 3.13 gives the 95% confidence intervals for the mean of the proposed metrics. As shown, a method may perform better than others with respect to some of the metrics, yet yields lower results with respect to others, *e.g.* the DTEA method gains the highest

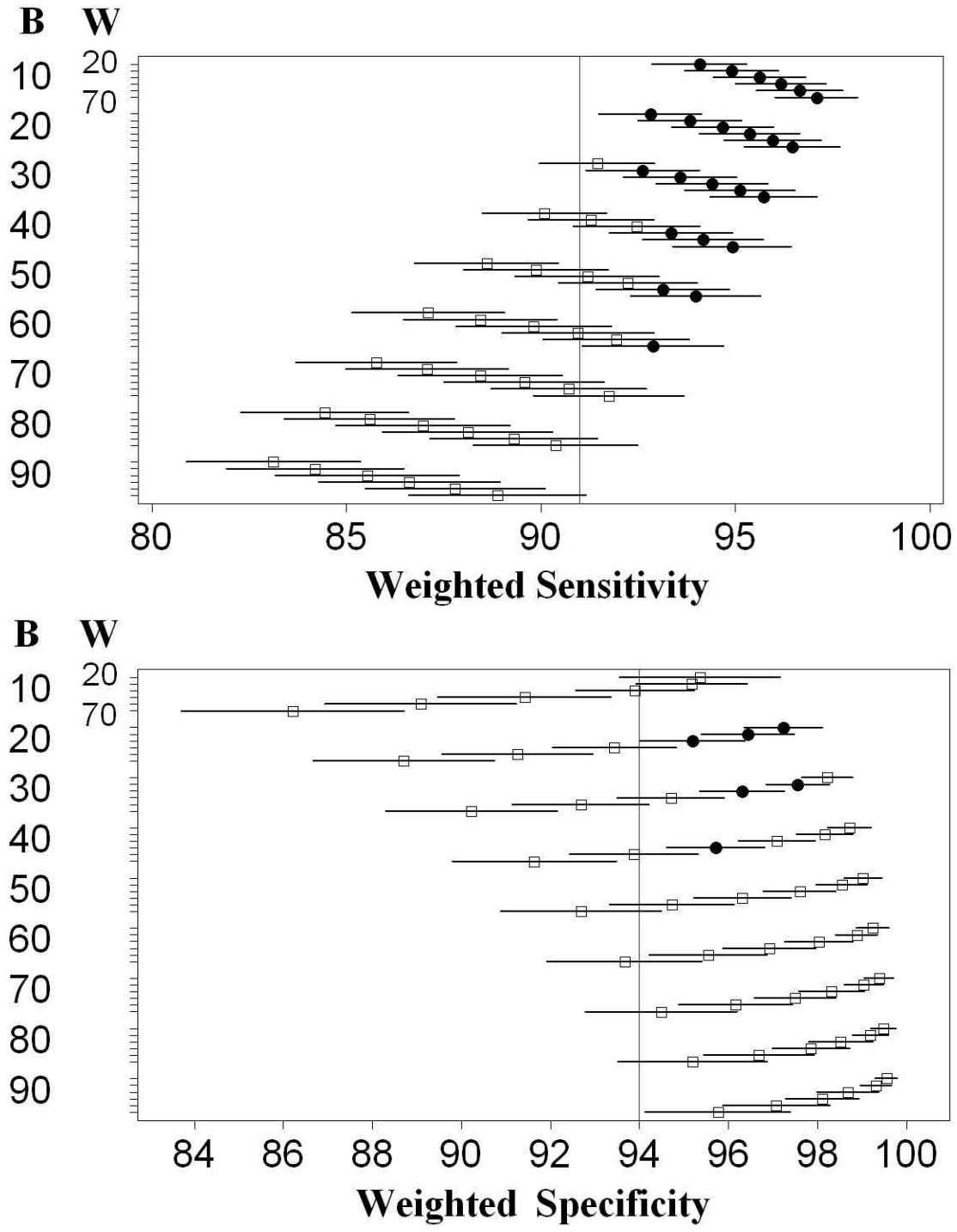

Figure 3.33: Mean and 95% confidence interval of weighted sensitivity and weighted specificity metrics for various W and B values over the image set.

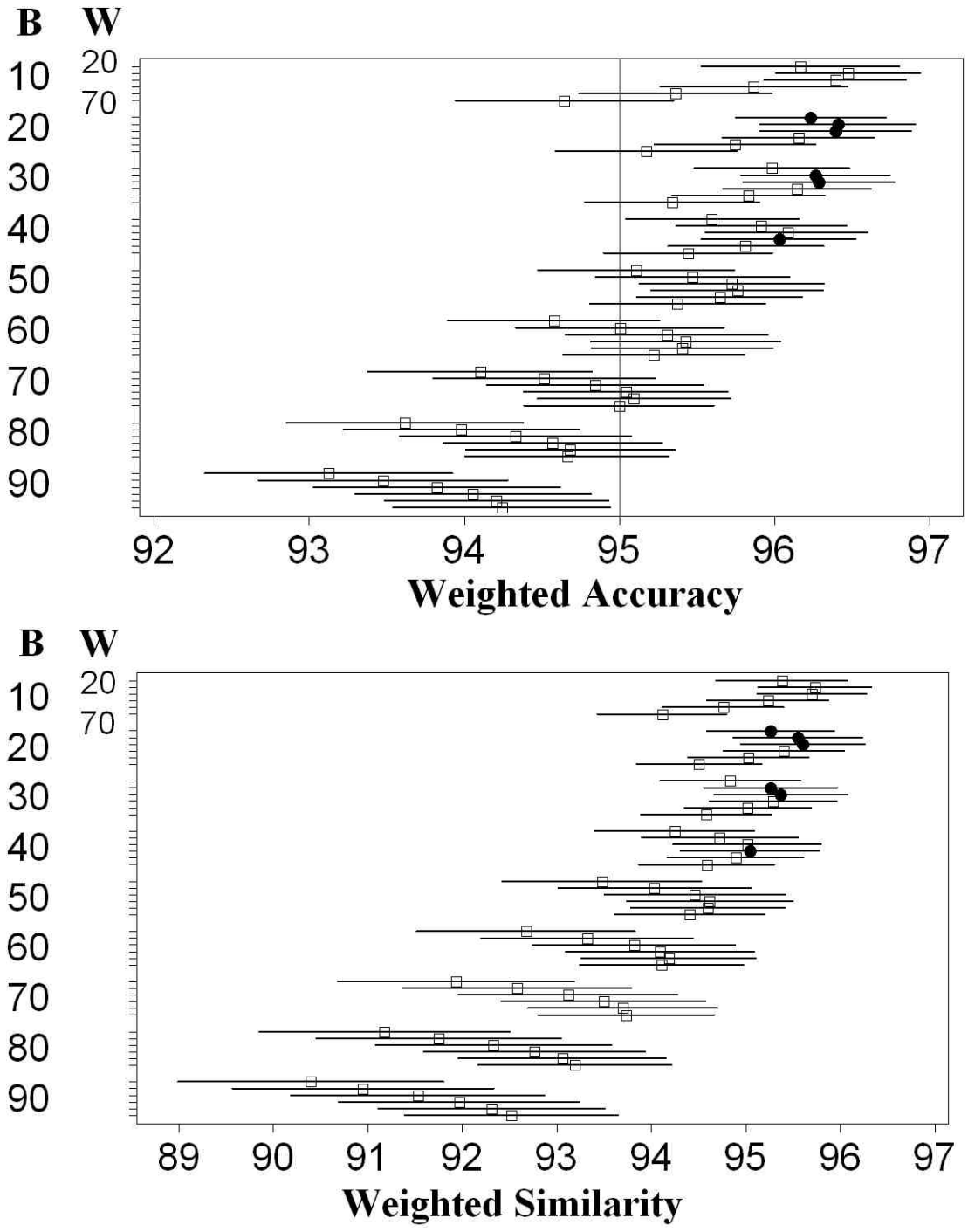

Figure 3.34: Mean and 95% confidence interval of weighted accuracy and weighted similarity metrics for various W and B values over the image set.

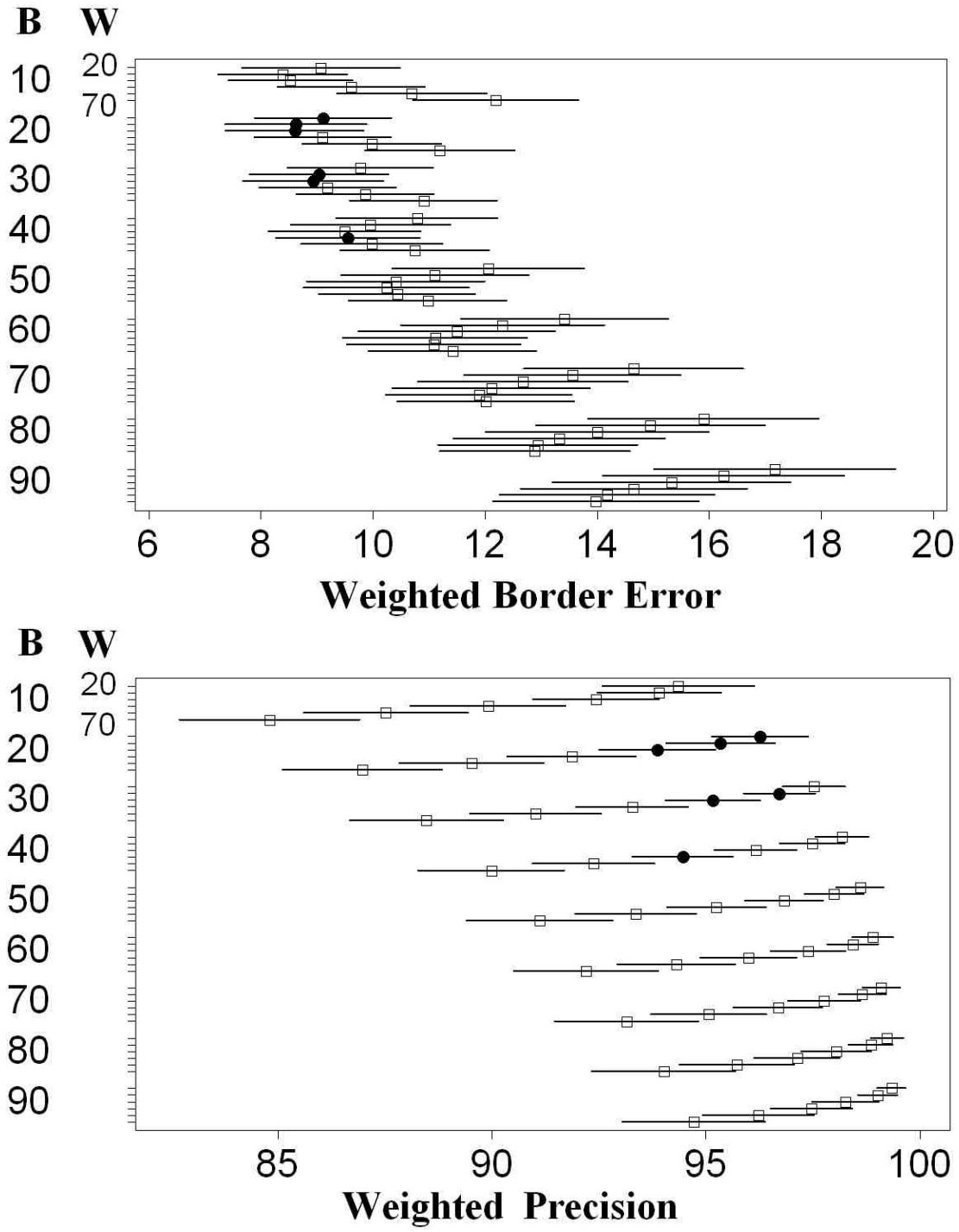

Figure 3.35: Mean and 95% confidence interval of weighted border error and weighted precision metrics for various W and B values over the image set.

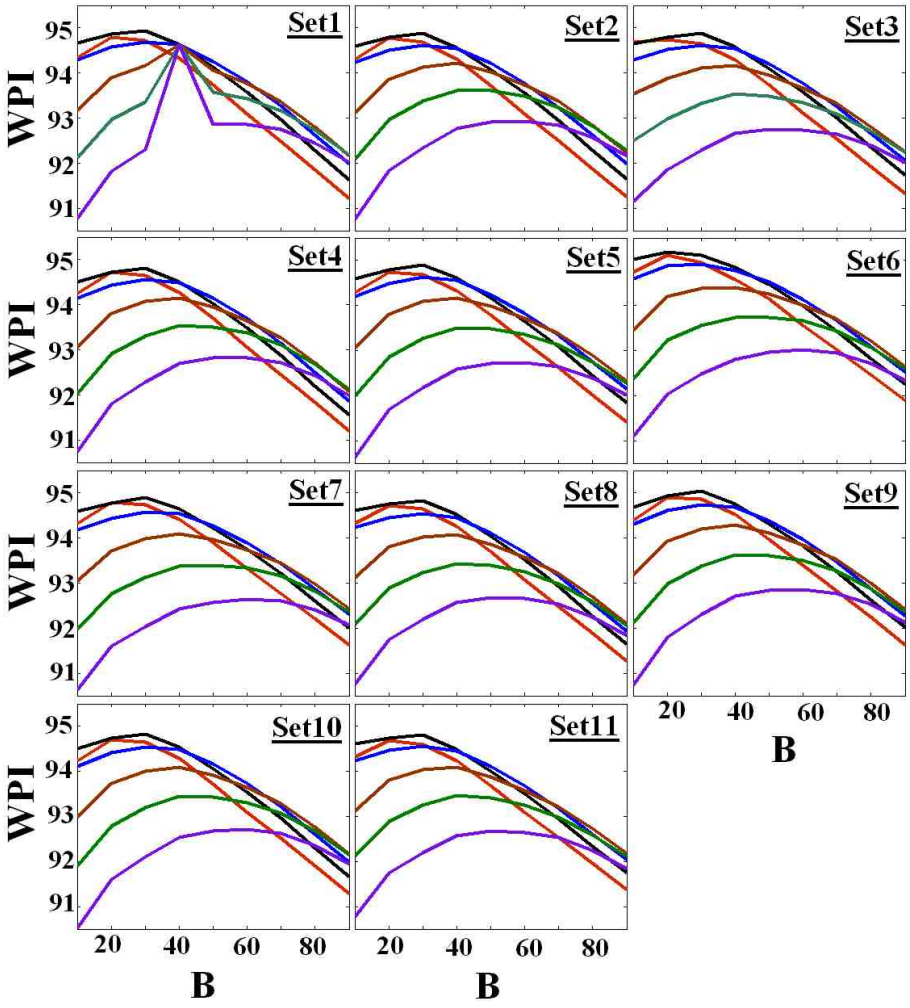

Figure 3.36: 11-fold cross validation using Weighted Performance Index evaluation metric

specificity and weighted-precision, yet it is overcome by the Hybrid method with respect to weighted metrics of sensitivity, accuracy, similarity and border error. For these reasons it has not hitherto been easy to provide an overall objective judgment as to which of the five methods is more suited for border detection of dermoscopy images. However, the use of the proposed weighted performance index, as shown in the bottom row of Table 3.13, facilitates such a judgment, which can easily be made by comparing values of the calculated WPI's. According to the obtained WPI, the Hybrid method provides the best segmentation result, followed by JSEG, DTEA, Global thresholding and KPP methods.

Table 3.14 gives the degree of superiority of each method over other methods, *e.g.* the second row and third column shows that Hybrid method overcomes JSEG (+2.8) such that we can be 95% confident that the true difference is between 1.80 and 3.80. The fourth

Table 3.12: Weighted Performance Index for different images in each test set.

WPI	**Image 1**	**Image 2**	**Image 3**	**Image 4**	**Image 5**	**Mean**	**Std.**
Set 1	94.29	93.49	95.22	94.63	94.80	94.49	0.64
Set 2	94.78	96.09	94.80	95.36	94.78	95.16	0.57
Set 3	97.13	96.59	95.29	92.43	94.51	95.19	1.85
Set 4	96.95	95.49	94.78	97.04	94.58	95.77	1.17
Set 5	93.50	94.79	95.10	93.56	97.66	94.92	1.69
Set 6	95.55	94.52	87.83	91.05	95.10	92.81	3.30
Set 7	95.34	94.30	96.42	90.75	98.02	94.96	2.73
Set 8	95.67	97.20	95.32	96.50	93.49	95.64	1.40
Set 9	93.91	97.87	89.95	96.36	89.51	93.52	3.74
Set 10	95.91	93.61	96.62	94.52	97.15	95.56	1.47
Set 11	96.95	96.70	91.97	96.45	97.39	95.89	2.22

Table 3.13: Evaluation metrics (mean ±margin of error) for five border detection methods; Hybrid, JSEG, DTEA, Global, KPP

	Hybrid	JSEG	DTEA	Global	KPP
Weighted Accuracy	96.27 ±0.4	93.73 ±0.9	92.49 ±1.0	92.29 ±0.7	91.28 ±1.3
Weighted Sensitivity	92.62 ±1.4	84.48 ±2.5	81.16 ±2.6	80.89 ±2.1	78.13 ±3.7
Specificity	97.56 ±0.7	99.59 ±0.2	99.79 ±0.1	99.63 ±0.2	99.17 ±0.6
Weighted Precision	96.72 ±0.8	99.43 ±0.3	99.70 ±0.1	99.45 ±0.2	98.99 ±0.7
Weighted Similarity	95.26 ±0.7	91.14 ±1.6	89.20 ±1.6	89.08 ±1.4	86.76 ±2.4
Weighted BorderError	9.04 ±1.2	15.79 ±2.5	18.98 ±2.5	19.35 ±2.1	22.39 ±3.5
WPI	94.90 ±0.5	92.09 ±1.2	90.56 ±1.2	90.33 ±1.0	88.65 ±1.7
PI	92.30 ±0.7	89.32 ±1.4	87.44 ±1.5	87.03 ±1.2	85.10 ±1.9

row and third column shows DTEA is defeated by JSEG (-1.53) with true mean difference between 0.43 and 2.63.

3.4.7 Conclusion

This section presented a novel approach for objective evaluation of border detection methods in dermoscopy images. In order to provide evaluation metrics meaningful in the context of melanoma application, we introduced specific weightings into the standard metrics of sensitivity, specificity, accuracy, border error, similarity and precision. Moreover, a comprehensive metric, Weighted Performance Index (WPI), was proposed to facilitate comparison between different methods. The proposed WPI has also been used for the optimisation of the Hybrid border detection method proposed in Section 3.2. The effectiveness of the proposed evaluation approach was demonstrated by applying five recent border detection methods on a set of 55 high-resolution dermoscopy images using the union of four sets of dermatologist-drawn borders as the ground truth. It was also shown that the weighted metrics, which are defined to reflect the dermatologists' perspectives, show a higher degree of agreement between automatic and manual borders, compared with

Table 3.14: Mean ±margin of error of Weighted Performance Index for comparisons of automated methods

	Hybrid	JSEG	DTEA	Global	KPP
Hybrid	0	2.80 ±1.0	4.34 ±1.0	4.57 ±0.8	6.24 ±1.5
JSEG	-2.80 ±1.0	0	1.53 ±1.1	1.76 ±1.1	3.43 ±1.5
DTEA	-4.34 ±1.0	-1.53 ±1.1	0	0.22 ±1.0	1.90 ±1.3
Global	-4.57 ±0.8	-1.76 ±1.1	-0.22 ±1.0	0	1.67 ±1.5
KPP	-6.24 ±1.5	-3.43 ±1.5	-1.90 ±1.3	-1.67 ±1.5	0

standard metrics.

3.5 Objective Evaluation: Optimised Weighted Performance Index

Prior to our reported results, no systematic evaluation procedure for the performance assessment of the border-detection methods in dermoscopy images, which could reflect dermatologists' perspective in the melanoma recognition process, had been thoroughly investigated in the literature. Quantitative evaluation of border detection methods is currently performed by using different statistical metrics of sensitivity, specificity, accuracy, border error, similarity, precision, etc. In the previous section, we addressed two issues with respect to current evaluation methodologies; (1) using various metrics is inherently problematic due to different characteristics of each metric, *i.e.*, a border detection method might overcome another method in terms of one metric, yet this effect can be reversed using a different metric, (2) the existing metrics are standard and do not reflect the particular specifications of melanoma diagnosis application.

We addressed these issues in the previous section by introducing two indices; namely, Performance Index and Weighted Performance Index. Those two indices, although having their own merits, can be improved upon by using the method proposed in this section. This is accomplished by formulating a constrained nonlinear multivariable optimisation problem and solving it by using the trust-region solver. A solution to the optimisation problem yields a set of weights to be introduced into the six standard evaluation metrics. The weights reflect the importance of each metric in the evaluation process, thus providing a basis for comparison with other methods. To demonstrate the effectiveness of the optimised index, it is used to evaluate five border detection methods applied on a set of 55 high-resolution dermoscopy images.

3.5.1 Proposed Evaluation Index

As discussed in Section 3.4, in order to customise the metrics to be meaningful for dermoscopy image analysis, and following consultation with expert dermatologists, the TP parameter was assigned the highest importance compared with the other three parameters

TN, FP and FN. On the other hand, FP (the areas included by the automated method, yet excluded by the gold standard) was deemed less important than FN (the area excluded by the automated method, yet included by the gold standard). So, a weight of 1.5 was assigned to TP to reflect its overall importance. Also, in those metrics which include both FN and FP, a factor of 0.5 was assigned to FP, to emphasise the importance of FN over FP. Consequently, we defined a weighted version of each of the six standard metrics of sensitivity, specificity, accuracy, border error, similarity and precision, and integrated them into the WPI metric (Equations 3.30) as a weighted average of those.

Obviously, the introduction of the above weights, even though based on expert opinion, is subjective and may not be the best possible way of reflecting the importance of such measures in the performance evaluation index. In this section we introduce a novel procedure based on constrained nonlinear optimisation to calculate these weights. In doing so, we attach specific weights to each of the TP, TN, FP and FN measures to result in the following set of weighted metrics (Equations 3.25–3.28). Then, we formulate and solve the optimisation problem to calculate an optimal set of weights as shown in the next section.

$$WSensitivity = \frac{\alpha TP}{\alpha TP + \sigma FN} \times 100\% \tag{3.31}$$

$$WSpecificity = \frac{\delta TN}{\delta TN + \beta FP} \times 100\% \tag{3.32}$$

$$WAccuracy = \frac{\alpha TP + \delta TN}{\alpha TP + \beta FP + \sigma FN + \delta TN} \times 100\% \tag{3.33}$$

$$WPrecision = \frac{\alpha TP}{\alpha TP + \beta FP} \times 100\% \tag{3.34}$$

$$WSimilarity = \frac{2\alpha TP}{2\alpha TP + \sigma FN + \beta FP} \times 100\% \tag{3.35}$$

$$WBorderError = \frac{\beta FP + \sigma FN}{\alpha TP + \sigma FN} \times 100\% \tag{3.36}$$

where α, β, σ and δ refer to weights of TP, FP, FN and TN, respectively.

Incorporating the weighted metrics defined above a comprehensive metric, Optimised Weighted Performance Index (OWPI), is defined to objectively evaluate border detection methods in dermoscopy Images. OWPI is in fact a weighted average, which takes into account the weighted metrics defined in Equations 3.31–3.36. The weights which are imbedded within each metric, would reflect the importance of each metric in the evaluation process.

$$OWPI = \frac{WSns + WSpc + WAcc + WPrc + WSml + WBErr}{6} \tag{3.37}$$

where WSns, WSpc, WAcc, WPrc, WSml and WBErr refer to WSensitivity, WSpecificity, WAccuracy, WPrecision, WSimilarity and 100%-WBorderError, respectively. By

substitution of the above-mentioned formula for each metric, OWPI is given by:

$$\begin{aligned} OWPI &= \frac{1}{6}\Big(\frac{\alpha TP}{\alpha TP+\sigma FN}+\frac{\delta TN}{\delta TN+\beta FP} \\ &+\frac{\alpha TP+\delta TN}{\alpha TP+\beta FP+\sigma FN+\delta TN} \\ &+\frac{\alpha TP}{\alpha TP+\beta FP}+\frac{2\alpha TP}{2\alpha TP+\sigma FN+\beta FP} \\ &+\frac{\alpha TP-\beta FP}{\alpha TP+\sigma FN}\Big)\times 100\% \end{aligned} \tag{3.38}$$

3.5.2 Determining the Weights

In order to determine the optimal weights for TP, FP, TN and FN in the proposed performance index, three methods have been applied, which are explained in the following.

Optimisation of OWPI:

In this method the optimal weights are computed by formulating a constrained nonlinear multivariable optimisation problem based on the trust-region approach [202] explained in the following section (Section 3.5.3). The solution to the optimisation problem finds the optimal set of weights that maximises the OWPI (Equation 3.38), taking into account certain constraints. The constraints are designed such that the dermatologists' requirements as discussed in Section 3.4.2 are met. A generalisation of the constraints used in the WPI problem in the previous section (Section 3.4), is made through the following set of inequalities:

$$\alpha+\epsilon_1\geq\beta\ ,\ \ \alpha+\epsilon_2\geq\sigma\ ,\ \ \alpha+\epsilon_3\geq\delta\ ,\ \ \sigma+\epsilon_4\geq\beta \tag{3.39}$$

These inequalities make sure that the right emphases and therefore weight is placed on the individual measure of TP, TN, FP, and FP, as recommended by the expert dermatologists. This set of inequalities can be expressed in a more compact way by the following matrix-vector inequalities:

$$\begin{bmatrix} -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 \\ 0 & 1 & -1 & 0 \end{bmatrix}\begin{bmatrix} \alpha \\ \beta \\ \sigma \\ \delta \end{bmatrix}\leq\begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \end{bmatrix} \tag{3.40}$$

The choice of the $\epsilon_i; i=1,2,3,4$ is a design consideration, where small values reflect less emphases on the importance of TP compared to large values. In this study it is set to 0.1. The optimisation problem may now be stated as follows:

$$max_{x\in R^4}\{OWPI(x)\} \tag{3.41}$$

Subject to the constraints

$$Cx<\epsilon \tag{3.42}$$

where

$$C = \begin{bmatrix} -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 \\ 0 & 1 & -1 & 0 \end{bmatrix} \quad , \quad x = [\alpha \;\; \beta \;\; \sigma \;\; \delta]^T \quad , \quad \epsilon = [\epsilon_1 \;\; \epsilon_2 \;\; \epsilon_3 \;\; \epsilon_4]^T \tag{3.43}$$

System of Nonlinear Equations:

The second approach is to determine the optimal weights of α, β, σ and δ, by solving the system of nonlinear Equations 3.44–3.47. These equations are derived by finding the maximum performance index with respect to each of TP, FP, FN and TN.

$$\begin{aligned} \frac{\partial OWPI}{\partial TP} &= \frac{2\alpha\sigma FN + \alpha\beta FP}{(\alpha TP + \sigma FN)^2} + \frac{\alpha\beta FP}{(\alpha TP + \beta FP)^2} \\ &\quad + \frac{\alpha(\beta FP + \sigma FN)}{(\alpha TP + \beta FP + \sigma FN + \delta TN)^2} \\ &\quad + \frac{2\alpha(\sigma FN + \beta FP)}{(2\alpha TP + \sigma FN + \beta FP)^2} \end{aligned} \tag{3.44}$$

$$\begin{aligned} \frac{\partial OWPI}{\partial FP} &= -\frac{\beta\delta TN}{(\delta TN + \beta FP)^2} \\ &\quad - \frac{\beta(\alpha TP + \delta TN)}{(\alpha TP + \beta FP + \sigma FN + \delta TN)^2} \\ &\quad - \frac{2\beta\alpha TP}{(2\alpha TP + \sigma FN + \beta FP)^2} \\ &\quad - \frac{\beta}{(\alpha TP + \sigma FN)^2} - \frac{\alpha\beta TP}{(\alpha TP + \beta FP)^2} \end{aligned} \tag{3.45}$$

$$\begin{aligned} \frac{\partial OWPI}{\partial FN} &= -\frac{2\alpha\sigma TP - \sigma\beta FP}{(\alpha TP + \sigma FN)^2} \\ &\quad - \frac{\sigma(\alpha FP + \delta TN)}{(\alpha TP + \beta FP + \sigma FN + \delta TN)^2} \\ &\quad - \frac{2\alpha\sigma TP}{(2\alpha TP + \sigma FN + +\beta FP)^2} \end{aligned} \tag{3.46}$$

$$\begin{aligned} \frac{\partial OWPI}{\partial TN} &= \frac{\delta\beta FP}{(\delta TN + \beta FP)^2} \\ &\quad + \frac{\delta(\beta FP + \sigma FN)}{(\alpha TP + \beta FP + \sigma FN + \delta TN)^2} \end{aligned} \tag{3.47}$$

As a maximum occurs when each derivative is set to zero, the optimisation problem becomes that of solving the system of nonlinear equations given in Equation 3.48, where the unknowns are the optimal weights.

$$\begin{cases} \frac{\partial OWPI}{\partial TP} = 0 \\ \\ \frac{\partial OWPI}{\partial FP} = 0 \\ \\ \frac{\partial OWPI}{\partial FN} = 0 \\ \\ \frac{\partial OWPI}{\partial TN} = 0 \end{cases} \Rightarrow \alpha, \beta, \sigma, \delta = ? \tag{3.48}$$

To solve this system of nonlinear equations, the SSE (Sum of Squares Error) as defined by Equation 3.49 is solved, subject to the constraints given in Equation 3.39, by using the interior point method [203] explained in the following section (Section 3.5.3).

$$SSE = \left(\frac{\partial OWPI}{\partial TP}\right)^2 + \left(\frac{\partial OWPI}{\partial FP}\right)^2 + \left(\frac{\partial OWPI}{\partial FN}\right)^2 + \left(\frac{\partial OWPI}{\partial TN}\right)^2 \tag{3.49}$$

The optimisation problem may now be stated as follows, subject to the constraints expressed in Equations 3.42 and 3.43:

$$min_{x \in R^4}\{SSE(x)\} \tag{3.50}$$

Grid Search:

Since the OWPI cost function has periodic behaviour (multiple local maxima), we have also applied a grid search to find the weights which lead to the maximum value of OWPI. The grid search is performed by the following steps:
(1) All possible combinations of $\alpha, \delta, \beta, \sigma$ within the range of [0.1,2] with the step of 0.1 and applying the constraints given in Equation 3.39, are produced.
(2) The OWPI for each combination of the weights for each image is calculated and averaged over the image set, to obtain an estimate of the population of the given samples (images).
(3) The maximum OWPI is determined and the respective weights are identified.

3.5.3 Optimisation Background

For completeness, in this section a summary of the optimisation algorithms used (trust-region and interior-point) is provided:

Trust-Region Methods (Un-constrained Optimisation)

Trust-region methods are an important class of iterative algorithms aimed at solving non-linear constrained optimisation problems. They are reliable, efficient, and free of ad-hoc decisions. In addition they exhibit strong convergence characteristics and are easy to implement in software. A distinguishing feature of trust-region methods is robust global behaviour, where mathematical assurance exists that starting from an arbitrary initial point the algorithms will converge to a stationary point. This is achieved by adjusting the size of the steps taken at each iteration, based on how well the local quadratic models are predicting decrease in the value of the function to be minimised, after each iteration. Let us consider the nonlinear function:

$$y = f(x) \tag{3.51}$$

We would like to find the vector arguments $x \in R^n$ so that

$$min_{x \in R^n} f(x) \tag{3.52}$$

Suppose that we are at the k^{th} iteration at which the vector argument is $x(k)$ and we want to move $x(k)$ to $x(k+1)$ so that

$$f(x(k+1)) < f(x(k)) \tag{3.53}$$

The trust-region algorithm insures inequality condition Equation 3.53 as follows. The function $f(x)$ is approximated by the first two terms of its Taylor approximation at the point $x(k)$, resulting in the a linear quadratic function $g(x)$. This function is then evaluated in a region R called *the trust-region* around (in the neighbourhood of) point x. If a trial point, say point q, in this region is tested and found to minimise the function

$$min_{q \in R^n} f(q) \tag{3.54}$$

then $x(k)$ is updated to this point and $x(k+1) = x(k) + q$, and the process continues. Otherwise the region is shrunk and the trial continues until a new point is found or the region collapses to null (a stationary point, which is the minima, is reached).

In mathematical terms the trust-region optimisation problem becomes

$$min_{|D|<|q|} \{\frac{1}{2} q^T H q + \alpha g\} \tag{3.55}$$

where g is the gradient of $f(x)$ at the current point x, H is a symmetric matrix whose elements are the second derivative of $f(x)$, D is a scaling matrix, and $|.|$ is the Euclidean norm.

Although algorithms that solve the optimisation problem Equation 3.55 exist, the time required may be too prohibitive for large scale (dimensional) problems. This problem can be overcome if the trust region is restricted to a two-dimensional space Q. Once this space is found, the solution to the optimisation problem Equation 3.55 can be approximated and easily found. The common approach to finding Q is to define it as a linear

space spanned by vectors q_1 and q_2. Vector q_1 is in the direction of the gradient g, and vector q_2 is obtained from:

$$H.q_2 = -g \text{ (Approximate Newton direction)} \tag{3.56}$$

or

$$q_2^T.H.q_2 < 0 \text{ (Negative curvature direction)} \tag{3.57}$$

The linear space Q above will force global convergence via the steepest decent direction or negative curvature direction.

Summary of algorithm:

1. Formulate the 2-dimensional trust region problem.
2. Solve Equation 3.55 to determine the trial step q.
3. If Equation 3.53 is satisfied then $x = x + q$.
4. Otherwise adjust D.

Trust-Region Methods (Constrained Optimisation)

Constrained nonlinear optimisation problems involve finding the points of minima subject to the following constraints:

$$C_i(x) \leq 0 \;\; i = 1, 2, ..., m \tag{3.58}$$

The most efficient available optimisation method, which is implemented in the Matlab function *fmincon-active set*, is based on the solution of the Karush-Kuhn-Tucker (KKT) equations. These equations state the necessary (and sufficient for convex functions) conditions for optimality of nonlinear constrained optimisation problems, *i.e.*, existence of global solution point. Therefore the problem involves the solution to the following set equations and inequalities.

$$\nabla f(x) + \sum_{i=1}^{m} \lambda_i \nabla C_i(x) = 0 \;\; ; \;\; \lambda_i \nabla C_i(x) \leq 0 \;\; ; \;\; \lambda_i \geq 0 \tag{3.59}$$

where ∇ is the gradient and λ_i are the Lagrange multipliers. There are a number of programming methods that can solve the set of Equations 3.59, among the most efficient and proven algorithm is the Sequential Quadratic (SQ), which allows for the problem to be solved in the same manner as unconstrained optimisation. This achieved by formulating SQ sub-problems where the nonlinear constraints are linearised. The SQ sub-problem is stated as follows:

$$min_{d \in R^n} \{ \frac{1}{2} d^T H_k d + \nabla f(x_k)^T d \}$$
$$\nabla g_i(x_k)^T d + g_i(x_k) = 0 \;\; ; \;\; i = 1, 2, ..., m \tag{3.60}$$

where d is the line search direction.

This problem can be solved using any QP algorithm, where the solution forms the basis for a new iteration

$$x(k+1) = x(k) + \alpha_k d_k \tag{3.61}$$

Another method for solving the constrained optimisation problem Equation 3.55 is by formulating the problem to include a so called *barrier function*. This method is implemented in the Matlab function *fmincon-interior point*, which solves a sequence of approximate minimisation problems, formulated as follows. Given the minimisation problem (Equation 3.52) subject to the set of constraints (Equation 3.58) for each μ solve the following problem:

$$min_{x,q} f_\mu(x,q) = min_{x,q} fx - \mu \sum_i ln(q_i) \tag{3.62}$$

subject to

$$g(x) \leq 0 \;\; ; g(x) = -q \tag{3.63}$$

where $\mu \sum_i ln(q_i)$ is the barrier function. Inequality in Equation 3.63 implies that $q_i > 0$ and therefore $ln(q_i) > 0$. As μ decreases, $f_\mu(x,q)$ approaches minimum of $f(x)$ (see Equation 3.62). At each iteration the algorithm decreases the following function:

$$f_\mu(x,q) + \beta|g(x) + q| \tag{3.64}$$

The parameter β is adjusted at each iteration to force the solution to feasibility.

3.5.4 Experimental Results: Determining the Weights

In order to determine the weights, the three different approaches explained in Section 3.5.2 are implemented using the results of five recent border detection methods; namely, Hybrid (W30B30) [204], DTEA [26], JSEG [95], KPP [89], and Global thresholding on XoYoR colour channel [205], on the image set of 55 dermoscopy images. For each of the five border detection methods, the border is detected for every image, the segmented image is formed and the four parameters of TP, FP, FN and TN are calculated by comparing the segmented image with the corresponding gold standard (ground truth). The obtained results for each approach are explained in the following.

Image Set and Gold Standard:

The proposed evaluation metric is tested on the same set of 55 high-resolution dermoscopy images applied in the previous section, with four sets of manual borders independently drawn by expert dermatologists from Royal Melbourne Hospital. As before, the union of those borders is taken as the ground truth. The choice of union for integrating the manual borders, is in line with the emphasis on the TP factor, reflected in the inequality constraints as expressed by Equations 3.39 and 3.40. The image standardisation procedure discussed in Section 3.3.2 has been also applied to balance the effect of large TN values and normalise the accuracy and specificity metrics.

Optimisation of OWPI:

After calculating the four parameters of TP, FP, FN and TN for each of the five border detection methods, for every image, 20 trials with different initial conditions are conducted and the weights (α, δ, β and σ) resulting in the highest OWPI value were obtained through the constrained nonlinear multivariable optimisation algorithm. The initial conditions used in trials were in the range of [0.1,2] with steps of 0.1, equally given to the four weights, *i.e.*, starting from $\alpha = 0.1, \delta = 0.1, \beta = 0.1, \sigma = 0.1$ and ending with $\alpha = 2, \delta = 2, \beta = 2, \sigma = 2$.

Experimental results show that for all 20 trials, for all five methods, the optimal weights for TP, FP and FN are 2, 0.1 and 0.2, respectively. For TN's weight, however, different values were returned by the optimisation algorithm, implying that the cost function has multiple local maxima. In order to determine the optimal weight for TN, the frequency of each of the suggested weights in different trials over the entire image set has been calculated, as shown in Tables 3.15–3.19 for the Hybrid, JSEG, DTEA, Global and KPP methods, respectively. The tables show that for all border detection methods in the majority of cases (56%, 61%, 68%, 75%, and 75%) the optimal value for TN's weight is 1.9. Figure 3.37 shows value frequencies of TN's weight obtained by the optimisation procedure on the 20 trials run over the entire image set, for the five border detection methods. As illustrated in the figure, for all methods the most frequent weight returned by the optimisation procedure for the TN parameter is 1.9, demonstrating the consistency of the optimal weights for all methods.

Solving the System of Nonlinear Equations:

To reconfirm the optimal weights α, β, σ and δ, we have also solved the system of nonlinear equations defined by Equation 3.48, subject to the constraints given in Equation 3.39, for the Hybrid method. Solving these equations was done by performing a constrained nonlinear optimisation procedure that minimises the defined SSE (Equation 3.49).

For each image in the set of 55 high-resolution dermoscopy images, 20 trials with different initial conditions were conducted. Starting from initial conditions of ($\alpha = 0.1$, $\beta = 0.1$, $\sigma = 0.1$, $\delta = 0.1$) and ending with ($\alpha = 2$, $\beta = 2$, $\sigma = 2$, $\delta = 2$), with steps of 0.1, the constrained optimisation procedure was applied. Using the optimal weights resulting from the optimisation, the OPWI was calculated. To check the accuracy of the optimal solution, the resultant weights (α, β, σ, δ) and the corresponding TP, FP, FN, and TN values were substituted in the four derivatives defined by Equations 3.44–3.47. The outcome was compatible with that obtained by the solution obtained in first method explained above.

Similar to the results obtained in the previous section (optimisation of OWPI), consistent weights of $\alpha = 2$, $\beta = 0.1$, $\sigma = 0.2$ were observed. For δ, it was observed that when δ's bounds were narrowed down from [0.1,2] to [1.9,2], the resultant OWPI increased and the derivative values decreased or remained the same. Table 3.20 shows the obtained

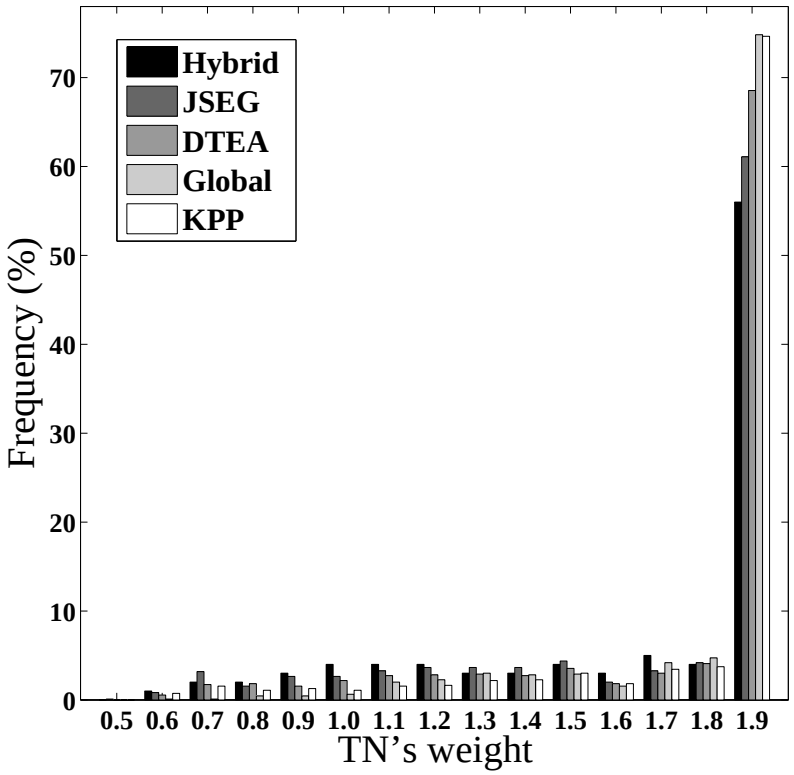

Figure 3.37: Frequencies of obtained values for TN's weight in 20 trials of optimisation over the entire image set, using the five border detection methods.

optimal weights and the corresponding OWPI values for a sample image, demonstrating that in the majority of cases the optimal value for TN's weight (δ) is 1.9. The same pattern was observed over the whole image set, consistent with the results obtained by optimisation of OWPI (the first approach of the previous section).

Figure 3.38 shows the increasing trend in OWPI value for the same image, as δ's bounds are narrowed down. Figure 3.39 shows the decreasing trend in the value of $\frac{\partial OWPI}{\partial TP}$ (Equation 3.44) for the same image, as δ's bounds are narrowed down from bound1=[0.1,2] to bound6=[1.9,2].

Grid Search:

A grid search, as explained in Section 3.5.2, is also applied on the Hybrid method to find the weights yielding to maximum OWPI. . Figure 3.40 shows the mean OWPI values over the image set, using all different combinations of α, β, σ and δ within the range of [0.1,2], subject to the constraints given in Equation 3.39. As the result of the grid search, the same optimal weights as those found by the optimisation processes, *i.e.*, ($\alpha = 2$, $\beta = 0.1$, $\sigma = 0.2$, $\delta = 1.9$).

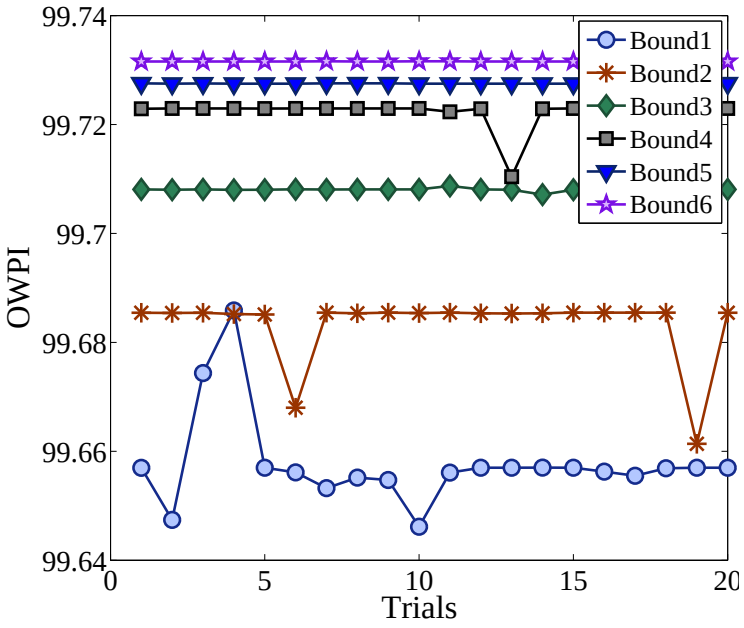

Figure 3.38: The increasing trend in OWPI value for a sample image, as δ's bounds are narrowed down from bound1=[0.1,2] to bound6=[1.9,2].

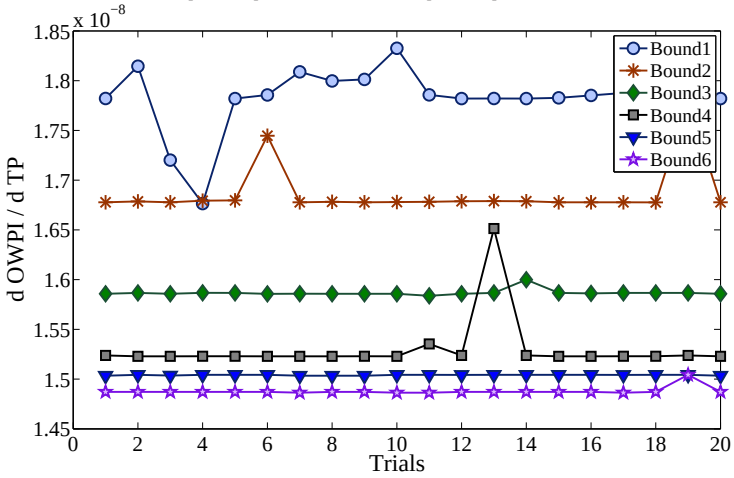

Figure 3.39: The decreasing trend in $\frac{\partial OWPI}{\partial TP}$ value for a sample image, as the δ's bound is narrowed down from bound1=[0.1,2] to bound6=[1.9,2].

Acceptability of the Obtained Weights Over the Entire Image Set:

Figure 3.41 shows the OWPI value using optimised weights of $\alpha = 2, \delta = 1.9, \beta = 0.1, \sigma = 0.2$, for our set of 55 dermoscopy images, using the Hybrid method. To make sure the applied weights are generic for different images we have considered 6 images as control samples and have produced their respective OWPI for different trials of the optimisation. The control images are those which produce the lowest OWPI. Thus, by calculating their

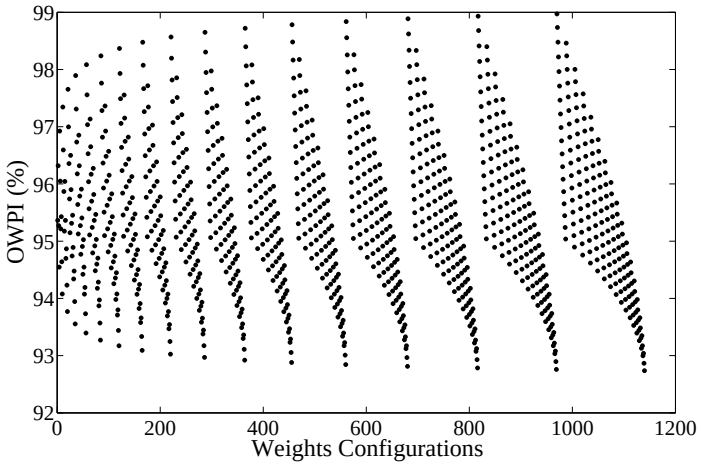

Figure 3.40: Mean OWPI values over the image set, using the constrained combinations of α, β, σ and δ within the range of [0.1,2].

OWI for different trials of optimisation we have shown that the obtained OWPI is still quite hight. These values are marked with O in Figure 3.41, demonstrating the small variability in the resultant OWPI values and the consistency of the obtained weights.

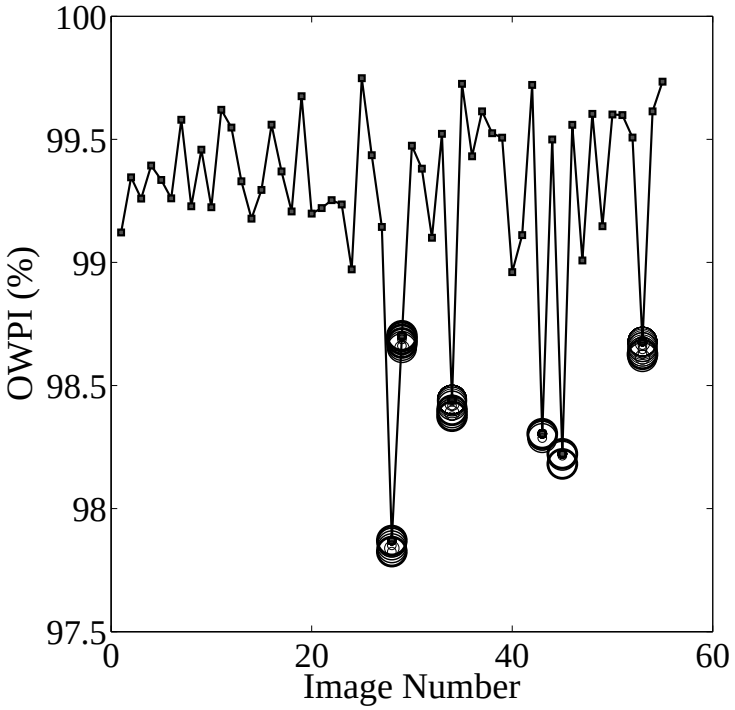

Figure 3.41: Optimised Weighted Performance Index for the set of 55 dermoscopy images using the Hybrid border detection method.

3.5.5 Applying OWPI for Comparison of Automated Methods

Figure 3.42 shows the five different borders resulting from applying the above-mentioned five border detection methods (DTEA [26], JSEG [95], KPP [89], global thresholding on XoYoR colour channel [205], and Hybrid [204]) on two sample dermoscopy images. Optimised Weighted Performance Index can be used as a mean to compare the effectiveness of these borders. The proposed Optimised Weighted Performance Index with the deter-

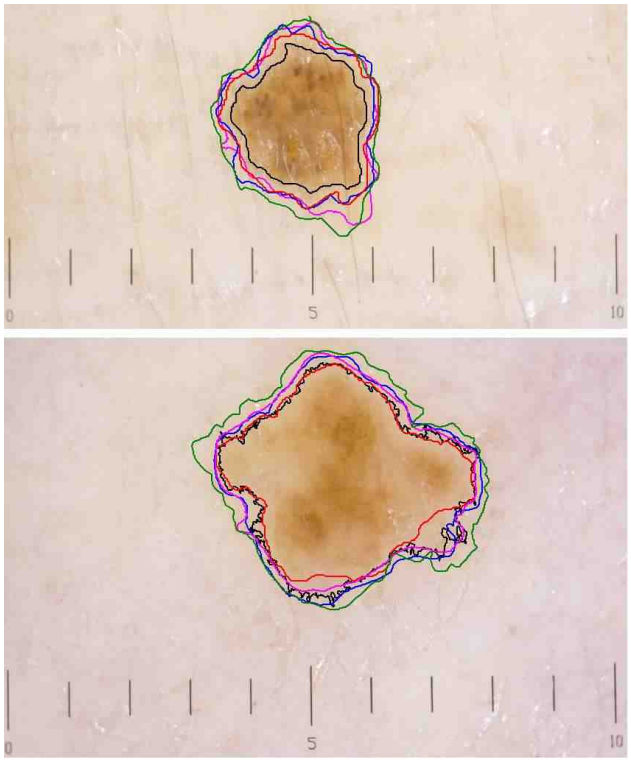

Figure 3.42: Different automatic borders for two sample dermoscopy images.

mined optimal weights is used to evaluate and compare the performance of the five border detection methods applied to the set of 55 high-resolution dermoscopy images.

Table 3.21 gives the mean and standard deviation of the metrics of weighted accuracy, weighted sensitivity, weighted specificity, weighted precision, weighted similarity, weighted border error and Optimised Weighted Performance Index. As shown, one method may perform better than others with respect to some of the metrics, yet yields lower results with respect to others, *e.g.* the DTEA method gains the highest weighted specificity and weighted precision, yet is bested by the Hybrid method with respect to weighted sensitivity, weighted accuracy, weighted similarity and weighted border error metrics. Therefore it has not been easy in the past to provide an overall objective judgment as to which of the five methods is more suited for border detection of dermoscopy images. However, the use of the proposed OWPI –since it incorporates all the metrics- facilitates such a judgment,

which can easily be made by comparing values of the calculated OWPI's (as shown in the last row of Table 3.21).

Table 3.21 demonstrates that the Hybrid (W30B30) method yields the highest performance with the average OWPI of 99.28%, followed by the JSEG method with average OWPI of 98.64%, the DTEA and Global methods with the average OWPI of 98.3%, and finally the KPP method with the average OWPI of 97.86%

Figure 3.43 shows different automatic borders drawn by the five border detection methods for a sample dermoscopy image and the ground truth drawn by dermatologists. As shown in the figure, the border produced by the Hybrid method is the closest to the ground truth.

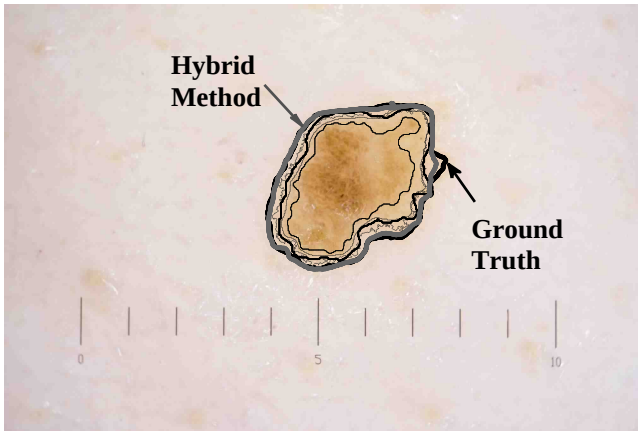

Figure 3.43: Different automatic borders for a sample dermoscopy image and the ground truth drawn by dermatologists. The border produced by the Hybrid method is the closest to the ground truth.

3.5.6 Extending the Optimisation Bounds

To study the effect of extending the bounds of the optimisation procedure, the range is extended to [0.05,2.1] and the process described in Section 3.5.2 is repeated for the all five border detection methods. Similar to the previous experiment with the initial bounds, the optimal weights for TP, FP and FN are constant for all 20 trials and for all five methods, as $\alpha = 2.1$, $\beta = 0.05$ and $\sigma = 0.15$.

Similarly, the optimisation procedure suggest various values for δ in the range of [0.5,2]. To determine the most consistent value for δ, the frequencies of various values obtained by the optimisation procedure on the 20 trials run over the entire image set are calculated for each of the five border detection methods, as shown in Figure 3.44. As illustrated in the figure, for all the methods the most frequent weight value chosen by the

optimisation procedure for the TN parameter is 2, which demonstrates the consistency of the optimal weights for all the five methods.

To confirm the new weights, the grid-search procedure explained in Section 3.5.2 is applied on the Hybrid method's results, using the extended bound of [0.05,1], which converged to the same weight values of $\alpha = 2.1$, $\beta = 0.05$, $\sigma = 0.15$ and $\delta = 2.0$ yielded by the optimisation procedure above.

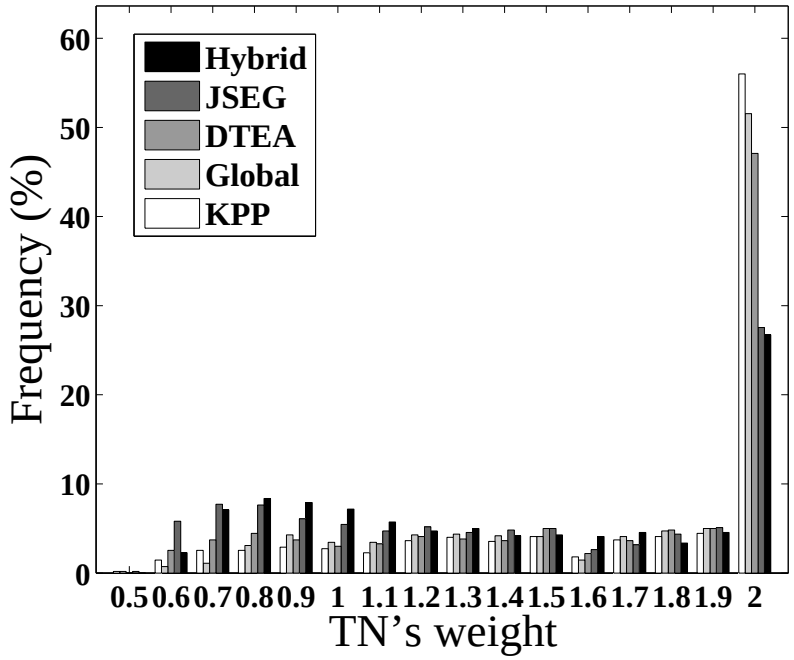

Figure 3.44: Frequencies of obtained values for δ in 20 trials of optimisation over the entire image set, for the five border detection methods using the extended bound of [0.05,2.1].

Figure 3.45 shows the average optimised weighted performance index over the image set, using $\alpha = 2.1$, $\beta = 0.05$, $\sigma = 0.15$ and various values of δ, for the five border detection methods. As illustrated in the figure, the trend is similar for all the five methods. Also, similar to the results observed in Table 3.21, the Hybrid border detection method provides the best performance, followed by JSEG, DTEA, Global and KPP methods, showing that although the obtained optimal weights for different bounds differs, the relative performance is stable which provides a sustained basis for the comparison of the border detection methods. Moreover, in extending the bounds it was observed that for the more relax bounds of [0.05,2.1] as shown in Figure 3.44, the distributions of weights flattens compare to the initial bounds of [0.1,2] as illustrated in Figure 3.37, thus the larger bound yields less consistency and sustainability in the obtained weights over the image set.

In this study, our optimisation process used constraints based on subjective concerns of dermatologists with respect to the application of melanoma diagnosis. The positive lower limit of 0.1 (or any positive number) reflects a desire to ensure nontrivial inclusion

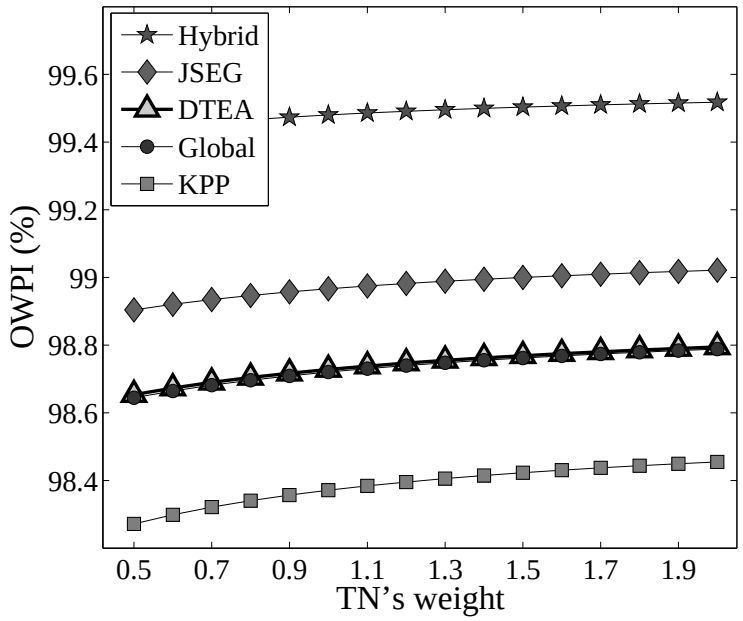

Figure 3.45: Average Optimised Weighted Performance Index over the image set, using $\alpha = 2.1$, $\beta = 0.05$, $\sigma = 0.15$ and various values of δ, for the five border detection methods.

of all four parameters FP, TP, FN, and TN in the metric. On the other hand, the upper limit of 2.0 (or any positive number), is a formal constraint needed to avoid unbounded weights, which does not in fact limit the form of the solution, since rescaling all weights by the same factor leaves the metric as given by Equation 3.38 unchanged.

3.5.7 Comparison between Fixed and Optimised Weights

Figure 3.46 shows the comparison of the three evaluation metrics of OWPI, WPI and PI for the five border detection methods, *i.e.*, Hybrid, JSEG, DTEA, Global and KPP, over the image set. Comparison between PI, WPI and OWPI shows the advantage of the latter. OWPI yields higher performance values for each of the five border detection methods, followed by WPI and PI comes last. Moreover, OWPI with optimal weights tailored to be meaningful in the melanoma application yields smaller differences between the border detection methods, implying the actual performance differences among the various methods are in fact smaller than those predicted by WPI and PI.

3.5.8 Conclusion

This section presented a new approach for evaluation of border detection methods in dermoscopy images. To provide evaluation metrics meaningful in the context of dermoscopy image analysis, we introduced specific weights into standard metrics of sensitivity, specificity, accuracy, border error, similarity and precision. The weights were determined by

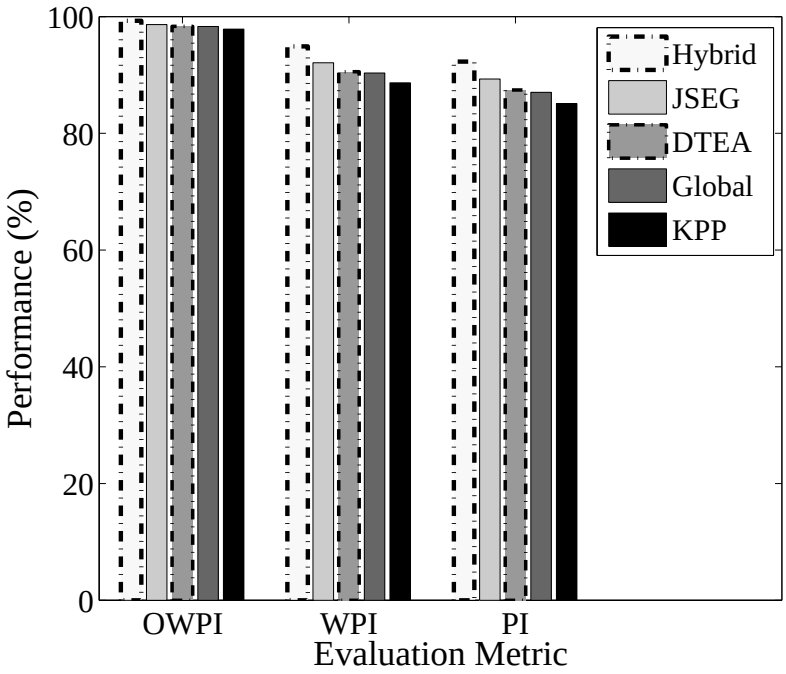

Figure 3.46: Comparison of the three evaluation indices of OWPI, WPI and PI for the five border detection methods: Hybrid, JSEG, DTEA, Global and KPP.

applying a constrained nonlinear multivariable optimisation method with trust-region approach, and the result was confirmed by a grid-search procedure. Moreover, a comprehensive metric, Optimised Weighted Performance Index, was proposed to facilitate comparison between different methods. The effectiveness of the proposed evaluation approach was demonstrated by applying five recent border detection methods on a set of 55 high-resolution dermoscopy images, using the union of four sets of dermatologist-drawn borders as the ground truth.

3.6 Concluding Remarks

The fist stage in development of a fully automated computer-based system for detecting melanoma is segmentation or border detection. This chapter presented two automated methods for detecting the borders in dermoscopy images; Global and Hybrid method, the latter being an enhancement of the former. The Global method applies colour space analysis and clustering-based histogram thresholding and determines X and XoYoR colour channels as the most effective and discriminative colour channels for detecting the borders in dermoscopy images, with an accuracy of approximately 97%. Quantitative evaluation is performed by comparing automated results to manual borders independently drawn by four dermatologists, with respect to four different metrics of accuracy, sensitivity, specificity and similarity, incorporating ROC analysis. Comparative study involving two state-

of-the-art automated methods demonstrated that with a proper choice of colour channels the proposed thresholding-based method, in spite of its simplicity, is highly competitive with well-known skin lesion segmentation methods, and outperforms them with respect to accuracy, specificity, and AUC metrics. Furthermore, the proposed method is potentially faster since it mainly involves scalar processing as opposed to vector processing performed in those methods. The Global border-detection method is published in [206, 205].

Considering the practical nature of the melanoma diagnosis, which calls for extreme caution when excluding portions of the image, and for the fact that dermatologists prefer to choose the outmost detectable pigment to minimise the risk of incorrect diagnosis, we improved the Global method with the goal of reducing the discrepancy between the manual and automatic borders by proposing the Hybrid border detection method. The method first applies the Global method on the optimal colour channel of XoYoR to detect an initial boundary, then integrates the local colour information around the initial boundary to expand the border and form the final segmentation result. 85 high-resolution dermoscopy images were analysed and the automatic borders were compared to borders drawn by two dermatologists using various evaluation metrics including accuracy, precision, sensitivity, specificity, and border error. The Hybrid method, which is comprised of two stages, is designed to increase specificity in the first stage (global thresholding) and sensitivity in the second stage (adaptive local thresholding). The proposed hybrid method was also compared to three state-of-the-art border detection methods and found to perform as well or better in terms of sensitivity, AUC, and border error. Furthermore, it was shown that the Hybrid method functions closer to the experienced dermatologist than the dermatology registrar. The Hybrid border-detection method is published in [204].

We then provided an analytical framework, including statistical analysis, optimisation and cross validation, to determine the optimal parameters for the proposed Hybrid border detection method. Moreover, the two proposed border detection methods; Global and Hybrid, were compared with each other and the superiority of the Hybrid method was demonstrated through a newly introduced evaluation metric, Performance Index, composed of standard metrics of sensitivity, specificity, accuracy, similarity, and border error. The comparison between the two methods, Global and Hybrid, is published in [207].

The last two sections of the chapter explore the objective evaluation of the border detection results in dermoscopy image analysis. The goal was to provide a comprehensive evaluation metric which reflects dermatologists' perspective in the melanoma recognition process. In doing so, specific weights are attached to the standard metrics of sensitivity, specificity, accuracy, border error, similarity and precision, and the Weighted Performance Index is defined as a weighted average of those. However, as the introduced weights were fixed and based on expert opinion only, and in order to provide weights which genuinely reflect the importance of each metric in the evaluation process we proposed a new metric; Optimised Weighted Performance Index. The Optimised Weighted Performance Index is also an average of the weighted metrics of sensitivity, specificity, accuracy, border error, similarity and precision, where the weights are obtained by formulating a constrained nonlinear multivariable optimisation problem. Comparison between Performance Index,

Weighted Performance Index and Optimised Weighted Performance Index shows the advantage of the latter. OWPI yields higher performance values for each of the five border detection methods, followed by WPI and PI comes last. Moreover, OWPI with optimal weights which are tailored to be meaningful in the melanoma application, yields smaller differences between the border detection methods, implying the actual performance differences among the various border detection methods are in fact smaller than those predicted by WPI and PI. Objective evaluation of border detection methods in dermoscopy images via Weighted Performance Index and Optimised Weighted Performance Index are published in Skin Research and Technology journal [208] and submitted to IEEE Transactions on Information Technology in Biomedicine [209], respectively.

Table 3.15: Frequencies of different suggested values for TN's weight (δ) in 20 trials of optimisation over the entire image set, using Hybrid border detection method.

δ	0.6	0.7	0.8	0.9	1.0	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9
Trial.1	0	0	1	3	5	5	3	1	0	3	3	1	1	29
Trial.2	4	2	0	0	1	1	0	0	0	0	0	0	0	47
Trial.3	5	2	0	1	2	0	0	0	0	0	0	0	0	45
Trial.4	6	3	0	0	0	0	0	0	0	0	0	0	0	46
Trial.5	0	10	0	0	0	0	0	0	0	0	0	0	0	45
Trial.6	0	8	4	0	0	0	0	0	0	0	0	0	0	43
Trial.7	0	0	15	0	0	0	0	0	0	0	0	0	0	40
Trial.8	0	0	0	19	1	1	0	0	0	0	0	0	0	34
Trial.9	0	0	0	0	28	2	1	0	0	0	0	0	0	24
Trial.10	0	0	0	0	0	36	5	1	0	1	0	0	0	12
Trial.11	0	0	0	0	0	0	37	4	2	0	0	1	0	11
Trial.12	0	0	0	0	0	0	1	37	6	2	0	1	0	8
Trial.13	0	0	0	0	0	0	0	0	29	21	0	1	0	4
Trial.14	0	0	0	0	0	0	0	0	0	17	28	4	1	5
Trial.15	0	0	0	0	0	0	0	0	0	0	4	38	5	8
Trial.16	0	0	0	0	0	0	0	0	0	0	0	1	31	23
Trial.17	0	0	0	0	0	0	0	0	0	0	1	3	2	49
Trial.18	0	0	0	0	0	0	0	0	0	0	0	0	0	55
Trial.19	0	0	0	0	0	0	0	0	0	0	0	0	5	50
Trial.20	0	0	0	5	10	1	1	0	0	0	0	0	0	38
Frequency	15	25	20	28	47	46	48	43	37	44	36	50	45	616
Percentage(%)	1	2	2	3	4	4	4	3	3	4	3	5	4	56

Table 3.16: Frequency of different suggested values for TN's weight (δ) in different trials of optimisation over the entire image set using JSEG border detection method.

δ	0.5	0.6	0.7	0.8	0.9	1.0	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9
Trial.1	0	0	3	2	3	1	0	1	0	0	2	0	1	0	42
Trial.2	0	4	1	0	0	1	0	1	0	0	0	0	0	0	48
Trial.3	1	2	1	0	0	0	0	0	0	0	0	0	0	0	51
Trial.4	0	2	8	0	0	0	0	0	0	0	0	0	0	0	45
Trial.5	0	1	11	0	0	0	0	0	0	0	0	0	0	0	43
Trial.6	0	0	11	1	0	0	0	0	0	0	0	0	0	0	43
Trial.7	0	0	0	12	0	0	0	0	0	0	0	0	0	0	43
Trial.8	0	0	0	1	15	0	1	0	0	0	0	0	0	0	38
Trial.9	0	0	0	0	1	25	0	1	0	0	0	0	0	0	28
Trial.10	0	0	0	0	0	1	34	1	0	1	0	0	0	0	18
Trial.11	0	0	0	0	0	0	0	36	1	0	1	0	0	0	17
Trial.12	0	0	0	0	0	0	0	0	38	1	0	1	0	0	15
Trial.13	0	0	0	0	0	0	0	0	0	38	4	0	1	0	12
Trial.14	0	0	0	0	0	0	0	0	0	0	41	2	2	0	10
Trial.15	0	0	0	0	0	0	0	0	0	0	0	19	28	1	7
Trial.16	0	0	0	0	0	0	0	0	0	0	0	0	3	41	11
Trial.17	0	0	0	0	0	0	0	0	0	0	0	0	1	2	52
Trial.18	0	0	0	0	0	0	0	0	0	0	0	0	0	1	54
Trial.19	0	0	0	0	0	0	0	0	0	0	0	0	0	1	54
Trial.20	0	0	0	1	10	1	1	0	1	0	0	0	0	0	41
Frequency	1	9	35	17	29	29	36	40	40	40	48	22	36	46	672
Percentage(%)	0.1	0.8	3.2	1.6	2.7	2.7	3.3	3.7	3.7	3.7	4.4	2	3.3	4.2	61.1

Table 3.17: Frequency of different suggested values for TN's weight (δ) in different trials of optimisation over the entire image set using DTEA border detection method.

δ	0.6	0.7	0.8	0.9	1.0	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9
Trial.1	0	0	3	0	2	1	1	0	0	0	0	0	0	48
Trial.2	1	3	1	0	0	0	0	0	0	0	0	0	0	50
Trial.3	2	3	0	0	0	0	0	0	0	0	0	0	0	50
Trial.4	3	3	0	0	0	0	0	0	0	0	0	0	0	49
Trial.5	0	6	0	0	0	0	0	0	0	0	0	0	0	49
Trial.6	0	4	3	0	0	0	0	0	0	0	0	0	0	48
Trial.7	0	0	11	0	0	0	0	0	0	0	0	0	0	44
Trial.8	0	0	1	11	0	0	0	0	0	0	0	0	0	43
Trial.9	0	0	0	1	16	0	0	0	0	0	0	0	0	38
Trial.10	0	0	0	0	1	28	0	1	0	0	0	0	0	25
Trial.11	0	0	0	0	0	1	29	1	0	0	0	0	0	24
Trial.12	0	0	0	0	0	0	1	30	0	0	0	0	0	24
Trial.13	0	0	0	0	0	0	0	0	30	5	0	0	0	20
Trial.14	0	0	0	0	0	0	0	0	0	34	5	0	0	16
Trial.15	0	0	0	0	0	0	0	0	0	0	15	31	2	7
Trial.16	0	0	0	0	0	0	0	0	0	0	0	2	39	14
Trial.17	0	0	0	0	0	0	0	0	0	0	0	0	2	53
Trial.18	0	0	0	0	0	0	0	0	0	0	0	0	1	54
Trial.19	0	0	0	0	0	0	0	0	0	0	0	0	1	54
Trial.20	0	0	1	5	5	0	0	0	0	0	0	0	0	44
Frequency	6	19	20	17	24	30	31	32	30	39	20	33	45	754
Percentage(%)	0.6	1.8	1.8	1.6	2.2	2.8	2.8	2.9	2.7	3.6	1.8	3	4	68.6

Table 3.18: Frequency of different suggested values for TN's weight (δ) in different trials of optimisation over the entire image set using Global border detection method.

δ	0.6	0.7	0.8	0.9	1.0	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9
Trial.1	0	0	1	0	0	0	1	0	0	0	0	0	0	53
Trial.2	0	0	0	0	0	0	0	0	0	0	0	0	0	55
Trial.3	0	0	0	0	0	0	0	0	0	0	0	0	0	55
Trial.4	0	0	0	0	0	0	0	0	0	0	0	0	0	55
Trial.5	1	0	0	0	0	0	0	0	0	0	0	0	0	54
Trial.6	0	1	0	0	0	0	0	0	0	0	0	0	0	54
Trial.7	0	0	2	0	0	0	0	0	0	0	0	0	0	53
Trial.8	0	0	1	3	0	0	0	0	0	0	0	0	0	51
Trial.9	0	0	0	1	6	0	0	0	0	0	0	0	0	48
Trial.10	0	0	0	0	1	21	0	0	0	0	1	0	0	32
Trial.11	0	0	0	0	0	1	23	1	0	0	0	0	0	30
Trial.12	0	0	0	0	0	0	1	32	0	0	0	0	0	22
Trial.13	0	0	0	0	0	0	0	0	31	2	0	0	0	22
Trial.14	0	0	0	0	0	0	0	0	0	30	8	0	0	17
Trial.15	0	0	0	0	0	0	0	0	0	0	7	43	0	5
Trial.16	0	0	0	0	0	0	0	0	0	0	0	2	47	6
Trial.17	0	0	0	0	0	0	0	0	0	0	1	1	2	51
Trial.18	0	0	0	0	0	0	0	0	0	0	0	0	1	54
Trial.19	0	0	0	0	0	0	0	0	0	0	0	0	2	53
Trial.20	0	0	1	1	0	0	0	0	0	0	0	0	0	53
Frequency	1	1	5	5	7	22	25	33	31	32	17	46	52	823
Percentage(%)	0.1	0.1	0.5	0.5	0.6	2	2.3	3	2.8	2.9	1.6	4.2	4.7	74.8

Table 3.19: Frequency of different suggested values for TN's weight (δ) in different trials of optimisation over the entire image set using KPP border detection method.

δ	0.6	0.7	0.8	0.9	1.0	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9
Trial.1	0	1	3	1	0	1	0	1	0	1	1	1	0	45
Trial.2	2	1	0	0	0	0	0	0	0	0	0	0	0	52
Trial.3	2	1	0	0	0	0	0	0	0	0	0	0	0	52
Trial.4	2	2	0	0	0	0	0	0	0	0	0	0	0	51
Trial.5	2	5	0	0	0	0	0	0	0	0	0	0	0	48
Trial.6	0	7	0	0	0	0	0	0	0	0	0	0	0	48
Trial.7	0	0	7	0	0	0	0	0	0	0	0	0	0	48
Trial.8	0	0	1	7	0	0	0	0	0	0	0	0	0	47
Trial.9	0	0	0	1	9	0	1	0	0	0	0	0	0	44
Trial.10	0	0	0	0	1	14	1	1	0	0	0	0	0	38
Trial.11	0	0	0	0	0	2	15	2	1	0	0	0	0	35
Trial.12	0	0	0	0	0	0	1	20	4	1	0	1	0	28
Trial.13	0	0	0	0	0	0	0	0	20	5	1	0	0	29
Trial.14	0	0	0	0	0	0	0	0	0	26	10	1	0	18
Trial.15	0	0	0	0	0	0	0	0	0	0	8	34	2	11
Trial.16	0	0	0	0	0	0	0	0	0	0	0	1	36	18
Trial.17	0	0	0	0	0	0	0	0	0	0	0	0	1	54
Trial.18	0	0	0	0	0	0	0	0	0	0	0	0	1	54
Trial.19	0	0	0	0	0	0	0	0	0	0	0	0	1	54
Trial.20	0	0	1	5	2	0	0	0	0	0	0	0	0	47
Frequency	8	17	12	14	12	17	18	24	25	33	20	38	41	821
Percentage(%)	0.7	1.6	1.1	1.3	1.1	1.6	1.6	2.2	2.3	3	1.8	3.4	3.7	74.6

Table 3.20: Optimal weights obtained in different trials and the corresponding OWPI value for a sample image.

Trial	α	β	σ	δ	OWPI
1	1.99	0.1	0.2	1.9	99.73
2	1.99	0.1	0.2	1.9	99.73
3	1.99	0.1	0.2	1.9	99.73
4	1.99	0.1	0.2	1.9	99.73
5	1.99	0.1	0.2	1.9	99.73
6	1.99	0.1	0.2	1.9	99.73
7	1.99	0.1	0.2	1.9	99.73
8	1.99	0.1	0.2	1.9	99.73
9	1.99	0.1	0.2	1.9	99.73
10	1.99	0.1	0.2	1.9	99.73
11	1.99	0.1	0.2	1.9	99.73
12	1.99	0.1	0.2	1.9	99.73
13	1.99	0.1	0.2	1.9	99.73
14	1.99	0.1	0.2	1.9	99.73
15	1.99	0.1	0.2	1.9	99.73
16	1.99	0.1	0.2	1.9	99.73
17	1.99	0.1	0.2	1.9	99.73
18	1.99	0.1	0.2	1.9	99.73
19	1.99	0.1	0.2	1.8	99.72
20	1.99	0.1	0.2	1.9	99.73

Table 3.21: Evaluation metrics (mean ±standard deviation) for five recent border detection methods.

	Hybrid	JSEG	DTEA	Global	KPP
WAccuracy	99.54 ±0.2	99.20 ±0.4	99.04 ±0.5	99.02 ±0.3	98.87 ±0.6
WSensitivity	98.77 ±0.9	97.11 ±2.2	96.42 ±2.2	96.42 ±2.0	95.45 ±3.7
WSpecifcity	99.87 ±0.1	99.98 ±0.1	99.99 ±0.0	99.98 ±0.0	99.95 ±0.1
WPrecision	99.74 ±0.2	99.96 ±0.1	99.98 ±0.0	99.96 ±0.1	99.92 ±0.2
WSimilarity	99.25 ±0.4	98.50 ±1.1	98.15 ±1.1	98.14 ±1.0	97.59 ±2.0
WBorderError	1.49 ±0.8	2.93 ±2.2	3.61 ±2.2	3.63 ±2.0	4.63 ±3.6
OWPI	99.28 ±0.3	98.64 ±1.0	98.33 ±1.0	98.31 ±0.9	97.86 ±1.6

Chapter 4

Feature Extraction, Selection and Classification

As mentioned in the first chapter, due to the enhancements in skin imaging technology and image processing techniques in the recent years, there has been a significant increase in interest in the computer-aided diagnosis of melanoma, aiming to remove subjectivity and uncertainty from the diagnostic process and provide a reliable second opinion to dermatologists. However, it is widely acknowledged that much higher accuracy is required for computer-based algorithms to be adopted routinely in the diagnostic process [24, 25].

A computer-aided diagnosis of melanoma generally comprises several components; image acquisition, segmentation or border detection (which was discussed in the previous chapter), feature extraction, feature selection, and classification; the latter three are the main focus of this chapter.

Feature extraction is used to extract the features; similar to those visually detected by dermatologists, that accurately characterises a melanoma lesion. The feature extraction methodology of many computerised melanoma detection systems has been largely based on the conventional clinical algorithm of ABCD-rule of dermoscopy due to its effectiveness and simplicity of implementation. Its effectiveness stems from the fact that it incorporates the main features of a melanoma lesion such as asymmetry, border irregularity, colour and diameter (or differential structures), where quantitative measures can be computed. Different image processing techniques have been used to extract such features, see for example [12, 60].

In this study we propose a new, comprehensive and highly effective feature extraction method which combines different types of features; novel ones and a few adopted from existing studies. In the proposed method texture and border (including geometry and time-series) features are extracted and combined.

The texture-based feature extraction method employs tree-structured wavelet decomposition. The previous studies [171, 169, 210] have used three-level decomposition

without further decomposition. In this study, we have elected to use a four-level wavelet tree in order to assess the richness of information content in another decomposition level. Moreover, we have incorporated colour information from different colour channels of Red, Green, Blue and Luminance into the texture analysis, as opposed to only luminance in the existing studies. This is done in order to investigate the textural information embedded in each colour channel independently as we suspect considering the Luminance channel only may disregard some of the content. We have also applied various statistical measurements and ratios on all the nodes of the wavelet-tree.

The border-based feature extraction method uses a variety of statistical measures. This is achieved via two distinctive processes: (i) various geometry-based features (standard and newly designed based on expert opinion of dermatologists) are defined to evaluate the properties of the overall shape of the lesion, and (ii) a time series model of the border is constructed and analysed in both spatial and frequency domains to extract main characteristics of the lesion boundary. To the best of our knowledge, the time-series analysis of process (ii) and some of the measures used in process (i) have not been previously applied on dermoscopy images.

Feature selection is an intermediate process that lies between the two steps of feature extraction and classification. In this process irrelevant, redundant, or noisy features are excluded from the initial feature vector constructed in the feature extraction step. Some of the prominent feature selection methods that have been used in existing computer-based melanoma recognition systems include sequential forward floating selection and sequential backward floating selection [58], genetic algorithms [211], ReliefF, mutual information and correlation-based methods used in [60], principal component analysis used in [169], statistical analysis used in [171, 210], etc. However, in this study the Gain Ratio-based Feature Selection (GRFS) method is employed for the following reasons: (i) the method is known to be computationally efficient, which is a vital factor because of the high dimensionality of the initial feature vector constructed in the feature extraction phase; (ii) It outperforms a sample of well-known filter-based feature selectors such as Information Gain, Chi-squared and ReliefF [212], as shown in this chapter. Due to computational efficiency, filter-based feature selection methods have been chosen over the wrapper-based models, and for this reason wrapper-based methods have not been considered in this study. With the use of GRFS, a 12% increment in the classification accuracy is obtained compared to using raw data (the whole of the constructed feature vector), as well as a considerable reduction in feature space dimension and computation time. To the best of our knowledge this method has not been applied in previous melanoma studies.

Moreover, in our initial investigations, we have integrated the GRFS method with a Correlation-based Feature Selection method (CFS) [213] and have proposed a two-stage feature selection scheme. The proposed scheme uses the GRFS method at the first stage, then applies the CFS method on the subset obtained from the first stage, at the second stage. However, experimental results (as will be discussed in detail in Experiment 1, Section 4.6.1 of this chapter) show that adding the CFS method does not provide a noticeable improvement to the classification performance over that obtained by using the

GRFS method only. Hence, the GRFS method was shown to be sufficient for the optimal and efficient feature selection. For this reason we have elected to use GRFS only as the feature selection method, for the rest of the experiments (Experiments 2 and 3).

Classification is the final step in the diagnosis process, where the extracted features are utilised to ascertain whether the lesion is cancerous or normal. The most common classification methods that have been applied in computer-based melanoma recognition systems [12] include Discriminate Analysis [171, 81, 113, 139, 173, 174, 175], Artificial Neural Networks [169, 59, 94, 113, 139, 177], K-nearest Neighbourhood [58, 59, 175], Support Vector Machine [12, 59, 60, 65, 113, 148, 181, 182], Decision Trees [12, 59, 81, 184, 185], Bayes Networks [12] and Logistic Regression [59, 172]. However some enhanced versions of these classifiers have been reported in literature. These include Random Forest (RF) [188], Logistic Model Tree (LMT) [189], and Hidden Naive Bayes (HNB) [1], which have been utilised in melanoma diagnosis for the first time, in this study. We have also applied the Support Vector Machine (SVM) classifier which turns out to be a very popular classifier in existing melanoma diagnosis applications according to the literature [12].

To evaluate the accuracy of our diagnosis system, which is comprised of the three phases of feature extraction, feature selection and classification, we have conducted a set of four experiments; a flowchart of which is shown in Figures 4.8 and 4.9. As shown in Experiment 3 (Section 4.6.5), the system is able to achieve the remarkable high result of 93.21% accuracy, 90.90% sensitivity and 94.92% specificity and Area Under Curve (AUC) of the Receiver Operating Characteristic (ROC) of 0.946. We have shown that the proposed system as presented in Experiment 3 is among the best and in some cases superior to other melanoma diagnosis systems that have been reported so far in the literature [12]. The proposed system encompasses most of the existing features proposed by other studies, blends them with novel ones carefully extracted to best reflect the property of the lesion, and outputs an optimal set of features which can accurately characterise any given lesion. In summary, the main contributions of this chapter are summarised as follows:

1. The use of four level of wavelet decomposition, incorporating four colour channels, as well as applying various statistical measurements and ratios on all the nodes of the fully extended wavelet tree for texture-based feature extraction.

2. The use of time-series analysis in spatial and frequency domains and new geometrical measures for border feature extraction.

3. Establishing that the Gain Ratio feature selection method can deliver substantial improvement in the accuracy of the system.

4. Combining different types of features in an optimal way to achieve higher performance for the diagnostic system.

The rest of the chapter is organised as follows. An overview of previous wavelet-based texture analysis methods in dermoscopy images and the proposed texture-based feature extraction method are provided in Section 5.1. The proposed geometry-based features

and time series-based features are detailed in Sections 5.2 and 5.3, respectively. Section 5.4 details the proposed feature selection method. The applied classification algorithms are discussed in Section 5.5. Experimental results and discussions are presented in Section 5.6. Section 5.7 provides a summary and concluding remarks.

4.1 Wavelet-based Texture Analysis in Dermoscopy Images

In clinical diagnostic approaches (*e.g.* ABCD rule of dermoscopy and pattern analysis) dermatologists look into the visual differences within the lesion and also changes in the appearance of the lesion over the time. These visual characteristics can be captured through texture analysis. Wavelet-based texture analysis provides a multiresolution analytical platform which enables us to characterise a signal (an image) in multiple spatial/frequency spaces. The multi-scale characteristics of wavelets can be very useful since dermoscopy images are taken under different circumstances such as various image acquisition set-ups (lighting, optical zooming, etc) and various skin colours.

The 2D wavelet transform has been widely applied in image processing-based biomedical applications. Mellat discussed the application of wavelets in image coding, texture discrimination and fractal analysis [214]. There exists two wavelet structures: (1) Recursive or pyramid-structured wavelet transform [214] which decomposes a signal into a set of frequency channels with narrower bandwidths in lower frequency channels, useful for signals whose important information lies in low frequency components, (2) Tree-structured wavelet analysis [215] which provides low, middle and high frequency decompositions which is done by decomposing both approximate (low-frequency) and detail (high-frequency) coefficients as shown in Figure 4.1. In dermoscopy image analysis, the lower frequency components reveal information about the general properties (shape) of the lesion, which is clinically important, and the higher frequency decomposition provides information about the textural detail and internal patterns of the lesion which is also significant in the diagnosis. Thus, the decomposition of all frequency channels are useful in this application, and the tree-structured wavelet analysis can be more informative for classification of skin lesions.

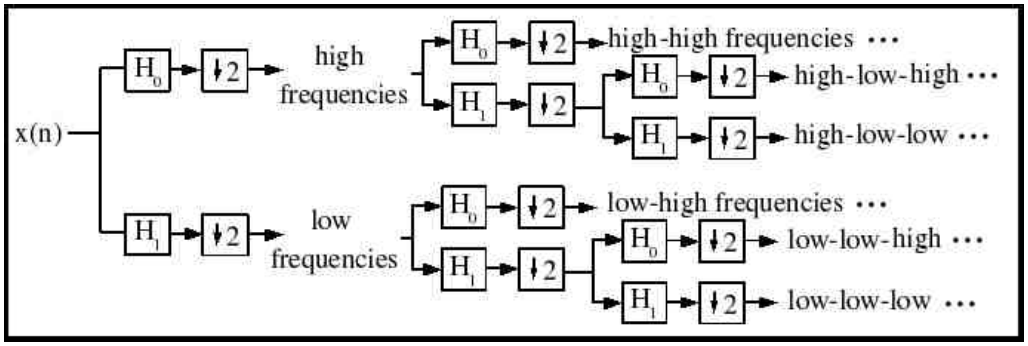

Figure 4.1: Tree-structured wavelet decomposition.

Some applications of wavelet transform to melanoma classification have been reported in the literature. Nimunkar *et al.* [171] applied the pyramid-structured wavelet transform in order to differentiate melanoma from dysplastic nevi on a set of 28 images. The luminance colour channel was decomposed into three levels and various ratios of the energy, entropy, mean and variance of the wavelet coefficients, plus the mean and standard deviation of the image intensity were measured to build up a vector of 34 features. Using a statistical feature selection approach (unpaired t-test for normally distributed features and Wilcoxon rank-sum test for non-normally distributed features), the feature vector was ultimately narrowed down to five optimal features. The classification was done by calculating the Mahanalobis distance between each feature for each image and the mean of the melanoma data (which is the mean value of features of images that fall within the melanoma class) and plotting the 2D scatter plot of the Mahanalobis distance values for each pair of features. This led to calculate ten discrimination functions and generate the respective ROC curves. Seven out of ten of the discrimination functions provided a true positive of 93% with false positives between 0–21%.

The drawback observed in their analysis is having multiple discrimination functions, each based on only two features, and lack of a discrimination function which takes into account all the features simultaneously. Moreover, despite the interesting approach in calculating various ratios of wavelet coefficients (which measures the spatial/frequency differences within the lesion) proposed in the study, we believe that including only the lower frequency components leads to missing significant diagnostic information which are stored in the higher frequency channels.

Inspired by Chang and Ku's method [215] with the tree-structured wavelet transform for image classification, Patwardhan *et al.* [210] proposed an adaptive wavelet-based tree-structured method to classify skin images. While Chang and Ku [215] utilised a fixed threshold for the average energy ratio as the stopping criterion for the tree decomposition, Patwardhan *et al.* [210] applied the three ratios of average energy, maximum energy and fractional energy ratios of the luminance image, to build up an initial feature vector of 231 features (three-level full tree decomposition). They, then applied a statistical analysis where the mean, variance, histogram of the feature values for each of the target classes (melanoma and dysplastic nevus) were used to select those features with bimodal distribution. For the features with bimodal distribution (using population mode), energy ratio thresholds were obtained for maximum separation between the two classes. These thresholds were then used to generate wavelet-based tree-structured models, or so-called signatures, for melanoma and dysplastic nevus. In the classification phase, images from a test set were semantically compared to the two built models of the two classes, resulting in a sensitivity of 90% and specificity of 90%. The training phase used 15 melanoma and 15 dysplastic nevus images, and test sets included 10 melanoma and 20 dysplastic cases which we believe may not be statistically sufficient for such a model-based approach.

The two feature sets proposed in [171] and [210] were later combined by Walvick *et al.* [169] and optimised using principal component analysis (over 33 images) to obtain an optimal set of four features. The classification was performed on a set of 27 images using a

back-propagation neural network, leading to a true positive rate of 100% and false positive rate of 23.53% which was shown to be higher than each individual feature set reported in [171] and [210]. We believe that the wavelet-based features proposed by Nimunkar *et al.* [171] and Patwardhan *et al.* [210] can be combined in a more effective way than what proposed by Walvick *et al.* [169]. As outlined in the following section, benefiting from ideas of the both studies [171, 210], we incorporate various colour channels, statistical measures and ratios to introduce a novel wavelet-based feature extraction approach that yields higher classification accuracy than those reported in [171], [210] and [169].

4.1.1 Proposed Texture-based Features

In this study, we propose a new, comprehensive and highly effective feature extraction method which combines different types of features; novel ones and a few adopted from existing studies [171, 210, 169]. The texture-based feature extraction method employs tree-structured wavelet decomposition. The novelty here lies in (i) the use of four-level decompositions as opposed to three level used in the existing wavelet-based methods [171, 169, 210], (ii) the use of four colour channels of red, green, blue and luminance as opposed to only luminance in existing studies, (iii) Applying various statistical measurements and ratios on all the nodes of the wavelet-tree. It is worth mentioning as far as we can ascertain no method has previously combined all three together as we have done in this study.

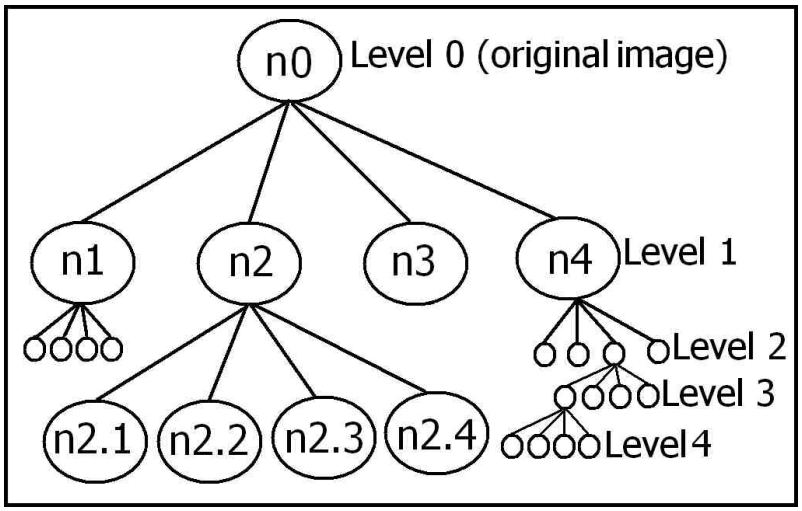

Figure 4.2: Schematic 4-level wavelet tree with nodes in circle.

The measures include energy (E), mean (M), standard deviation (Std), entropy (H), average-energy (AvgE), skewness (S), kurtosis (K), and norm (N), expressed in Equations 4.1–4.8. Figure 4.2 displays a schematic illustration of a wavelet tree with nodes marked by circles. For each colour channel, the measures are applied on the original image (called level 0) and also on wavelet coefficients of each node (sub-image) of the wavelet

tree. The wavelet tree has 340 nodes in total (4 nodes in the first level, 16 nodes in the second level, 64 nodes in the third level and 256 nodes in the fourth level). The eight measures therefore yield a total of 8×341 features, per colour channel.

$$E(n_i) = \frac{\sum_{j=1}^{J}\sum_{k=1}^{K} x_{jk}^2}{J \times K} \tag{4.1}$$

$$M(n_i) = \frac{\sum_{j=1}^{J}\sum_{k=1}^{K} x_{jk}}{J \times K} \tag{4.2}$$

$$Std(n_i) = \sqrt{\frac{\sum_{j=1}^{J}\sum_{k=1}^{K}(x_{jk} - M(n_i))^2}{J \times K}} \tag{4.3}$$

$$S(n_i) = \frac{\sum_{j=1}^{J}\sum_{k=1}^{K}(\frac{x_{jk}-M(n_i)}{Std(n_i)})^3}{J \times K} \tag{4.4}$$

$$K(n_i) = \frac{\sum_{j=1}^{J}\sum_{k=1}^{K}(\frac{x_{jk}-M(n_i)}{Std(n_i)})^4}{J \times K} \tag{4.5}$$

$$N(n_i) = max(\sqrt{eig(X_i \times X_i')}) \tag{4.6}$$

$$H(n_i) = \frac{\sum_{j=1}^{J}\sum_{k=1}^{K}\left(x_{jk}^2 \times log(x_{jk}^2)\right)}{J \times K} \tag{4.7}$$

$$AvgE(n_i) = \frac{\sum_{j=1}^{J}\sum_{k=1}^{K}|x_{jk}|}{J \times K} \tag{4.8}$$

where $i = 0, 1, ..., 340$ is an index of wavelet tree nodes starting sequentially from node n_0 of colour channel 1 and ending at node $n_{4.4.4.4}$ of colour channel 4 (colour channels: Red = 1, Green = 2, Blue = 3, Luminance = 4). X_i is a $J_i \times K_i$ matrix of the i-th node, X_i' is its transpose, x_{jk} is its jk-th element, and $eig(X_i)$ are its eigenvalues.

The ratios used in this study are maximum ratio (R_m), fractional ratio (R_f) and sub-level ratio (R_s), defined in Equations 4.9–4.11, respectively for a sample node $n_{2.3}$. As each decomposition sub-tree has four nodes (as illustrated in Figure 4.2), the maximum ratio divides each node by the maximum node among the four. The fractional ratio divides each node by the summation of the other three nodes. The sub-level ratio divides each node of the tree by its parent node. These ratios are calculated for each of the above-mentioned 8 measures, which will produce a total of $6120 = 8 \times 765$ (340 maximum ratios, 340 fractional ratios, 85 sub-level ratios) features.

$$R_m(n_{2.3}) = \frac{m(n_{2.3})}{max(m(n_{2.1}), m(n_{2.2}), m(n_{2.3}), m(n_{2.4}))} \tag{4.9}$$

$$R_f(n_{2.3}) = \frac{m(n_{2.3})}{m(n_{2.1}) + m(n_{2.2}) + m(n_{2.3}) + m(n_{2.4})} \tag{4.10}$$

$$R_s(n_{2.3}) = \frac{m(n_{2.3})}{m(n_2)} \tag{4.11}$$

where $n_{2.3}$ refers to node 2.3 and $m(n_{2.1})$ refers to the measure (energy, entropy, etc.) applied on the node.

Moreover, the feature *ln(std+1)* suggested by [171] and [169] has also been used. All of these features were calculated for four colour channels of red, green, blue and luminance (Equation 4.12) and a set of 35396 features is obtained.

$$luminance = (0.29 \times R) + (0.59 \times G) + (0.11 \times B) \tag{4.12}$$

where R, G and B represent red, green and blue colour channels, respectively.

As outlined in the introduction this is the first time four colour channels with four levels of wavelet decomposition have been used in melanoma application.

The features proposed in this section also account for colour information of the lesion (in addition to the textural properties). The colour features, which include the eight measures of energy, mean, standard deviation, entropy, average-energy, skewness, kurtosis, and norm of the four colour channels red, green, blue and luminance of the image, are measured by employing these eight measures on the node-0 (parent node) of the wavelet tree, *i.e.*, the original image.

4.2 Geometry-based Features in Dermoscopy Images

Border formation and geometric properties of the shape of the lesion provide significant diagnostic information in the detection of melanoma. The main prerequisite for extracting border features is the segmentation or border detection step, where the lesion is separated from the surrounding normal skin. The output of the segmentation step is the segmentation plane, which is a black-white image. In this study, after forming the lesion pixels into a 2D matrix and the corresponding boundary pixels into a vector, the following set of 11 geometry-based features are extracted from each dermoscopy image:

Area *(A)*: Number of pixels of the lesion.

Perimeter *(P)*: Number of pixels along the detected boundary.

Greatest Diameter *(GD)*: The length of the line which connects the two farthest boundary points and passes across the lesion centroid (C), which is given by:

$$(x_c, y_c) = \left(\frac{\sum_{i=1}^{n} x_i}{n}, \frac{\sum_{i=1}^{n} y_i}{n} \right) \tag{4.13}$$

where n is the number of pixels inside the lesion, and (x_i, y_i) are the coordinates of the i-th lesion pixel.

Shortest Diameter *(SD)*: The length of the line connecting the two closest boundary points and passes across the lesion centroid.

Circularity Index *(CRC)*: It explains the shape uniformity.

$$CRC = \frac{4A\pi}{P^2} \tag{4.14}$$

Irregularity Index A *(IrA)*:

$$IrA = \frac{P}{A} \tag{4.15}$$

Irregularity Index B *(IrB)*:

$$IrB = \frac{P}{GD} \tag{4.16}$$

Irregularity Index C *(IrC)*:

$$IrC = P \times (\frac{1}{\mathrm{SD}} - \frac{1}{\mathrm{GD}}) \tag{4.17}$$

Irregularity Index D *(IrD)*:

$$IrD = GD - SD \tag{4.18}$$

Major and Minor Asymmetry Indices: These indices are defined as the area difference between the two halves of the lesion, taken the principal axes (obtained by Equation 4.19) as the major symmetry axis, and its 90 degree rotation as the minor axes of the symmetry.

$$tan\ \ 2\theta = \frac{2\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} y_i^2} \tag{4.19}$$

where θ is the orientation of the principal axis. Figure 4.3 shows a sample dermoscopy image with obtained principal axes (major and minor axes of symmetry). After calculating the major and minor symmetry axes, the lesion is folded along the axes and the differences between the two halves of the lesion are calculated by applying the XOR operation on the binary segmentation plane. The asymmetry index is measured by:

$$\text{Asymmetry Index} = \frac{A_D}{A} \times 100 \tag{4.20}$$

where A_D denotes the difference between the two halves. Figure 4.4 shows a dermoscopy image with the obtained major principal axes (symmetry axes), and the process of calculating the major asymmetry index.

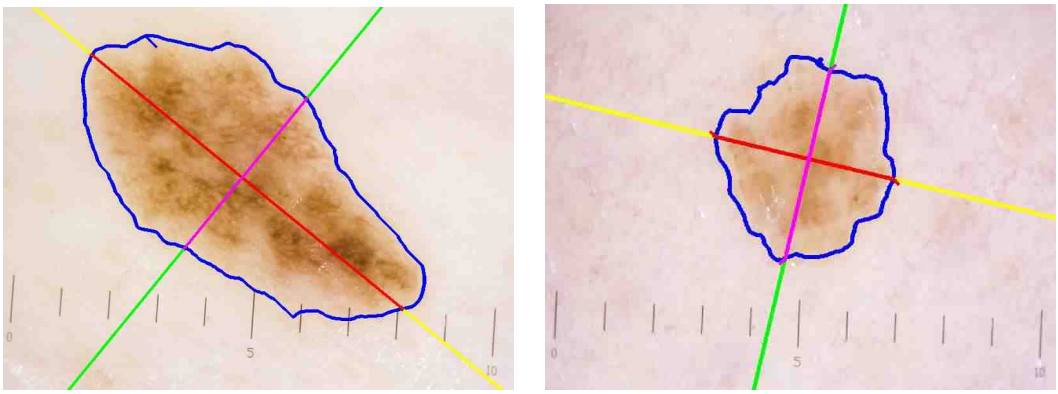

Figure 4.3: Major and minor symmetry axes for sample dermoscopy images.

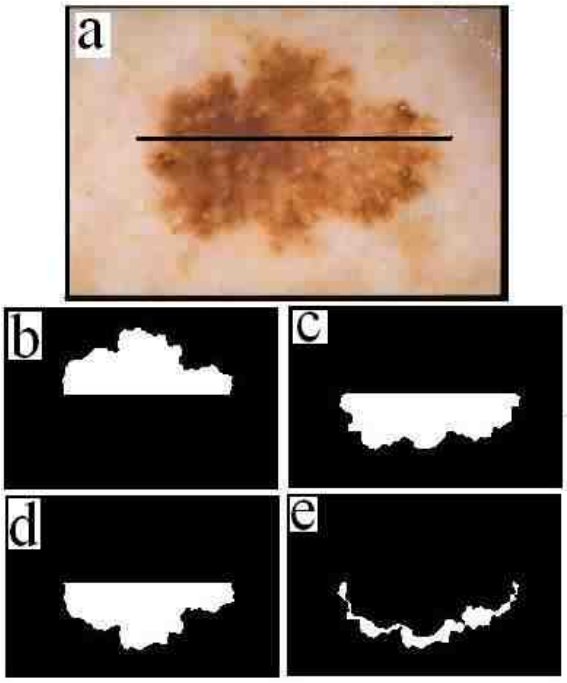

Figure 4.4: Calculating the major symmetry index: (a) major symmetry axis, (b) upper half, (c) lower half, (d) folded upper half, (e) difference.

4.3 Time-Series Analysis in Dermoscopy Images

A time-series is an ordered sequence of values or observations x that are measured and recorded at successive equally spaced time intervals t [216]. Time-series has been widely applied in statistics, signal processing, data mining, mathematical finance and economy. To the best of our knowledge they have not been previously applied to the application of melanoma recognition. In this study, we propose to build the time series by traversing

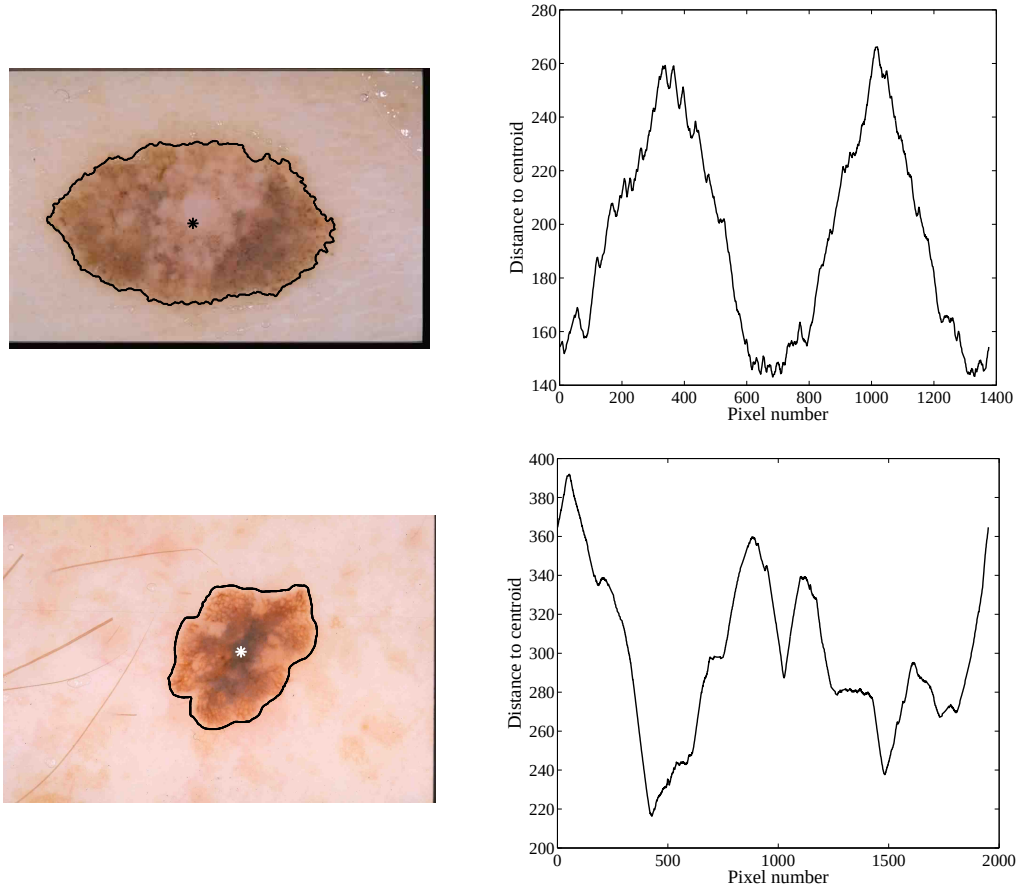

Figure 4.5: Time-series for sample dermoscopy images.

the boundary of the lesion (starting from an arbitrary pixel on the lesion boundary[1]), and calculating the distance between each border pixel and the centroid of the lesion, as expressed by Equation 4.21.

$$TS = d_1, d_2, ..., d_P \tag{4.21}$$

where P is the number of pixels along the boundary (previously defined as perimeter of the lesion) and

$$d_j = \sqrt{(x_j - x_c)^2 + (y_j - y_c)^2} \tag{4.22}$$

where (x_j, y_j) and (x_c, y_c) are the coordinates of the j-th boundary pixel and lesion centroid, respectively. Figure 4.5 shows two dermoscopy images and the corresponding time-series curve.

[1]The choice of the starting point is not our concern because the measures applied on the resultant time-series, including histogram and wavelet analysis, are rotation invariant.

4.3.1 Proposed Time-series Features

The obtained time series is analysed in both spatial and frequency domains. The former is done by calculating the histogram of the time series (using 10 bins), and the latter is performed via applying pyramid-structured wavelet decomposition, where the signal (the time series) is decomposed into an approximate (low-frequency) and a detail (high-frequency) component, and as in the pyramid-structured wavelet transform, the approximate component is used for the next level of decomposition. The wavelet decomposition was performed up to the third level, after which the time-series loses its time dependency and does not carry useful information regarding the border, for the image set under study. A set of six statistical measures of Energy, Mean, Standard deviation, Skewness, Kurtosis and Norm (defined by Equations 4.1–4.8) is then applied on the resultant eight signals of time series; histogram of time series, and the three pairs of approximate and detail components of the wavelet transform, to produce a vector of 48 features.

4.4 Feature Selection

Feature selection is the process of finding an optimal subset of features which provides the highest discriminating power when employed by the classifier. Exclusion of less significant features in the problem space can help the classifier by removing the noisy evidence in the training stage. Feature selection can also reduce the number of features to be considered in the problem space, thus reducing computational cost.

Feature selection algorithms can be categorised into three classes [3]: (1) the filter model, which uses the general characteristics of the data to evaluate the features and selects the optimal feature subset(s) without using any classification/learning algorithm, (2) the wrapper model, which uses a predetermined learning algorithm and searches for features which improve the learning performance, (3) the hybrid model which combines the two approaches. The filter model feature selectors are less computationally expensive than wrapper-based approaches. In this study, due to the high dimension of the feature vector, we have adopted the filter-based model, using the two measures of Gain Ratio and Correlation, which are described in the following.

4.4.1 Gain Ratio-based Feature Selection

To perform feature selection, we have used the Gain Ratio-based feature selection method. It is an entropy-based feature valuator which assesses the worth of an attribute (feature) by measuring its Gain Ratio (GR) with respect to the class (Equation 4.23). The GR is a measure of goodness of the feature and illustrates how precisely the feature predicts the class label. The Ranker search method [217] is used to rank and sort features according to their individual evaluated Gain Ratio. In GRFS, feature selection is done by specifying

a cut-off threshold of the GR, below which features are discarded by the Ranker.

$$Gain - Ratio(C, A) = \frac{H(C) - H(C|A)}{H(A)} \tag{4.23}$$

where A is the attribute or feature, C is the class label, and H specifies the entropy.

4.4.2 Correlation-based Feature Selection

We have also used CFS method which evaluates the worth of a subset of features by considering the individual predictive ability of each feature along with the degree of redundancy between them. This feature evaluator aims to select the subset of features that are highly correlated with the class while having low inter-correlation. The correlation between the two attribute (features) A and B are calculated using the Symmetric Uncertainty measure [217]:

$$Uncertainty(A, B) = 2 \times \frac{H(A) + H(B) - H(A, B)}{H(A) + H(B)} \tag{4.24}$$

where H (entropy) is regarded as a measure of uncertainty about a random variable. *H(A,B)* is the joint entropy. It is worth to mention that the joint entropy of a set of variables is greater than or equal to all of the individual entropies of the variables in the set 4.25. Also, the joint entropy of a set of variables is less than or equal to the sum of the individual entropies of the variables in the set 4.26.

$$H(A, B) \geq \max\{H(A), H(B)\} \tag{4.25}$$

$$H(A, B) \leq H(A) + H(B) \tag{4.26}$$

The Best First method [218] has been applied to search for the best correlated features. It searches the feature space using a greedy hill-climbing algorithm with backtracking, which can either start with the empty set of features (forward direction), or with the full set of features (backward direction), or at any point, and search in both directions.

In our initial investigations (see Section 4.6.1 for details), we have proposed a two-stage feature selection scheme, which uses the GRFS method in the first stage, and applies the CFS method on the subset obtained from the first stage, in the second stage. However, experimental results show that adding the CFS method does not provide a noticeable improvement to the classification performance over that obtained by using the GRFS method alone. Therefore, we have elected to use GRFS only as the feature selection method for all experiments conducted in this study, except for the first experiment as explained above.

Our choice of the GRFS over other compatible filter-based methods[2] is vindicated by a comparison study where its performance is compared with three others; namely, Information Gain, Chi-squared and ReliefF [212]. In this comparison study the entire

[2]Due to computation efficiency, filter-based feature selection methods were chosen and the wrapper-based model has not been considered in this study.

feature vector comprising 35455 features (including 35396 texture, 11 geometry, and 48 time-series features) is ranked individually by the four feature selectors. Then the first 100 highest ranked features by each selector are fed individually into the Random Forest classifier and the accuracy is computed. For all four feature selectors the optimal number of trees of the Random Forest is determined as 6. The results are shown in Figure 4.6, which demonstrates the superiority of the GRFS over the other three.

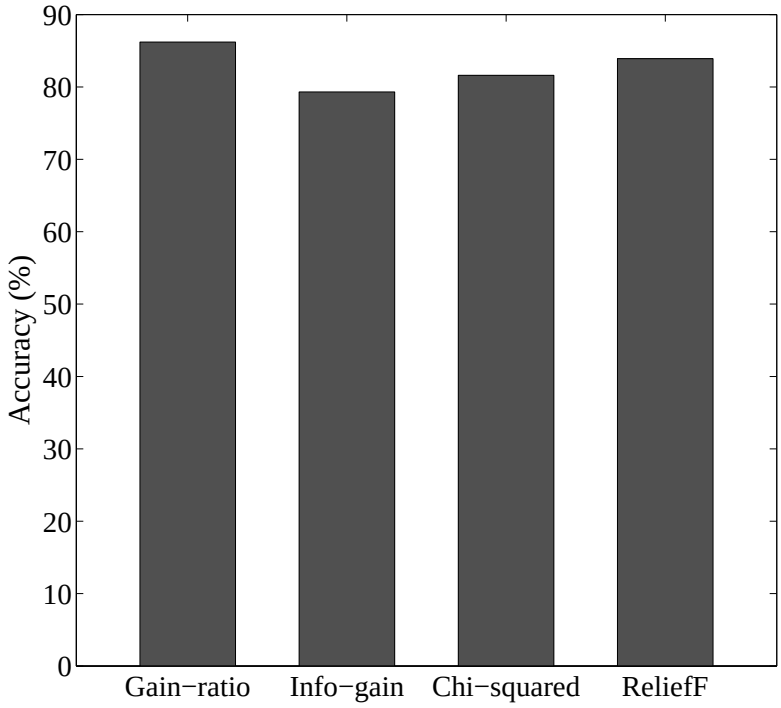

Figure 4.6: Comparying various filter-based feature selection methods.

In the experimental results (Section 4.6.5) we have shown that applying the GRFS method results in significant improvement in the system performance, and a great reduction in the dimension of the feature vector as well as the required time for classification.

4.5 Classification

In this study four different classes of classifier are applied: Support Vector Machine [219] with SMO implementation [220], Random Forest [188], Logistic Model Tree [189] and Hidden Naive Bayes [1]. A brief introduction about each classifier is provided in the following.

4.5.1 Support Vector Machine

Due to its strong theoretical foundation and splendid practical performance, Support Vector Machines have recently drawn a great attention in the machine learning community [60] and have been applied to a versatile range of biomedical applications and computer-based recognition systems of various diseases [221]; namely, oral cancer in optical images, polyps in CT colonography, microclassification detection in mammograms, and more particularly melanoma detection in dermoscopy images [12, 60]. We have also adopted this classifier in our study.

SVMs as proposed by Vapnik are a machine learning paradigm based on statistical learning theory [219]. SVMs perform classification by constructing a set of N-dimensional hyper-planes that optimally separate the given data into classes, using the largest possible margin. Margin is the distance between optimal hyperplane and the nearest training data points of any class, and the larger the margin the lower the generalisation error of the classifier.

The original optimal hyperplane algorithm suggested by Vapnik in 1963 was a linear classifier. For solving linearly non-separable problems, Boser, Guyon and Vapnik in 1992 proposed an approach to create nonlinear classifiers by applying kernel functions [222], which allows the algorithm to fit the maximum-margin hyperplane in a transformed feature space. Some common kernels include Linear, Polynomial, Gaussian or Radial Basis Function (RBF), and Sigmoid. In this study, the RBF kernel is adopted for several reasons [60]: (i) the Linear kernel is not able to handle nonlinearly separable classification problems, and it is in fact a special form of the RBF kernel [223]; (ii) computation of the RBF kernel is more stable than that of the Polynomial kernel, which may return values of zero or infinity in certain cases; (iii) the Sigmoid kernel is only valid (*i.e.*, satisfies Mercer's conditions [224]) for certain parameters; (iv) the RBF kernel (Equation 4.39) has fewer parameters (γ) to be calculated when compared to the Polynomial (γ,r,d) and Sigmoid kernels (γ,r).

Training a support vector machine requires the solution of a very large Quadratic Programming (QP) optimisation problem. We apply the Sequential Minimal Optimisation (SMO) algorithm [220] for solving this optimisation problem. SMO breaks this large QP problem into a series of smallest possible QP problems. These small QP problems are solved analytically, consuming less time, and the amount of memory required scales linearly, which allows SMO to handle very large training sets.

SVM Theoretical Background [224, 220]:

Consider a given set D of n training data points:

$$D = \left\{(x_i, y_i)|x_i \in \Re^d, y_i \in \{-1, +1\}, i = 1, 2, ..., n\right\} \tag{4.27}$$

where x_i indicates a point in d-dimensional space (d-dimensional vector), and y_i refers to a two-class label. Suppose we have a hyperplane that separates the positive examples from

the negative ones. The points x on such hyperplane would satisfy the following condition:

$$w.x + b = 0 \tag{4.28}$$

where w is the normal vector to the hyperplane and dot(.) is the dot products of the two vectors. The parameter $\frac{|b|}{||w||}$ is the perpendicular distance (offset) from the hyperplane to the origin along the normal vector ($||w||$ is the Euclidean norm of w). If we take two such hyperplanes between the positive and negative samples, the SVM task is to maximise the distance (margin) between them. Assuming d_+ and d_- to be the shortest distance from the hyperplanes to the closest positive and negative samples, the margin is defined as $d_+ + d_-$. For a linearly separable set of samples, the training data would satisfy the following constraints:

$$x_i.w + b \geq +1 \;\; for \;\; y_i = +1 \tag{4.29}$$

$$x_i.w + b \leq -1 \;\; for \;\; y_i = -1 \tag{4.30}$$

which can be combined to yield:

$$y_i(x_i.w + b) - 1 \geq 0 \;\; \forall i \tag{4.31}$$

The samples satisfying Equation 4.29 lie on the hyperplane $H1 : x_i.w+b = 1$ with distance $\frac{|1-b|}{||w||}$ from the origin, and those satisfying Equation 4.29 are a member of hyperplane $H2 : x_i.w + b = -1$ with distance $\frac{|-1-b|}{||w||}$. Thus, $d_+ = d_- = \frac{1}{||w||}$ and the margin is $\frac{2}{||w||}$. In order to find the pair of hyperplanes with the maximum margin, $||w||^2$ is minimised, subject to constraints given in Equation 4.31 (the optimisation problem). The training samples satisfying those are called support vectors.

Using the following Lagrangian formulation, the optimisation problem is converted into a dual form which is a QP problem:

$$L_{linear} = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} y_i y_j \alpha_i \alpha_j x_i x_j \tag{4.32}$$

subject to the inequality constraints:

$$\alpha_i \geq 0 \quad \forall_i \tag{4.33}$$

and one linear equality constraint:

$$\sum_{i=1}^{n} \alpha_i y_i = 0 \tag{4.34}$$

For non-linear separable data, positive slack variables ξ_i, $(i = 1, 2, ..., n)$ are added to the constraints. Also the user chosen parameter C may be placed as an upper bound on the parameter α to act as a penalty parameter associated with classification errors. A larger C value corresponding to a higher penalty for errors ($0 \leq \alpha_i \leq C$).

$$y_i(x_i.w + b) \geq 1 - \xi_i \;\; , \;\; \xi \geq 0 \forall_i \tag{4.35}$$

Moreover, a general kernel function K replaces the dot products:

$$L_{kernel} = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} y_i y_j \alpha_i \alpha_j K(x_i, x_j) \tag{4.36}$$

subject to

$$0 \leq \alpha_i \leq C \tag{4.37}$$

and

$$\sum_{i=1}^{n} \alpha_i y_i = 0 \tag{4.38}$$

The RBF kernel is given by:

$$K(x_i, x_j) = e^{\frac{-\gamma ||x_i - x_j||^2}{2\sigma^2}} \quad , \quad \gamma > 0 \tag{4.39}$$

Detailed explanation of the algorithm can be found in [224, 220].

4.5.2 Random Forest

Random forest is an ensemble classifier which grows a number of decision trees. To classify an unknown object (image, in our application), the constructed feature vector is sent down every tree in the forest. Each individual tree independently classifies the given feature vector (each tree votes for a class) and the forest chooses the classification result with the most votes, over all the trees in the forest. Thus, the RF's output is the class label which is the statistical mode of the class's output by individual trees [188].

Random Forest is an appealing choice for applications with noisy data, datasets containing many missing values, when some of the attributes are categorical or semi-continuous, and also for high-dimensional feature vectors with highly correlated features (the situation which often occurs in bioinformatics and medical diagnosis applications [225]). RF is very robust with respect to noise, runs efficiently on large data bases and can handle thousands of input variables [188]. Each tree is grown as follows [188]:

(1) A random sample is taken from the data set (including N observations) with replacement. This process is called bagging. Some observations will be selected more than once, and others will not be selected. On average, about 2/3 of the data will be selected by the sampling. The remaining 1/3 of the data are called the Out Of Bag (OOB). A new random selection of data is performed for each tree constructed.

(2) Using the data selected in step 1, a decision tree is constructed. The tree is built to the maximum size, with no pruning. As the tree is built, only a subset of the total set of predictor variables (features) are considered as possible splitters for each node; *i.e.*, if there are M features, a number $m < M$ is specified such that at each node, m features are selected at random out of the M and the best split on these m features is used to split the node. The value of m is held constant during the forest growing.

(3) Steps 1 and 2 are repeated for a large number of times to construct a forest of trees.

(4) In order to score a given data, the data is run through each tree in the forest and the predicted class label (*i.e.*, terminal node) that the data ends up in, is recorded. For a regression analysis, the average score predicted by all of the trees is calculated. For a classification analysis, the predicted class for each tree is considered as votes, and the class with the most votes is used as the predicted class label for the data.

Random forests involve two stochastic (randomising) elements: the selection of data used as input for each tree, and the set of predictor variables (features) considered as candidates for each node split. A remarkable property of RF is that they always converge so that overfitting is not a problem [188]. Overfitting is a problem in large, single-tree models where the model begins to fit noise in the data. When such a model is applied to data not used to build the model, the model does not perform well (*i.e.*, it does not generalise well). To avoid this problem, single-tree models must be pruned to the optimal size. In nearly all cases, decision tree forests do not have a problem with overfitting, and there is no need to prune the trees in the forest. Generally, the more trees in the forest, the better the fit. More information of the method can be found in [188].

4.5.3 Logistic Model Tree

Logistic Model Trees combine logistic regression and tree induction to perform supervised learning. LMTs are classification trees with logistic regression functions at the leaves. Landwehr *et al.* [189] proposed this approach for building logistic models, using a CART-based algorithm [226] for pruning the tree. It is shown [189] that LMT is more accurate than C4.5 decision trees[3] [227] and standalone logistic regression on real-world datasets, and even competitive with boosted C4.5 trees.

As mentioned, a Logistic Model Tree basically consists of a standard decision tree structure with logistic regression functions at the leaves. Hence, similar to ordinary decision trees, a test on one of the attributes (features) is associated with every inner node. For a nominal attribute with k values, the node has k child nodes, and instances are sorted down one of the k branches depending on their value of the attribute. For numeric attributes, the node has two child nodes and the test consists of comparing the attribute value to a threshold: an instance is sorted down the left branch if its value for that attribute is smaller than the threshold and sorted down the right branch otherwise. LMT consists of a tree structure that is made up of a set of inner or non-terminal nodes N and a set of leaves or terminal nodes T. Suppose S denotes the whole instance space, spanned by all attributes that are present in the data. Then the tree structure gives a disjoint subdivision of S into regions S_t, and every region is represented by a leaf in the tree:

$$S = \bigcup_{t \in T} S_t \quad , \quad S_t \cap S_{t'} = \phi \quad for \quad t \neq t' \tag{4.40}$$

[3] C4.5, developed by Ross Quinlan, is an algorithm used to generate a decision tree. It is an extension of Quinlan's earlier ID3 algorithm which can be used for classification [227].

Unlike in ordinary decision trees, the leaves $t \in T$ have an associated logistic regression function f_t instead of just a class label. If V is the subset of all attributes present in the data, then the regression function ft is a subset of V, *i.e.*, the regression function f_t, $V_t \subseteq V$ (where we assume that nominal attributes have been binarised for the purpose of regression), and models the class membership probabilities as:

$$Pr(G = j|X = x) = \frac{e^{F_j(x)}}{\sum_{k=1}^{J} e^{F_k(x)}} \tag{4.41}$$

where

$$F_j(x) = \alpha_0^j + \sum_{v \in V_t} \alpha_v^j . v \tag{4.42}$$

or, equivalently,

$$F_j(x) = \alpha_0^j + \sum_{k=1}^{m} \alpha_{vk}^j . v_k \tag{4.43}$$

If $\alpha_{vk}^j = 0$ for $v_k \notin V_t$. Model represented by LMT is then given by:

$$f(x) = \sum_{t \in T} f_t(x) . I\left(x \in S_t\right) \tag{4.44}$$

where

$$I\left(x \in S_t\right) = \begin{cases} 1 & if \;\; x \in S_t \\ 0 & \text{otherwise} \end{cases} \tag{4.45}$$

It is worth to remember that both standalone logistic regression and ordinary decision trees are special cases of LMT, the first is a logistic model tree pruned back to the root, the second a tree in which $V_t = \phi$ for all $t \in T$. Further details of the method may be found in [189].

4.5.4 Hidden Naive Bayes [1, 2]

Hidden Naive Bayes is a promising attempt to improve the performance of Naive Bayes method. The conditional independence assumption of Naive Bayes, which ignores attribute dependencies, is rarely true in real-world applications. On the other hand, although a Bayesian network can represent arbitrary attribute dependencies, learning an optimal structure of a Bayesian network classifier is an NP[4]-hard problem [228] and extremely time-consuming. Restricting the structures of Bayesian networks, similar to what is done in TAN (extended Tree-like Naive Bayes) [229], results in acceptable computational complexity and a significant improvement over naive Bayes. However, the main issue in learning TAN is that only one attribute parent is allowed for each attribute and the method disregards influences from other attributes. In addition, in a TAN learning

[4]non-deterministic polynomial-time

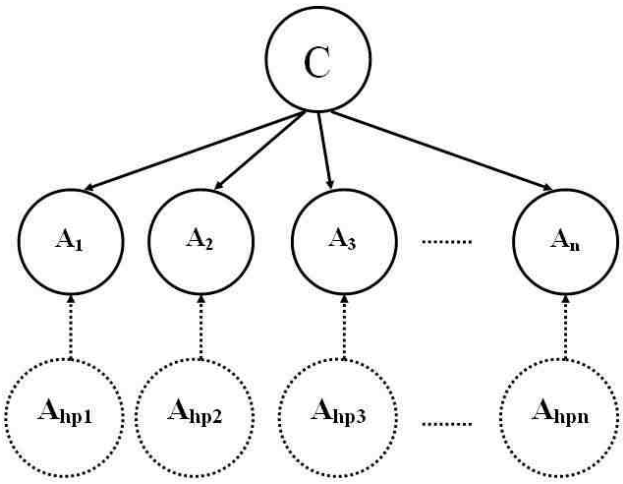

Figure 4.7: The structure of HNB [2].

algorithm, structure learning still exists. Yet, HNB aims to represent attribute dependencies to some proper degree, without structure learning. This is done by creating a hidden parent for each feature (attribute), which combines the influences of all other features using conditional mutual information. The conditional mutual information is defined as:

$$I_P(X;Y|Z) = \sum_{x,y,z} P(x,y,z) log \frac{P(x,y|z)}{P(x|z)P(y|z)} \tag{4.46}$$

where x, y and z are the values of variables X, Y and Z respectively.

Figure 4.7 illustrates the structure of an HNB [2], where C, the class node, is the parent of all attribute nodes. Each attribute A_i has a hidden parent A_{hp_i} $(i = 1, 2, ..., n)$ shown by a dashed circle. The joint distribution represented by an HNB is defined as:

$$P(A_1, A_2, ..., A_n, C) = P(C) \prod_{i=1}^{n} P(A_i|A_{hp_i}, C) \tag{4.47}$$

where

$$P(A_i|A_{hp_i}, C) = \sum_{j=1, j\neq i}^{n} W_{ij} \times P(A_i|A_j, C) \tag{4.48}$$

The HNB classifier on an example $E = (a_1, ..., a_n)$ is given by:

$$c(E) = \arg\max_{c \in C} P(c) \prod_{i=1}^{n} P(a_i|a_{hp_i}, c) \tag{4.49}$$

where

$$P(a_i|a_{hp_i}, c) = \sum_{j=1, j\neq i}^{n} W_{ij} \times P(a_i|a_j, c) \tag{4.50}$$

The hidden parent A_{hp_i} for A_i is in fact a mixture of the weighted influences from all other attributes. In order to determine the weights, the conditional mutual information between two attributes A_i and A_j is considered, defined as follows:

$$W_{ij} = \frac{I_P(A_i; A_j|C)}{\sum_{j=1,j\neq i}^{n} I_P(A_j; A_j|C)} \tag{4.51}$$

where $I_P(A_j; A_j|C)$ is the conditional mutual information given by:

$$I_P(A_i; A_j|C) = \sum_{a_i,a_j,c} P(a_i, a_j, c) \log \frac{P(a_i, a_j|c)}{p(a_i|c)P(a_j|c)} \tag{4.52}$$

Further details of the method may be found in [2].

4.6 Experimental Results

In this section, a series of six different experiments is carried out to evaluate the optimal set of features, the most effective feature selection approach and the best classification method for classification of dermoscopy images. Three image sets have been used in this study; *Image-set1* is used in the first experiment, *Image-set2* is utilised in the next three experiments (Experiments 2a, 2b and 2c), and *Image-set3* is used in last two Experiments (3a and 3b). Table 4.1 shows the distribution of images in each image set.

Table 4.1: Image sets used in the study; M : Malignant, B : Benign.

	Train set	Validation set	Test set	Experiment
Image-set1: 96 M, 109 B	48 M, 55 B	-	48 M, 54 B	1
Image-set2: 88 M, 109 B	48 M, 55 B	-	40 M, 54 B	2a, 2b, 2c
Image-set3: 114 M, 175 B	40 M, 59 B	30 M, 57 B	44 M, 59 B	3a, 3b

The first experiment (Experiment 1) is designed to investigate the effectiveness of (i) the proposed wavelet-based texture features, (ii) two-stage feature selection method based on Gain Ration and Correlation, (iii) the four classifiers SVM, RF, LMT and HNB, (iv) the optimal texture features as opposed to some existing texture features utilised in melanoma diagnosis applications.

The purpose of next three experiments (Experiments 2a, 2b and 2c) is to compare the classification performance resulting from the three sets of features extracted from (i) texture analysis, (ii) texture and geometrical measurements analysis, and (iii) texture, geometrical and time-series. Figure 4.8 illustrates the flowchart of these three experiments.

The last two experiments (Experiment 4a and 4b as labelled in Figure 4.9) involve two streams of feature optimisation to assess the performance of the developed diagnostic system in an unbiased way.

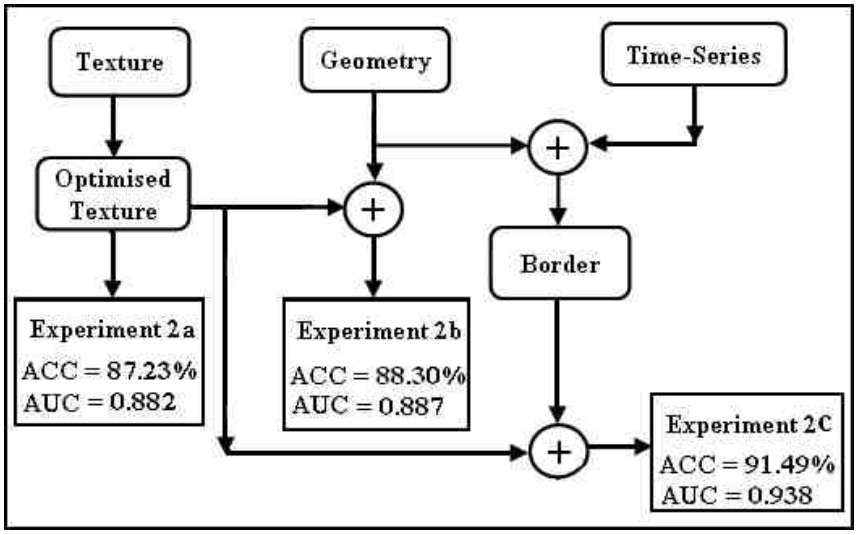

Figure 4.8: Experiments flowchart: investigating the effectiveness of three sets of features extracted from: texture analysis (Experiment 2a), texture and geometrical measurements analysis (Experiment 2b), and (iii) texture, geometrical and time-series (Experiment 2c).

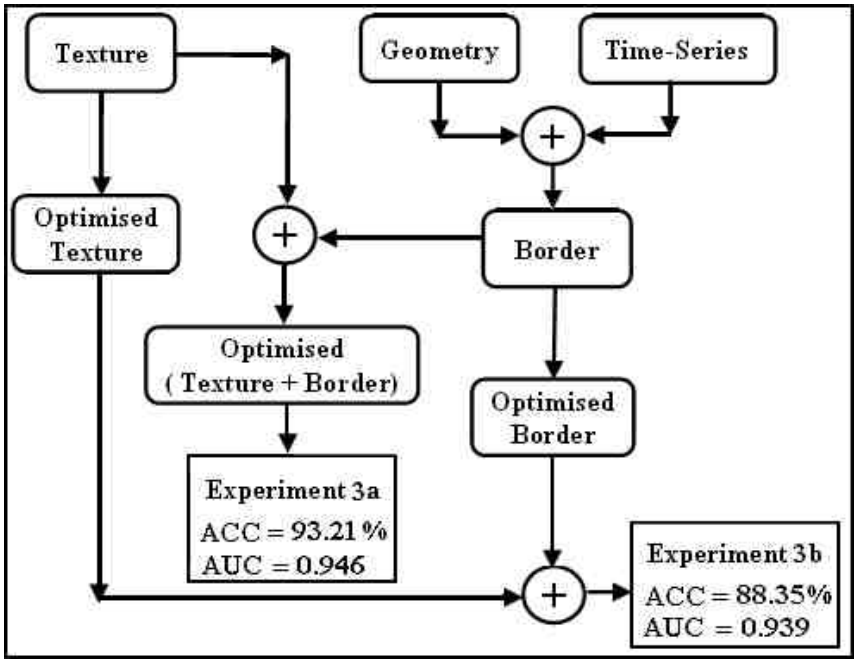

Figure 4.9: Experiments flowchart: Performance evaluation of the proposed computer-aided Diagnosis of melanoma including the two streams of optimisation.

4.6.1 Experiment 1: Analysing Texture Features

The purpose of this experiment is to study the effectiveness of the texture-based features proposed in Section 4.1.1 in the classification of melanoma. An image set of 205 digital dermoscopy images (Image-Set1 from Table 4.1) obtained from the Interactive Atlas of Dermoscopy [29], is applied in this experiment. These were JPEG images with a resolution of 768×512 pixels. The diagnosis distribution include 96 melanoma and 109 benign (65 Clark nevi, 14 blue nevi, 30 Reed/Spitz nevi), where the histopathological results have been determined by biopsy or excision.

Preprocessing

In this experiment, the lesion borders are obtained manually, according to practice used by dermatologists. Moreover, a rectangle frame enclosing the lesion with horizontal and vertical sides is set around the lesion and the rest is cropped. Also, within this frame, pixels not belonging to the lesion are set to black. This process is done to minimise the impact of background skin on the analysis and maximise the focus on lesion pixels.

Feature Extraction and Feature Selection

In order to determine the most significant colour and texture features for classification of melanoma, we have designated half of the images as *train set* (the first row of Table 4.1) to apply the feature selection. The texture-based feature extraction method (explained in Section 4.1.1) is applied to images of the train set and the corresponding feature vectors are constructed. MATLAB Daubechies-3 wavelet function is used for the wavelet decomposition.

For selecting the optimal features a two-stage approach is used; at the first stage, the GRFS method is applied on the constructed feature vectors, resulting in various Gain Ratio values or evaluation rank. We consider these values as threshold; each corresponds to a certain subset of texture features. Figure 4.10 shows the number of features selected by the Gain-ratio feature selection method, at different feature evaluation ranks (thresholds). Figure 4.11 shows the obtained accuracy and the required time for building the classification model at different feature evaluation ranks (set as threshold for the Ranker search method) using the SVM classifier with RBF kernel (gamma=0.1). As illustrated in the figure, by reducing the number of features the accuracy is increased and then decreased after a certain threshold; the so-called peak phenomenon [230]. Therefore, the process of feature selection or optimisation is to find the optimal threshold. As expected and also shown in Figure 4.11 the time required for performing classification is proportional to the dimension of the optimal feature vector and thus decreases considerably as a result of the feature selection process.

By visual inspection, we have chosen eight thresholds and the eight resultant feature subsets, as shown in Table 4.3. At the second stage of feature selection correlation-based

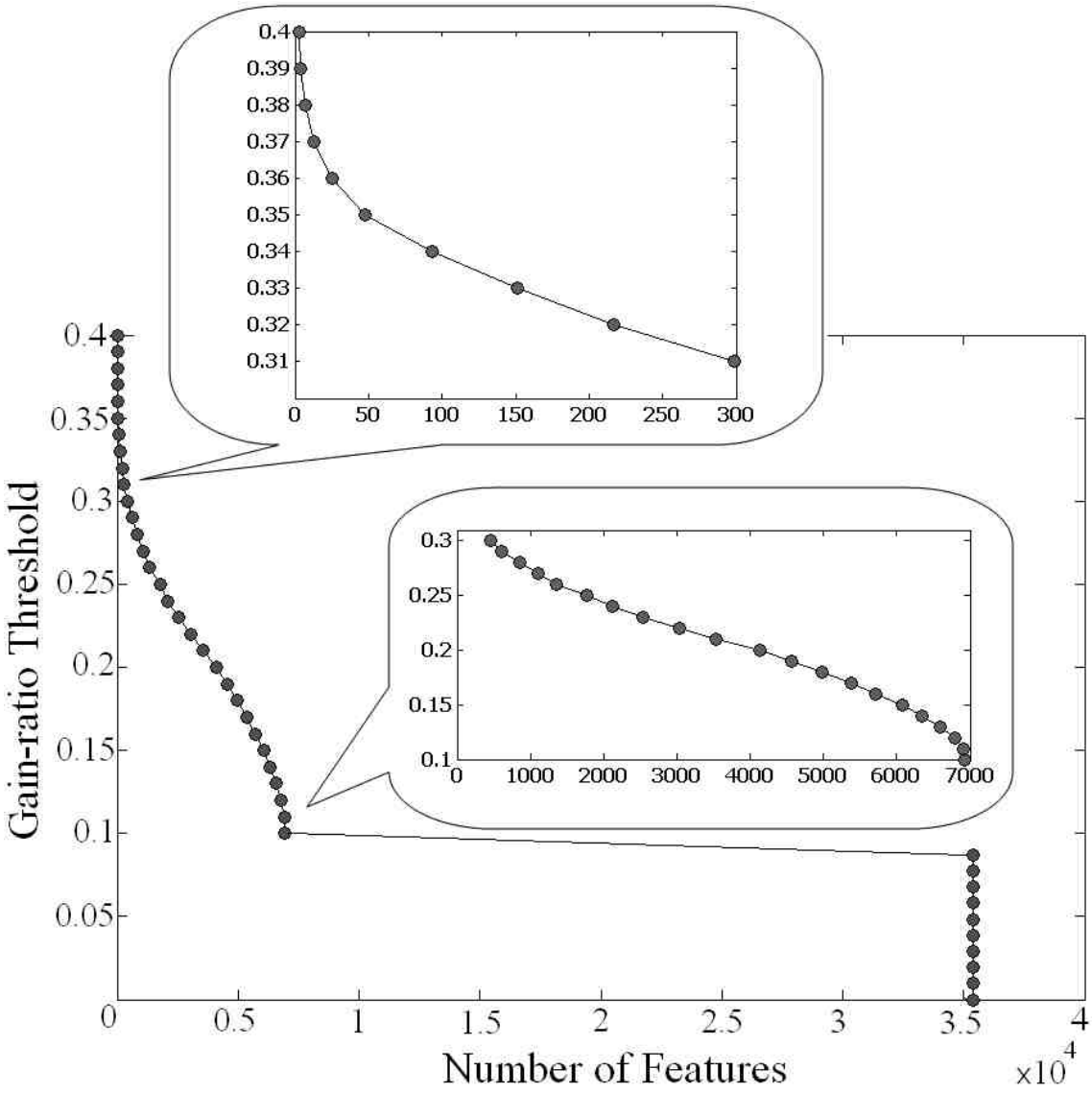

Figure 4.10: Number of features selected by Gain Ratio feature selection method at different thresholds.

feature evaluation along with best-first search method in forward and backward directions (explained in Section 4.4.2) are applied on the eight subsets resulted from the first stage, yielding 16 feature configurations, shown in Table 4.4.

Classification

To test the effectiveness of the feature subsets selected in the previous step, we have applied the four classifiers of SVM, RF, MLT and HNB described in Section 4.5, using the

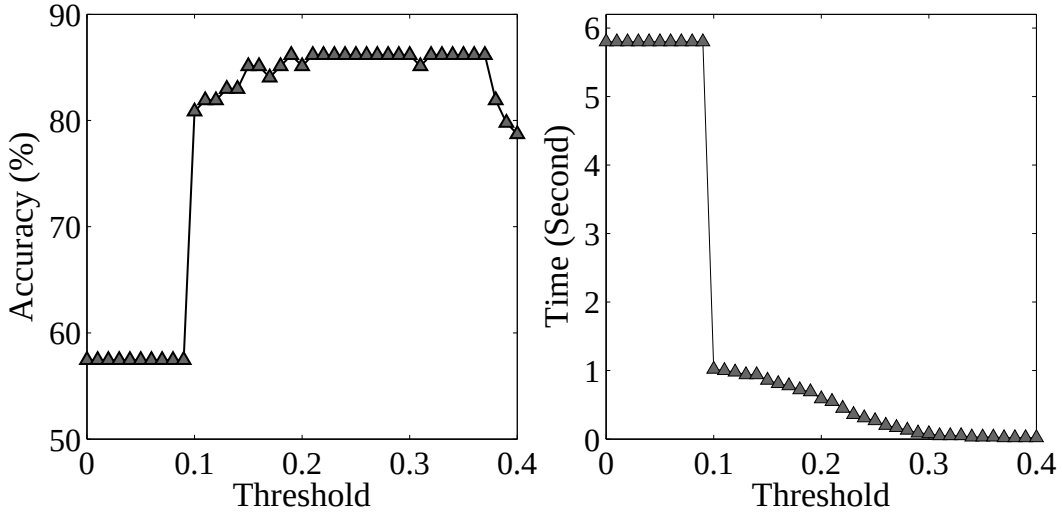

Figure 4.11: Obtained accuracy and required time for building the classification model for different thresholds using SVM classifier with RBF kernel.

corresponding parameter range shown in Table 4.2, on the *test set* (See the first row of Table 4.1); parameters outside the indicated regions did not produce useful results. In the classification experiments, the highest accuracy obtained by each classifier is used as the measure of its performance. The Weka data mining toolkit [217, 231] has been utilised in this study. Having the two independent image sets for feature selection and classification removes any bias from the feature subsets used for classification.

Table 4.2: Parameters setting for the classifiers

Classifier	Parameter range
Support Vector Machine (RBF)	Gamma = [0.001, 0.5]
Random Forest	number of trees = [2, 200]
Logistic Model Tree	weightTrimBeta = [0, 0.9]
Hidden Naive Bayes	No parameters

Table 4.3 shows the 10-fold cross validation [199] of performance (accuracy and AUC value) of different classifiers on the test set, using the eight feature configurations resulting from the first stage of feature selection on the train set. The "T" column shows the ranking threshold (Gain Ratio value) for including the features in the subset and column "N" shows the number of features. Table 4.4 shows the 10-fold cross validation of performance (accuracy and AUC value) of the four classifiers on the test set, using the 16 feature configurations resulting from the second stage of feature selection on the train set. The "D" column refers to the direction of the best-first method, and F and B show the forward and backward directions, respectively. According to Tables 4.3 and 4.4,

Table 4.3: 10-fold cross validation of performance (accuracy and ROC area) of different classifiers on the test set, using the 8 feature configurations resulting from the first stage of feature selection on the train set.

Conf.	T	N	SVM	RF	LMT	HNB
1	0.30	451	**86.27**, 0.862	**86.27**, 0.917	**88.24**, 0.918	**86.27** , 0.896
2	0.31	299	**86.27**, 0.862	**86.27**, 0.922	82.35, 0.886	**86.27** , 0.892
3	0.32	217	**86.27**, 0.862	85.29, 0.923	**86.27**, 0.908	84.31 , 0.91
4	0.33	151	**86.27**, 0.862	85.29, 0.917	83.33, 0.902	85.91,0.907
5	0.34	93	85.29, 0.853	**86.27**, 0.921	**87.25**, 0.917	**86.27**, 0.921
6	0.35	48	85.29, 0.852	85.29, 0.901	**86.27**, 0.871	85.29, 0.885
7	0.36	25	84.31, 0.843	85.29, 0.897	83.33, 0.865	**86.27**, 0.878
8	0.37	13	**86.27**, 0.862	**87.25**, 0.893	84.31, 0.891	82.35 , 0.87

Table 4.4: 10-fold cross validation of performance (accuracy and ROC area) of different classifiers on the test set, using the 16 feature configurations resulting from the second stage of feature selection on train set.

Conf.	Thrsh	D	N	SVM	RF	LMT	HNB
9	0.30	F	28	85.29 , 0.852	81.37 , 0.894	81.37 , 0.88	85.29, 0.899
10	0.30	B	27	85.29 , 0.852	82.35 , 0.897	81.37 , 0.877	83.33 , 0.888
11	0.31	F	33	85.29 , 0.852	82.35 , 0.893	80.39 , 0.843	85.29 , 0.893
12	0.31	B	25	84.31 , 0.843	80.39 , 0.875	82.35 , 0.89	84.31 , 0.883
13	0.32	F	29	84.31 , 0.844	82.35 , 0.896	82.35 , 0.877	**86.27** , 0.896
14	0.32	B	29	85.29 , 0.854	83.33 , 0.902	80.39 , 0.878	84.31 , 0.898
15	0.33	F	29	84.31 , 0.843	85.29 , 0.906	84.31 , 0.892	84.31 , 0.9
16	0.33	B	29	85.29 , 0.853	85.29 , 0.898	81.37 , 0.88	84.31 , 0.901
17	0.34	F	17	82.35 , 0.822	84.31 , 0.9	82.35 , 0.866	83.33 , 0.907
18	0.34	B	23	85.29 , 0.853	85.29 , 0.914	82.35 , 0.868	83.33 , 0.907
19	0.35	F	21	84.31 , 0.844	**86.27** , 0.895	85.29 , 0.907	82.35 , 0.881
20	0.35	B	20	85.29 , 0.854	85.29 , 0.891	**86.27** , 0.906	81.37 , 0.876
21	0.36	F	8	**86.27** , 0.863	**86.27**, 0.903	83.33 , 0.885	82.35 , 0.874
22	0.36	B	8	**86.27** , 0.863	**86.27** , 0.894	81.37, 0.87	82.35 , 0.875
23	0.37	F	6	80.39 , 0.803	83.33 , 0.893	85.29 , 0.887	82.35 , 0.87
24	0.37	B	6	80.39 , 0.803	83.33 , 0.902	85.29 , 0.881	82.35 , 0.87

the highest accuracy (88.24%) and AUC value (0.918) are obtained by the first feature configuration, *i.e.*, applying the GRFS method at the threshold of 0.3 which results in 451 features, when classified by LMT. This is followed by 5-th and 8-th configurations, where the GRFS method was used with thresholds of 0.34 (using the LMT classifier) and 0.37 (using the RF classifier) resulting in 93 and 13 optimal features. Both feature sets obtain an accuracy of 87.25%. However, the AUC value for 93-element subset is 0.917, and 0.893 for 13-element feature set.

Multiple feature configurations; *e.g.* 2, 3, 4, 6, 7, 19, 20, 21 and 22 exhibit an accuracy of 86.27%, among which 21 and 22 have the smallest number of features, *i.e.*, eight optimal features. The decision of which configuration is best to use is a trade-off between having a more accurate classification decision and the expense of increased complexity and computational cost.

Table 4.5 shows the 13 optimal features selected by the 8-th feature configuration, which resulted in the accuracy of 87.25%. As indicated in Table 4.5, certain measures and nodes appear more frequently in the optimal feature subset suggested by this configuration, such as the kurtosis and average-energy measures, and 4.3.4.4 and 4.1.4 nodes of the wavelet tree. Moreover, 60% of the optimal features happen to lie in the fourth level of the wavelet decomposition. In this feature configuration, the Luminance channel is the dominant one, following by Red. Blue and Green each have two occurrences.

Table 4.5: The 13 optimal features selected by the 8-th feature configuration, resulted in the accuracy of 87.25%.

Measure	colour Channel	Tree Node
Kurtosis	Red	4.3.4.4
Kurtosis	Red	2.4.1
Kurtosis	Luminance	4.3.4.4
Kurtosis	Green	4.3.4.4
Kurtosis	Blue	4.1.4
Average-energy	Luminance	3.2.1.3
Kurtosis	Green	2.2.4
Average-energy	Luminance	4.1.4
Average-energy	Luminance	4.1.4.4
Average-energy	Red	4.1.4
Average-energy	Blue	4.1.4.4
Standard deviation	Luminance	4.2.4.1
Engergy	Luminance	4.2.4.1

Comparison with others methods

The proposed feature extraction method has been compared with three other wavelet-based melanoma classification methods proposed by Nimukar *et al.* [171], Patwardhan *et al.* [210] and Walvick *et al.* [169]. In order to make this comparison we have applied the suggested feature sets by [171], [210] and [169] on our test set of 102 images (the first row of Table 4.1). Table 4.6 shows the classification accuracy and obtained AUC value of our best three feature configurations along with the result obtained by the above-mentioned methods. The parameters of the classifiers have been tuned in a way that the highest performance for the given feature set is gained. According to the table, our proposed method (with three different feature configurations) outperforms the other three wavelet-

based methods. However, the best performance among the three other methods is obtained by Walvick *et al.* [169] with the accuracy of 86.27% and AUC value of 0.901 when the RF classifier is used. It is followed by Nimukar *et al.* [171] with the highest accuracy of 85.29% and Patwardhan [210] 's methods come last.

Table 4.6: Comparative study: 10-fold cross validation of performance (accuracy and AUC value) of different classifiers on test set, using features proposed by [171], [210] and [169], and our best three feature configurations.

Method	N	SVM	RF	LMT	HNB
configuration 1	451	**86.27**, 0.862	**86.27**, 0.917	**88.24**, 0.918	**86.27** , 0.896
configuration 5	93	85.29, 0.853	**86.27**, 0.921	**87.25**, 0.917	**86.27**, 0. 921
configuration 8	13	**86.27**, 0.862	**87.25**, 0.893	84.31, 0.891	82.35 , 0.87
Nimukar [171]	5	83.33, 0.834	85.29 , 0.91	85.29 , 0.879	82.35 , 0.81
Patwardhan [210]	5	52.94, 0.5	62.74, 0.623	62.74, 0.692	52.94, 0.507
Walvick [169]	4	82.35, 0.822	**86.27**, 0.901	82.35, 0.904	79.41, 0.759

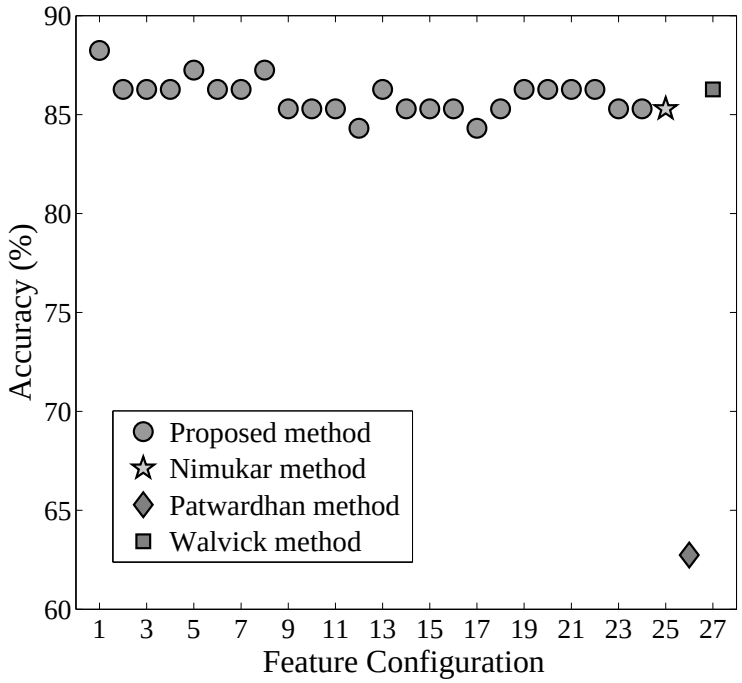

Figure 4.12: Accuracy of various feature configurations proposed by the proposed texture-based feature configurations as opposed to three other feature extraction methods by [171], [210] and [169].

Figure 4.12 shows the obtained accuracies of various feature configurations proposed by our method as opposed to other three feature extraction methods by [171], [210]

and [169]. These accuracies are the highest of all four classifiers (SVM, HNB, LMT and RF) for each feature extraction/selection method. As shown in the figure, our method outperforms the other three wavelet-based methods.

In the following we further extend the feature vector and evaluate the usefulness of border-based features for classification of melanoma. As illustrated in Figure 4.8, the purpose of next three experiments (Experiments 2a, 2b and 2c) is to compare the classification performance resulting from the three sets of features extracted from (i) texture analysis, (ii) texture and geometrical measurements analysis, and (iii) texture, geometrical and time-series. According to the experimental results obtained in Experiment 1, adding the CFS method does not provide a noticeable improvement to the classification performance over that obtained by using the GRFS method only. Hence, we will use a one-stage feature selection method and will apply the Gain Ration feature selection method in the following experiments of this chapter.

4.6.2 Experiment 2a: Optimal Texture Features

This experiment which investigates the effect of texture-based features on classification of melanoma, is in fact a replication of the previous experiment (Experiment 1), but using the new image set of 197 dermoscopy images (Image-set2 from Table 4.1). This repetition is done in order to make a valid comparison between the result of this experiment and the two following experiments; Experiment 2b and 2c. The difference in the two image sets emanates from excluding eight lesions, whose poor quality made it impossible to perform accurate border-based measurements. Similar to the previous experiment, the following initial steps are conducted:

(i) Dermoscopy images undergo pre-processing in order to minimise the impact of background skin in the analysis, wherein the lesion borders are obtained manually by using dermatologists' expertise. Then, an enclosing rectangle frame is set around the lesion, the rest is cropped and the non-lesion pixel is set to black. This pre-processing step will be used in all of the four experiments of this chapter.

(ii) Using the MATLAB Daubechies-3 wavelet function, the texture-based features (explained in Section 4.1 of this chapter) are extracted from half of the *Image-Set2* designated as the *train image set* (See the second row of table 4.1).

(iii) The GRFS method is applied on the constructed feature vectors, resulting in various Gain Ratio values or evaluation rank. We consider these values as threshold; each corresponding to a certain subset of texture features.

(iv) By visual inspection, eleven of these thresholds are empirically chosen in the range of [0.27 , 0.37] and the corresponding texture features are extracted from the second half of the dermoscopy images designated as the *test set* (the second row of table 4.1). The four classifiers of SVM, FR, LMT and HNB are then employed. This has resulted in the classification accuracy shown in Table 4.7, which indicates that the best classification result (accuracy of 87.23% and AUC value of 0.882) is obtained by applying the Random

Forest classifier using the 9-th configuration with 48 texture-based features. The "T" column shows the ranking threshold (Gain Ratio value) for including the features in the subset and column "N" shows the number of features.

Table 4.7: 10-fold cross validation of performance (accuracy and AUC) of different classifiers on the test set, using the 11 texture feature configurations resulting from the GRFS method.

Configuration	T	N	SVM (RBF)	RF	LMT	HNB
1	0.27	1108	86.17, 0.854	86.17, 0.889	82.98, 0.886	86.17, 0.885
2	0.28	852	85.11, 0.841	84.04, 0.887	84.04, 0.836	88.30, 0.887
3	0.29	609	86.17, 0.854	86.17, 0.899	80.85, 0.825	87.23, 0.882
4	0.30	451	87.23, 0.866	85.11, 0.868	85.11, 0.871	86.17, 0.881
5	0.31	299	85.11, 0.844	85.11, 0.885	82.98, 0.855	85.11, 0.873
6	0.32	217	86.17, 0.854	82.98, 0.887	84.04, 0.870	86.17, 0.888
7	0.33	151	86.17, 0.857	85.11, 0.901	84.04, 0.881	86.17, 0.894
8	0.34	93	87.23, 0.869	86.17, 0.883	79.79, 0.842	84.04, 0.902
9	0.35	48	87.23, 0.869	**87.23, 0.882**	84.04 , 0.846	85.11, 0.880
10	0.36	25	86.17, 0.863	85.11, 0.856	82.98, 0.890	82.98, 0.855
11	0.37	13	86.17, 0.854	85.11, 0.859	84.04, 0.871	84.04, 0.852

In the next two experiments of this study, we expand the feature vector by augmenting the optimal texture features with geometry and time series features. We then investigate the impact of such an expansion by comparing the classification results obtained from (1) texture, (2) texture and geometry, and (3) texture, geometry and time-series features. As mentioned earlier, test image set used in these experiments (40 melanoma and 55 benign, second row of Table 4.1) is slightly different from the test image set used in Experiment 1 (48 melanoma and 55 benign, first row of Table 4.1).

4.6.3 Experiment 2b: Optimal Texture and Geometry Features

The purpose of this experiment as depicted in Figure 4.8 is to analyse the effect of integrating geometry with textural features and assess the classification performance. The geometry features are extracted from the images in the *test set* of *Image-set2* (See the second row of Table 4.1). The optimal texture feature sets are obtained from applying the GRFS method as per experiment 2a. The optimal texture feature sets are obtained from applying the GRFS method as per experiment 1. Eleven optimal features sets are chosen corresponding to the eleven thresholds shown in Table 4.7.

The whole appended feature vector (texture plus geometry) is then fed into the same four classifiers to ascertain the status of the 103 lesions. Table 4.8 shows the 10-fold cross-validation of classification performance where the highest accuracy for each of the classifier is quoted. The "T" column shows the ranking threshold (Gain Ratio value) for including the features in the subset and column "N" shows the number of features. Examination of

Tables 4.7 and 4.8 shows that there is an overall improvement in the classification results as a consequence of augmenting texture with geometry-based features; when using the RF classifier, the maximum classification accuracy has increased from 87.23% (texture) to 88.30% (texture and geometry). Figure 4.13 illustrates this effect more clearly on the RF classifier, which shows that for most of the thresholds the classifier accuracy increases when the geometry-based features are added.

Table 4.8: 10-fold cross validation of performance (accuracy and AUC) of different classifiers on the test set, using the 11 texture feature configurations resulting from the Gain Ratio feature selection, combined with geometry features.

Configuration	T	N	SVM (RBF)	RF	LMT	HNB
12	0.27	1119	86.17, 0.854	86.17, 0.874	82.98, 0.886	86.17, 0.878
13	0.28	863	85.11, 0.841	87.23, 0.878	84.04, 0.836	88.30, 0.856
14	0.29	620	86.17, 0.854	85.11, 0.887	80.85, 0.825	86.17, 0.881
15	0.30	462	87.23, 0.866	84.04, 0.884	82.98, 0.863	86.17, 0.883
16	0.31	310	85.11, 0.844	85.11, 0.837	82.98, 0.855	85.11, 0.873
17	0.32	228	85.11, 0.844	85.11, 0.884	84.04, 0.858	86.17, 0.888
18	0.33	162	86.17, 0.857	86.17, 0.913	84.04, 0.882	86.17, 0.894
19	0.34	104	87.23, 0.869	86.17, 0.891	81.91, 0.875	84.04, 0.898
20	0.35	59	87.23, 0.869	87.23 , 0.881	82.98, 0.816	85.11, 0.879
21	0.36	36	85.11, 0.848	**88.30, 0.887**	82.98 , 0.853	84.04, 0.853
22	0.37	24	88.30, 0.882	86.17, 0.859	81.91, 0.889	85.11, 0.849

In the next experiment the feature vector is further extended to include the time series-based features. Since the RF classifier exhibits the best overall results we proceed with this classifier in the remaining experiments.

4.6.4 Experiment 2c: Texture, Geometry and Time-series Features

As shown in flowchart of Figure 4.8, the purpose of this experiment is to analyse the effect of integrating border-based features (geometry and time-series) with texture features, and to investigate the classification performance. This is done by appending the 59 border features (the same 11 geometry features as per experiment 1 and 48 time series features, Section 4.3.1) to the same 11 optimal texture feature subsets listed in Table 4.7 which were previously used in Experiments 2a and 2b. The RF classifier is then applied on the test set (the second row of Table 4.1), resulting in the classification accuracies for each of the 11 feature vectors shown in Figure 4.13. The figure also shows the results of the highest accuracies obtained in Experiments 2a and 2b for easy comparison, where the highest accuracy (91.49%) is obtained by choosing the feature vector corresponding to the 0.33 threshold (the feature vector for this threshold comprises 151 texture, 11 geometry and 48 time series features). The obtained AUC value for this feature configuration is 0.938.

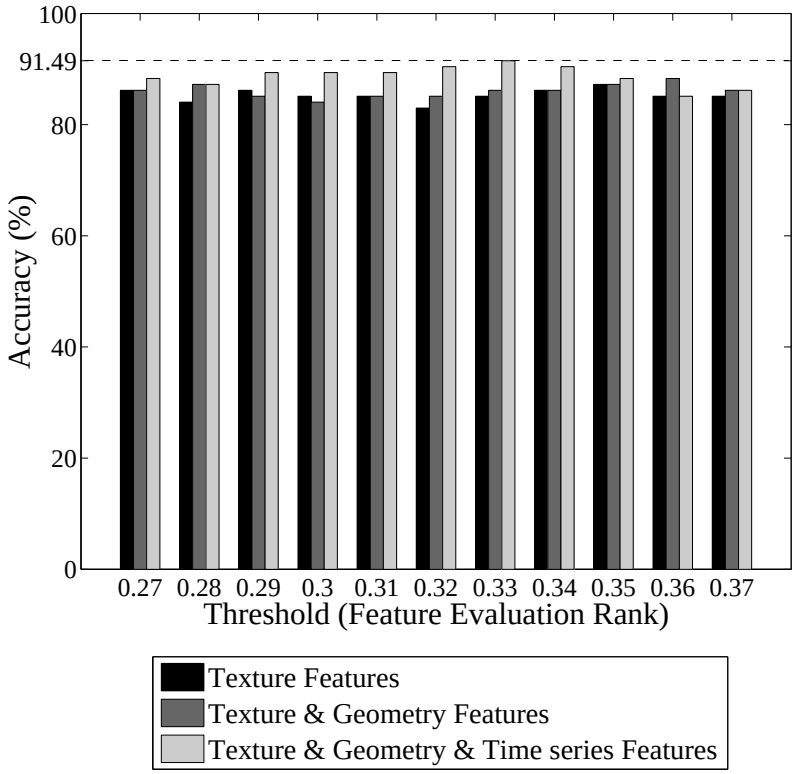

Figure 4.13: 10-fold cross validation of accuracy of texture, geometry and time series-based features at various thresholds of the Gain Ratio feature selection method, using the Random Forest classifier.

4.6.5 Experiment 3: Performance Evaluation of the Proposed Computer-aided Diagnosis of Melanoma

For this experiment the *Image-set3* from Table 4.1 is used. In order to assess in an unbiased way the performance of the developed diagnostic system, we consider three independent image sets; a train, a validation and a test set of 99, 87 and 103 dermoscopy images, respectively (See the third row of Table 4.1). As these images are in different conditions, the border detection step in two-thirds of them is done manually and the automated Hybrid border-detection method (explained in Chapter 3, Section 3.2 of this thesis) is applied on the remainder to separate the lesion from the surrounding skin. This is followed by the pre-processing step explained in Section 4.6.1 to prepare the images for performance evaluation of the diagnosis system. In order to evaluate the robustness of our system, we have elected to use both manual and automated borders. Furthermore, in order to bias the experiment towards the manual borders (worst case scenario) we have designated two-thirds of the images to manual borders.

As illustrated in Figure 4.9, we use two streams of optimisation. The first employs

Global optimisation (labelled as experiment 3a in the flowchart) wherein the whole feature vector, consisting of texture and border (geometry and time series) features is optimised. The second uses individual optimisations (labelled as experiment 3b in the flowchart) wherein the texture and border features are optimised individually and the resultant optimal features are appended. The details of these two streams are explained and the outcomes are compared next.

Global Optimisation

In this experiment the whole feature vector, consisting of texture and border (geometry and time series) is optimised, using the following three-step procedure.

(i) Feature selection: feature extraction is performed on images from the designated *train set* (the third row of Table 4.1) and for each image a vector of 35455 features are extracted, including 35396 wavelet-based texture features and 59 border features (11 geometry and 48 time series features). The extracted features are evaluated and ranked by using the Gain Ratio method. This results in various evaluation ranks in the range of [0, 0.6], each corresponding to a subset (a configuration) of features.

(ii) Parameter tuning: 26 various evaluation ranks, in the range of [0.35 , 0.6] with step size of 0.1 are empirically chosen as thresholds and the corresponding feature subsets are extracted from the images of the designated *validation set* (the third row of Table 4.1). Then, the RF classifier is applied on each feature configuration to classify the images of the validation set. In each classification process, various parameters of the classifier (number of trees) are tested and the parameter which resulted in the highest classification accuracy is selected.

Table 4.9 shows the number of features for various thresholds, the resultant 10-fold cross-validation of the classification performance (accuracy and AUC) and the corresponding classification parameter; number of trees of the RF classifier. Three feature configurations at thresholds of 0.57 (10 features), 0.44 (91 features) and 0.4 (166 features) obtain the highest accuracy of 87.23%. The best configuration out of these three is the feature subset at the threshold 0.40 with AUC value of 0.913. This optimal feature set includes 166 features and the highest accuracy and AUC values are obtained using 15 trees in the RF classifier.

(iii) Classification: the parameters (Gain Ratio, threshold and number of trees of the RF) tuned in the previous step are then used to evaluate the overall performance of the proposed diagnostic system. The optimal feature configuration at threshold of 0.4 with 166 features is extracted from the images designated as the *test set* (the third row of Table 4.1). The RF classifier with 15 trees is then applied resulting in an accuracy of 93.21% (using 10-fold cross-validation). Other performance measures of the proposed method are listed in the first row of Table 4.10.

Table 4.9: Number of features and the resultant 10-fold cross-validation of accuracy and AUC value using Random Forest classifier at various thresholds of the Gain Ratio feature selection method, applied on validation image set.

Threshold	Feature No.	No. of trees	Accuracy	AUC
0.60	2	10	66.67%	0.662
0.59	2	10	66.67%	0.662
0.58	8	50	64.37%	0.618
0.57	10	10	87.36%	0.827
0.56	12	5	86.21%	0.828
0.55	14	10	85.06%	0.829
0.54	18	5	86.21%	0.820
0.53	18	5	86.21%	0.820
0.52	23	2	86.21%	0.832
0.51	23	2	86.21%	0.832
0.50	29	10	85.06%	0.843
0.49	44	2	86.21%	0.83
0.48	51	3	86.21%	0.848
0.47	62	10	86.21%	0.872
0.46	67	90	85.06%	0.887
0.45	78	70	85.06%	0.823
0.44	91	5	87.36%	0.829
0.43	99	125	86.21%	0.891
0.42	122	55	83.91%	0.847
0.41	149	200	85.05%	0.900
0.40	**166**	**15**	**87.36%**	**0.913**
0.39	202	50	85.06%	0.887
0.38	225	220	85.06%	0.915
0.37	275	10	85.06%	0.883
0.36	302	5	86.21%	0.879
0.35	387	20	85.06%	0.886

Individual Optimisation

In this experiment the texture and border features are optimised individually and the resultant optimal features are combined. In order to do that, the first two steps of the procedure performed in the previous section (global Optimisation) are conducted independently for each of the border and texture features. This involves performing feature selection and parameter tuning using *train* and *validation image sets* (the third row of Table 4.1) to determine the optimal border features and optimal texture features. The optimal texture feature vector includes 379 features, and the optimal border feature vector includes 8 features. Then, the two optimal sets are combined yielding a total 387 optimal features, and the classifier parameter is tuned using the validation image set (number of

trees = 35). Finally the RF classifier is applied on the test image set, using the optimal feature set and the classification parameter tuned in the previous step. This results in an accuracy of 88.35% (using 10-fold cross-validation). Other performance measures of the proposed method are listed in the second row of Table 4.10.

Comparison and Observations

Table 4.10 shows the 10-fold cross-validation of performance (evaluated in terms of accuracy, sensitivity, specificity and the area under the ROC curve) of the proposed diagnostic system for the two optimisation streams; global and individual. As indicated in the table, the global optimisation performs better with respect to all of the measures. The obtained performance of our system is in fact highly comparable with the state-of-the-art melanoma diagnosis systems reported in the recent survey by Maglogiannis *et al.* [12] and outperforms some of them.

Table 4.10: 10-fold cross-validation of performance of the proposed diagnostic system, using train set of 99 images, validation set of 87 images and test set of 103 dermoscopy images, applying the two optimisation streams.

Stream	Feature No.	Accuracy	Sensitivity	Specificity	AUC
Global	166	93.21%	90.90%	94.92%	0.946
Individual	387	88.35%	88.64%	89.06%	0.939

An interesting observation of this experiment is the distribution of features in the optimal feature vector, which shows that out of 166 selected features only eight (one is geometry based and the rest are time series based) are border (4.82%) and the majority are texture-based features. Table 4.11 shows the ranking of these border-based features.

Table 4.11: Rankings of the eight optimal border features appearing in the optimal feature set.

Rank	feature	Description
9	Norm(TS)	TS: Time-series
10	Norm(WA1(TS))	WAn:Wavelet approximation component
13	Norm(WA3(TS))	level n
14	Norm(WA2(TS))	
31	Norm(Hist(TS))	Hist: Histogram
34	Eng(Hist(TS))	Eng: Energy
139	Perimeter	belongs to the Geometry category
140	Mean(Hist(TS))	

Figure 4.14 shows the Gain Ratio values of the features nominated as optimal. We have estimated the contribution of the border features in the optimal feature set by calculating the accumulative border features' Gain Ratio divided by accumulative features'

Gain Ratios (Equation 4.53). Considering the 35 highest-ranked features, which consists of six border features and 29 texture features, the approximate contribution of border-based features is 17% , however its overall contribution in the optimal set of 166 features is only 5%.

$$\text{Approximate Contribution} = \frac{\sum GainRatio_{(Border)}}{\sum GainRatio_{(All)}} \times 100 \tag{4.53}$$

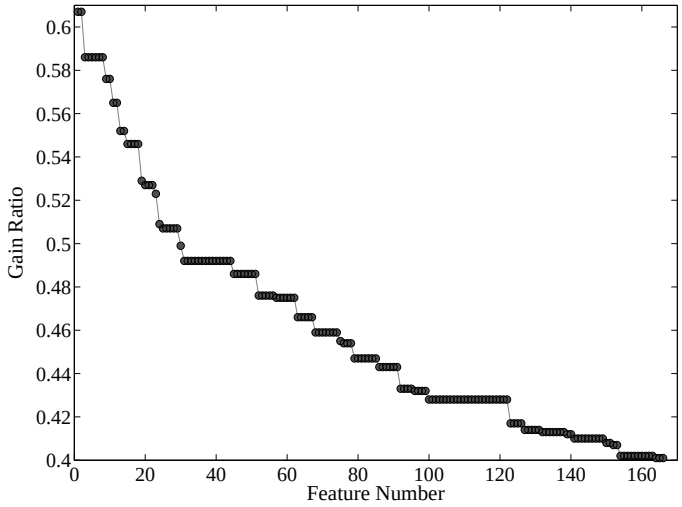

Figure 4.14: The Gain Ratio values of the features in the optimal feature subset.

According to the estimated contribution, we suggest that texture-based features play the dominant role in the classification of melanoma. This finding is in fact fully aligned with the clinical knowledge we obtained from our meetings with experienced dermatologists at the Royal Melbourne Hospital, Melbourne, Australia. We are informed that, when they examine an image, they focus on the changes in the internal structure of the lesion while its external attributes such as border and shape are considered secondary.

The 20 first-ranked texture-based features are listed in Table 4.12. The following conclusions are derived from this table: (i) six out of 20 features belong to the nodes from the forth level, which vindicates our use of four-level wavelet decomposition as opposed to the three-level decomposition applied in the other studies, (ii) texture features extracted from colour channels of Green and Blue are among the first 20 optimal texture features, which justifies our decision to incorporate the independent colour channels of Red, Green, Blue along with the Luminance, as opposed to other studies where Luminance only is used, (iii) the optimal features are derived from both low and high frequencies which justifies our choice of tree-structured decomposition over the pyramid-structured wavelet, (iv) particular measures we proposed to apply in this diagnostic system (*e.g.* Kurtosis) are prominent in the final optimal feature set.

Table 4.13 shows the advantages of applying the GRFS method in our system, which

Table 4.12: The 20 first-ranked texture-based features.

Rank	Measure	colour Channel	Level	Tree-node
1	Average-energy	Luminance	3	4.1.3
2	Average-energy	Green	3	4.1.1
3	Kurtosis	Blue	1	4
4	Average-energy	Green	2	4.1
5	Average-energy	Luminance	3	4.1.1
6	Energy	Luminance	3	4.1.1
7	Standard deviation	Luminance	3	4.1.1
8	Kurtosis	Green	1	4
11	Average-energy	Luminance	4	4.1.1.2
12	Average-energy	Luminance	2	4.1
15	Kurtosis	Blue	2	4.2
16	Standard deviation	Luminance	4	1.1.1.3
17	Energy	Luminance	4	1.1.1.3
18	Average-energy	Luminance	4	2.3.1.1
19	Average-energy	Luminance	3	4.1.2
20	Average-energy	Luminance	2	4.2
21	Kurtosis	Luminance	1	4
22	Sub-level ratio of Average-energy	Blue	4	1.1.1.1
23	Average-energy	Blue	4	4.1.1.1
24	Average-energy	Luminance	3	4.4.4

is a significant reduction (by a factor of 214) in the dimension feature space and in the time required for classification (50 times). Moreover, the classification accuracy is improved by 12%. Note that the figures listed in Table 4.13 are obtained by applying the RF classifier on the test image set, using the classifier parameters tuned over the validation image set.

Table 4.13: Effect of Gain Ratio-based feature selection method.

	Feature No.	Accuracy	Time (sec)
without feature selection	35455	81.55%	2.45
with feature selection	166	93.21%	0.05

4.7 Summary and Concluding Remarks

In this chapter we have presented a computer-aided diagnosis system for melanoma. The system covers the main diagnosis components of segmentation, feature extraction, feature selection and classification. The segmentation step was carried out both manually and automatically in various experiments, using different image sets, to investigate the ability of diagnostic system in both semi and fully automated manners.

The feature extraction process involved a guided generation of categorised features, extracted from (i) wavelet-based texture analysis, (ii) geometrical measurements, and (iii) time-series analysis in spatial and frequency domains. The texture features were extracted from application of four-level wavelet decomposition analysis on Red, Green, Blue and Luminance colour channels in addition to employing a comprehensive set of statistical measures on all the nodes of the wavelet-tree. The geometry measures were extracted from indices that are commonly used in the literature as well as novel ones defined in this study to reflect the expert opinion of dermatologists. Time-series analysis was applied on dermoscopy images for the first time in this study, wherein features were extracted by constructing a time-series model of the lesion border and analysing it in spatial and frequency domains by applying histogram analysis and wavelet transform.

A computationally efficient feature selection method, based on Gain Ratio, is adopted and demonstrated to outperform three other well-known feature selection methods, namely; Information Gain, Chi-squared and ReliefF. Moreover, a two-stage feature selection scheme was conducted which applied the Gain Ratio at first stage to reduce the feature space, then used Correlation at the second stage, on feature vector optimised at the first stage. However, experimental results showed that adding the CFS method does not provide a noticeable improvement to the classification performance over that obtained by using the GRFS method only, and the GRFS method was shown to be sufficient and efficient for the optimal feature selection. It has been shown that the use of the GFRS method provides significant improvement in the classification performance (12% increment) as well as a great reduction in the dimension of the feature vector (by a factor of 214), and the required classification computational time (50 times).

Furthermore, four advanced classes of classification techniques; namely, Support Vector Machine, Random Forest, Logistic Model Tree and Hidden Naive Bayes were employed in this study, among which the Random Forest classifier was shown to be highly effective in the classification of melanoma lesions.

Comprehensive sets of experiments were conducted in this study to analyse the effectiveness of applying various combinations of features in the classification of melanoma.

Firstly, we investigated the effectiveness of using texture-based feature extraction method (explained in Section 4.1 of this chapter) in the classification of melanoma lesions. The texture features were extracted from four-level Wavelet-decomposition in colour channels of red, green, blue and luminance. The feature selection was performed using the Gain Ratio and correlation-based feature selection methods on a train set of 103 dermoscopy images. Classification was conducted by the four classifiers of SVM, RF, HNB, and LMT. When an optimal subset of texture features was extracted from a test set of 102 dermoscopy images, LMT was able to successfully differentiate between melanoma and benign lesions with 88.24% accuracy, followed by the RF classifiers with an accuracy of 87.25%. Moreover, the optimal subset of texture features outperformed three other wavelet-based approaches [171, 210, 169].

Second, a ternary set of experiments were conducted wherein the effectiveness of three sets of features extracted from (i) texture analysis, (ii) texture and geometrical

measurements analysis, and (iii) texture, geometrical and time-series was investigated, and the results were compared. These experiments were performed on the same image set (197 dermoscopy images including 98 malignant and 109 benign lesions) in order to provide a valid base for comparison. 103 of the images were designated as train set for the purpose of feature selection and the remaining 94 images were designated as test set to test the classification performance. Experimental results showed that the in the first experiment where the texture-based features were analysed the resultant classification accuracy was 87.23%. In other two experiments with the extended feature vectors, where the geometry-based features and then the time series-based features were appended to the optimal texture-based features, improvements in the classification accuracy (88.30% and 91.49%) were obtained. These results have demonstrated the superiority of incorporating border (geometry and time-series) with texture features over using exclusively texture.

In the last experiment the overall performance of the proposed system was evaluated by conducting two streams of optimisation on a set of 289 dermoscopy images (114 malignant and 175 benign) subdivided into three image sets; train, validation and test. Two optimisation streams were studies: (i) the global optimisation, in which the whole feature vector, consisting of texture and border features is optimised, (ii) Individual optimisation wherein the texture and border features are optimised individually and the resultant optimal features are integrated. The experimental results showed that by applying the Gain Ratio feature selection method and the Random Forest classifier, global optimisation outperforms individual optimisation, with 93.21% accuracy, 90.90% sensitivity and 94.92% specificity. Although the results cited above were obtained through experiments conducted on a particular image set, they indicate that high diagnostic accuracy can be achieved when existing features are optimally blended with novel ones. In fact these results compare quite well with, and in most of the cases superior to the state-of-art melanoma diagnosis systems reported in the literature [12].

An important finding is the clear advantage gained in optimally complementing texture with border and geometry features compared to using texture information only. Another interesting finding of this study is that in the melanoma classification application texture features play the dominant role and make much higher contribution compared to border-based features. Among border features, which include geometry and time series features, the latter has been shown to be more informative and obtain higher evaluation ranks (Gain Ratio values) than geometry in this application.

The proposed feature extraction and classification method in this chapter is published in [232] and an extended version has been submitted to IEEE Transactions on Information Technology in Biomedicine [233].

Chapter 5

Summary and Conclusion

Malignant melanoma is the deadliest form of skin cancer and its early diagnosis is of extreme importance for public health and the basis for reduction of mortality rates. Despite advancements in imaging technology such as dermoscopy, clinical diagnosis of melanoma is challenging and suffers from inherent subjectivity, particularly for primary care physicians – who are the initial contact points (gateway) for patients to the dermatology community. As a result, computer-based melanoma diagnosis has become a major research area in recent years with the aim of assisting physicians through the provision of quantitative reproducible analysis of the skin lesion. Computer-based diagnosis systems for melanoma have been acknowledged by physicians to play a useful role in producing quantitative results, recording patient follow-ups, and monitoring remedial progress. However, their adoption in daily practice is still a work in progress. For that to happen, physicians need higher accuracy and reliability than what has so far been offered.

With the aim of improving some of existing methods and developing new techniques to facilitate accurate, fast and reliable computer-based diagnosis of melanoma, this thesis presented new and effective approaches to development of an integrated computer-aided diagnostic system of melanoma. A computer-aided diagnosis of melanoma involves several components including image acquisition, image segmentation or border detection, feature extraction, feature selection, and classification. In this chapter we conclude the thesis by explaining the contributions of this study with respect to each component of the computer-based melanoma diagnostic system. This is followed by some discussions on possible directions for future work on this topic.

5.1 Main Contributions

In the following a summary of the main contributions of this thesis in the area of computer-based diagnostic system of melanoma is provided.

5.1.1 Colour-channel optimisation for border-detection in dermoscopy images.

Colour information plays a significant role in dermoscopy image analysis. Variegation of colour within the lesion and the diversity found in human skin colour makes border detection a challenging process. Although numerous skin lesion segmentation methods have been developed, few studies have focused on determining the most effective colour space and colour channel for melanoma application. In this study, we have proposed a comprehensive colour optimisation procedure, which (i) transforms dermoscopy images into different colour channels from various colour spaces (as illustrated in Table 3.1), (ii) applies histogram thresholding on the transformed image, (iii) uses mathematical morphology and connected component analysis to detect the border of the lesion and separate the lesion from background skin, (vi) performs a quantitative comparison between each resulting border and the corresponding ground-truth drawn by four dermatologists. Applying statistical and ROC analysis on the resulting borders in this thesis, demonstrated that the optimal colour channels for border-detection in dermoscopy images are X colour channel from the XYZ colour space, and XoYoR which integrates the segmentation results from X and Y colour channels (from the XYZ colour space) with the R colour channel (from the RGB colour space). The optimal colour channel(s) determined via the above-mentioned optimisation process are the main building block for the border-detection methods proposed in this study, and provide a remarkable added value to the final segmentation result.

5.1.2 Automatic border detection in dermoscopy images: Global and Hybrid methods.

Segmentation or border detection – which separates the lesion from the background skin – is a fundamental step towards development of a computer-aided diagnosis of melanoma and its accuracy is essential for accurate implementation of the subsequent steps. In this thesis, we have proposed two automated methods for detecting the borders in dermoscopy images; Global and Hybrid method, the latter being an enhancement of the former.

The proposed Global border-detection method applies Otsu's global histogram thresholding on the optimal colour channels derived from the above-mentioned colour optimisation process. This is followed by application of connected component analysis and morphological operation which resulted in automatic lesion borders with an accuracy of approximately 97%. Comparative study involving two state-of-the-art automated methods demonstrated that with a proper choice of colour channels, the proposed thresholding-based method in spite of its simplicity is highly competitive with well-known skin lesion segmentation methods, and outperforms them with respect to accuracy, specificity, and AUC metrics (as shown in Table 3.6). The proposed method is potentially faster since it mainly involves scalar processing as opposed to vector processing performed in those methods. The proposed Global border-detection method along with the colour optimisation process is published in [205, 206].

Considering the practical nature of melanoma diagnosis which calls for extreme caution when excluding portions of the image, dermatologists prefer to choose the outmost detectable pigment to minimise the risk of incorrect diagnosis. In contrast, existing border detection methods generally fail to accurately detect the marginal areas of the lesion, which are faded into the background skin, and mistakenly exclude it from the segmentation result. In order to address this issue, we have improved the above-mentioned Global method into the Hybrid border-detection method and have shown that it reduces the discrepancy between the manual and automatic borders to a great extent (Table 3.11). The Hybrid method first applies the Global method on the optimal colour channel of XoYoR (derived from the colour optimisation process) to detect an initial boundary for the lesion, then integrates the local colour information around the initial boundary to expand the border toward the background skin, and form the final segmentation result. 85 high-resolution dermoscopy images were analysed and the automatic borders were compared to borders drawn by two dermatologists using various quantitative evaluation metrics including accuracy, precision, sensitivity, specificity, and border error. The proposed Hybrid method, which is comprised of two stages, is designed to increase specificity in the first stage (global thresholding) and sensitivity in the second stage (adaptive local thresholding). The proposed method is shown to be highly effective in detecting the borders of the lesion with an accuracy of 98%. Moreover, it is demonstrated that the Hybrid method is more effective than a less-experienced dermatologist and functions closer to the experienced dermatologist (Figure 3.22). The proposed border detection method has been compared to three state-of-the-art border detection methods and found to be superior than (with respect to some of the evaluation metrics) or highly competitive with those methods (with respect to other metrics) as shown in Tables 3.7 and 3.8. The proposed Hybrid border-detection method is published in [204].

We have also provided an analytical framework, including statistical analysis (Figures 3.25– 3.27), optimisation (Figure 3.28), cross validation (Figure 3.29 and Table 3.10) to determine the optimal parameters for the proposed Hybrid border detection method and vindicated the results obtained by the Hybrid method. The optimisation is done using a newly introduced evaluation metric; Performance Index (PI), which is composed of standard metrics of sensitivity, specificity, accuracy, similarity, and border error. Moreover, the two proposed border detection methods, Global and Hybrid, were compared and the superiority of the Hybrid method was demonstrated objectively through the proposed PI (Table 3.11). The proposed systematic procedure for comparison between the two proposed border-detection methods is published in [207].

5.1.3 Objective evaluation of border-detection method in dermoscopy images.

Prior to our reported metrics, no systematic evaluation procedure for the performance assessment of the border-detection methods in dermoscopy images, which could reflect dermatologists' perspective in the melanoma recognition process, had been thoroughly investigated in the literature. Also, there is a need for a comprehensive and practical

assessment procedure, because in using the standard metrics of sensitivity, specificity, accuracy, border error, similarity, and precision, some border-detection methods show superior results to other method with respect to some metrics but inferior with respect to the other metrics. A problem therefore arises in determining which method yields better overall results. In this thesis two quantitative metrics of *Weighted Performance Index* and *Optimised Weighted Performance Index* have been proposed as useful means for assessing performance of a given border detection method, and also facilitate comparison between different border detection methods. The manner in which dermatologists expertise is incorporated into the standard metrics is briefly outlined in the following.

All six standard metrics of sensitivity, specificity, accuracy, border error, similarity and precision are functions of four measures of True Positive, False Positive, True Negative and False Negative. In order to provide evaluation metrics meaningful in the context of melanoma application and at the same time be easily computable, we introduced specific weights, drawing from the expertise of the dermatologists, into the standard metrics by attaching weights to these four measures. The Weighted Performance Index is defined as a weighted average of the six metrics, wherein the weights are imbedded within each metric. Therefore, the proposed evaluation index has two main objectives: (i) it is customised for melanoma diagnosis application as it accommodates dermatologists's diagnostic perspectives, and (ii) it takes into account various evaluation metrics and provides a comprehensive yet easy means for comparison between different border-detection methods.

The effectiveness of the proposed evaluation index was demonstrated by applying five recent border detection methods on a set of 55 high-resolution dermoscopy images using the union of four sets of dermatologist-drawn borders as the ground truth (shown in Table 3.13). It was also shown that the weighted metrics, which are defined to reflect the dermatologists' perspectives, show a higher degree of agreement between automatic and manual borders, compared with standard metrics . Objective evaluation of border detection methods in dermoscopy images via Weighted Performance Index is published in [208].

The proposed evaluation metric WPI is later improved by proposing Optimised Weighted Performance Index. In contrast to WPI wherein the weights were fixed and based on expert opinion only, in OWPI the optimal weight for each measure of TP, FP, TN and FN was obtained by formulating a constrained nonlinear multivariable optimisation problem. The optimisation is done in order to provide weights which genuinely reflect the importance of each metric in the evaluation process. The optimal weights were calculated by applying trust-region and interior-point optimisation methods on five recent border detection methods on a set of 55 high resolution dermoscopy images, using the union of four sets of dermatologist-drawn borders as the ground truth. The automated borders drawn by all five border-detection methods have been used to eliminate any bias towards any particular border-detection method and provide a fair basis for the comparison as done in Table 3.21. Experimental results showed that the optimisation performed on all five border-detection methods resulted in similar set of weights (Figure 3.37 and Tables 3.15–3.19) which vindicates the robustness of our proposed procedure. Moreover, the resulting

optimised weights were confirmed by a grid-search procedure.

Comparison between Performance Index, Weighted Performance Index and Optimised Weighted Performance Index shows the advantage of the latter. OWPI yields higher performance values for each of the five border detection methods, followed by WPI and PI comes last (shown in Figure 3.46). Moreover, OWPI with optimal weights, tailored for the melanoma application yields smaller differences between the border detection methods, implying the actual performance differences to be in fact smaller than those predicted by WPI and PI. As shown in Figure 3.46 all three metrics of PI, WPI and OWPI coincide with the superiority of the proposed Hybrid border-detection method over the four other border-detection methods. Objective evaluation of border detection methods in dermoscopy images via Optimised Weighted Performance Index has been submitted to IEEE Transactions on Information Technology in Biomedicine in July 2010, has received very good feedback from the review committee and the revision was submitted to the journal in February 2011 [209].

5.1.4 Feature extraction in dermoscopy images using wavelet, geometry and time-series analysis.

In a computer-based melanoma diagnosis system, feature extraction is applied to extract the features that accurately characterise a melanoma lesion. These features, which are extracted from the ROI determined through the segmentation (border-detection) process, are similar to those visually detected by dermatologists in their clinical practice. As a result of studying existing clinical diagnostic approaches such as the ABCD rule of dermoscopy and pattern analysis as well as consultation with expert dermatologists from Royal Melbourne Hospital, we have developed a novel, comprehensive, and highly effective feature extraction method which combines different types of features. The proposed method captures (i) the changes in the internal texture of the lesion, and (ii) the external attributes of the lesion such as border and shape, and combines the resulting features.

The multi-scale properties of wavelets, makes wavelet-based texture analysis a useful technique for analysing dermoscopy images, as the dermoscopy images are often taken under different circumstances, *i.e.*, various image acquisition set-ups (lighting, optical zooming, etc) and diverse skin and lesion colours of the population. Among the two wavelet structures; Pyramid-structured and Tree-structured wavelet analysis, the latter was found to be more informative for classification of melanoma images, as it can capture information of both general/low-frequency (overall shape) and detail/high-frequency (internal texture) of the lesion. For this reason, the proposed texture-based feature extraction method employs tree-structured wavelet decomposition. The novelty of the texture-based feature extraction method proposed in this study lies in (i) the use of four-level decompositions, (ii) the use of four colour channels of Red, Green, Blue and Luminance, (iii) applying various statistical measurements and ratios on all nodes of the wavelet-tree. It is worth mentioning that the existing methods have used only one of the three (we will further complement the feature extraction method by integrating border features to establish an

umbrella feature extraction method).

Experimental results demonstrated that features contained in the fourth level of the wavelet tree, those extracted from the Blue and Green colour channels, those derived from the high frequency channels of the wavelet tree, as well as particular measures we proposed to apply in this diagnostic system (*e.g.* Kurtosis) are prominent in the final optimal feature set (see Table 4.12). These observations demonstrate the advantage of applying the four-level wavelet decomposition as opposed to three levels used in existing wavelet-based methods, incorporating various colour channels into the texture analysis as opposed to only adopting the commonly used Luminance, applying tree-structures as opposed to pyramid-structured in some of the studies, and applying various statistical measures and ratios, for classification of melanoma. The effectiveness of the proposed texture-based features and its superiority over three recent wavelet-based methods is published in [232].

In order to measure border-based features (i) we have applied geometrical measurements, such as area, perimeter, circularity index, irregularity index, etc. to analyse the overall shape of the lesion. In doing so, we have used standard geometry features existing in the literature complemented by other new ones, drawing from our discussions with expert dermatologists, and (ii) we have constructed a time-series model of the border and analysed it in both spatial and frequency domains. Time-series has been widely applied in statistics, signal processing, data mining, mathematical finance and economy. To the best of our knowledge it has not been previously applied to the application of melanoma diagnosis. In this thesis we have applied histogram analysis and wavelet transform on the time series constructed from the lesion boundary to extract the main characteristics of the border of the lesion. Experimental results carried out in this thesis show that the proposed time series-based features have higher contribution in classification of melanoma, than the geometry-based features used in the previous computer-based melanoma studies (see Table 4.11).

5.1.5 Selection and optimal integration of features and classification of dermoscopy images.

Feature selection is the process of eliminating irrelevant, redundant, or noisy features from the initial feature vector constructed in the feature extraction step. In this study the Gain Ratio feature selection method, which is an entropy-based feature valuator, is employed for the following reasons: (i) the method is known to be computationally efficient, which is a vital factor because of the high dimensionality of the initial feature vector constructed in the feature extraction phase; (ii) it outperforms a sample of well-know filter-based feature selectors such as Information Gain, Chi-squared and ReliefF (illustrated in Figure 4.6). Gain Ratio has been successfully applied in various applications, yet to the best of our knowledge it has not been applied in previous melanoma-diagnosis studies. This study establishes that the Gain Ratio feature selection method can deliver substantial improvement to system performance compared to using raw data (the whole of the constructed feature vector). As shown in Table 4.13, it yields a 12% increment in

the classification accuracy, as well as a considerable reduction in feature space dimension (by a factor of 214) and computation time (50 times).

Classification is the final step in the diagnosis process, wherein the extracted features are utilised to ascertain whether the lesion is malignant or benign. In this study, four advanced classes of classification techniques are employed; Support Vector Machine has been very popular and shown to be efficient in computer-based melanoma diagnosis studies in the literature. We have elected to complement this choice by the use of the three other classifiers; namely, Random Forrest, Logistic Model Tree and Hidden Naive Bayes which are applied to the melanoma classification problem for the first time in this thesis. All these classifiers and more particularly the Random Forest classifier were shown to be highly effective in the classification of melanoma lesions.

Comprehensive sets of experiments were conducted in this thesis to analyse the effectiveness of applying various combinations of features in the classification of melanoma. These experiments cover the main diagnosis components of segmentation (border detection), feature extraction, feature selection and classification. In order to evaluate the robustness of our system, we have elected to use both manual (under dermatologists' guidance) and automated (by applying the Hybrid border-detection method proposed in the first part of this thesis) borders.

First, we investigated the effectiveness of using wavelet-based texture features in the classification of melanoma lesions. Feature extraction and a two-stage feature selection method, based on Gain Ratio and Correlation, were applied to a train set of 103 images. The resultant feature subsets were then fed into four different classifiers: SVM, RF, HNB, LMT to classify melanoma in a test set of 102 images, which resulted in an accuracy of 88.24% and AUC value of 0.918. Comparative study carried out in this thesis has shown that the proposed feature extraction method outperforms three other wavelet-based approaches (Table 4.6 and Figure 4.12).

Second, a ternary set of experiments were conducted (see Figure 4.8) wherein the effectiveness of three sets of features extracted from (i) texture analysis, (ii) texture and geometrical measurements analysis, and (iii) texture, geometrical and time-series were investigated, and the results were compared. These experiments were performed on the same image set (197 dermoscopy images including 98 malignant and 109 benign lesions) in order to provide a valid base for comparison. As illustrated in Figure 4.13, experimental results showed improvement in classification performance upon extending the feature vector and appending the border-based features to the optimal texture-based feature vector; from 87.23% (using texture only) to 88.30% (adding the geometry features) and 91.49% (appending the time-series features). These results demonstrate the clear advantage gained in optimally complementing texture with border and geometry features compared to using texture information only.

Finally, we evaluated the overall performance of the proposed system by proposing two streams of optimisation and comparing the results of the two streams (Figure 4.9). The experiment was carried out on a set of 289 dermoscopy images (114 malignant and 175 benign) subdivided into three sets; train, validation and test, in order to evaluate the

system's performance in an unbiased way. Two optimisation streams were studied: (i) global optimisation, in which the whole feature vector, consisting of texture and border features is optimised, (ii) individual optimisation wherein the texture and border features are optimised individually and the resultant optimal features are integrated. The experimental results showed that by applying the Gain Ratio feature selection method and the Random Forrest classifier, global optimisation outperforms individual optimisation, with 93.21% accuracy, 90.90% sensitivity and 94.92% specificity (Table 4.10). Although the results cited here were obtained through experiments conducted on a particular image set, they indicate that high diagnostic accuracy can be achieved when existing features are optimally blended with novel ones. In fact these results compare quite well with, and in most cases are superior to state-of-art melanoma diagnosis systems reported in the literature [12].

Another interesting finding of this study is that in the melanoma classification application texture features play the dominant role and make much higher contribution than border-based features. Among border features, which include geometry and time-series features, the latter has been shown to be more informative and obtain higher evaluation ranks (Gain Ratio values) than geometry in this application.

The proposed computer-based diagnostic system [233] has been submitted to IEEE Transactions on Information Technology in Biomedicine.

5.2 Directions for Future Work

Throughout the thesis we have presented new ideas and approaches to address the key components of a computer-based system for diagnosis of melanoma. We now canvass several possible directions that can be pursued to further advance the state of this project.

In order to develop a fully automated computer-aided diagnosis system of melanoma, which can reliably operate in real-world and be convincing enough for people (including dermatologists and patients) to rely upon, it is required that the system be rigorously tested on a large number of images with different conditions, *i.e.*, on thousands of lesions acquired from people with different skin colours, with images taken under various lighting conditions. This would necessitate a well-organised participation of various medical institutes.

The automated melanoma diagnosis system should be tailored to suit the needs of: (i) expert users, including experienced dermatologists, who may seek a quantitative analysis of the given lesion as a second opinion to their clinical diagnosis, and (ii) non-expert users. including dermatology registrars or general practitioners, who need to improve the accuracy of their diagnosis and gain more confidence. The diagnosis system should be designed to serve as an educational tool for medical students during the course of their studies. The ultimate target of this software is people/patients who should be able to perform a preliminary analysis of their skin at home, and according to the results would seek medical advice from medical centres for further investigations. This would be very

useful for people who need regular follow-ups, thus saving time and cost of multiple travels, particularly if they live in rural areas.

In order to accommodate certain requirements for each group of users, various user interfaces need to be designed for the software system. For example, for an expert user inter-active software is desirable, where the expert dermatologist can conduct the quantitative analysis for different input parameters (for example the *window-size* parameter in the border detection method according to the lighting and colouring condition of the lesion under assessment) and investigate the results. On the other hand, for non-expert users fully-automated software with high sensitivity and reasonable specificity is required.

An extensive image-acquisition procedure should be established, wherein various imaging modalities are used to generate multiple images of the same lesion, each conveying certain aspect of the lesion (*e.g.* ultra sound to extract information about the depth of melanoma, dermoscopy to provide detailed inspection of the surface of the lesion, etc.). This information can be then combined in order to provide a more accurate computer-based analysis of the given lesion. Moreover, for patients with suspicious lesions a series of images can be taken over a period of time in order to provide a follow-up quantitative analysis of the lesions. Furthermore, having a large set of images with diverse pathological distribution of lesions can lead to "multi-level" or "multi-class" classification of the lesions, where instead of classifying a lesion as either malignant or benign it is classified according to its level of malignancy or its histopathology. This will allow us to (i) detect different stages of melanoma (ii) differentiate between dysplastic nevi (which is often mistaken for melanoma), benign nevi, and melanoma (iii) detect various types of benign lesions.

Further improvements along this line of research include extending the colour optimisation procedure by investigating other colour spaces and examining the impact on the border detection results. Moreover, for objective evaluation of border-detection methods, other metrics can be incorporated in the performance index. The feature extraction technique can also be further advanced by quantifying internal texture and colour symmetry of the lesion, or applying supervised pattern recognition techniques to detect certain differential structures such as globules, dots, radial streaming, etc.

Bibliography

[1] H. Zhang, L. Jiang, and J. Su. Hidden naive bayes. In *Twentieth National Conference on Artificial Intelligence*, pages 919–924, 2005.

[2] L. Jiang, H. Zhang, and Z. Cai. A novel bayes model: Hidden naive bayes. *IEEE Transactions on Knowledge and Data Engineering*, 21:1361–1371, 2009.

[3] H. Liu and L. Yu. Toward integrating feature selection algorithms for classification and clustering. *IEEE Transactions on Knowledge and Data Engineering*, 17:491–502, 2005.

[4] American cancer society, cancer facts and figures 2010. Available at: http://www.cancer.org/Research/CancerFactsFigures/CancerFactsFigures/cancer_-facts-and-figures-2010. Accessed April 2011.

[5] Australia skin cancer facts and figures. Available at: http://www.cancer.org.au/cancersmartlifestyle/SunSmart/Skincancerfactsandfigur_es.htm. Accessed September 2009.

[6] Australia skin cancer facts and figures. Available at: http://www.cancer.org.au/Healthprofessionals/cancertypes/melanoma.htm. Accessed September 2009.

[7] A. C. Geller, S. M. Swetter, K. Brooks, M. Demierre, and A. L. Yaroch. Screening, early detection, and trends for melanoma: Current status (2000-2006) and future directions. *Journal of the American Academy of Dermatology*, 57:555–572, 2007.

[8] R. P. Braun, L. E. French, and J. H. Saurat. Dermoscopy of pigmented lesions: a valuable tool in the diagnosis of melanoma. *Swiss Medical Weekly*, 134:83–90, 2004.

[9] C. M. Balch, A. C. Buzaid, S. J. Soong, N. Cascinelli M. B. Atkins, D. G. Coit, I. D. Fleming, J. E. Gershenwald, A. J. Houghton, J. M. Kirkwood, K. M. McMasters, M. F. Mihm, D. L. Morton, D. S. Reintgen, M. I. Ross, A. Sober, J. A. Thompson, and J. F. Thompson. Final version of the american joint committee on cancer staging system for cutaneous melanoma. *Journal of Clinical Oncology*, 19:3635–3648, 2001.

[10] C. M. Balch, S. J. Soong, J. E. Gershenwald, J. F. Thompson, D. S. Reintgen, , and N. Cascinelli et al. Prognostic factors analysis of 17,600 melanoma patients:

validation of the american joint committee on cancer melanoma staging system. *Journal of Clinical Oncology*, 19:3622–3634, 2001.

[11] R. Braun, H. Rabinovitz, M. Oliviero, A. Kopf, and J. Saurat. Dermoscopy of pigmented lesions. *Journal of the American Academy of Dermatology*, 52(1):109–121, 2005.

[12] I. Maglogiannis and C. Doukas. Overview of advanced computer vision systems for skin lesions characterisation. *IEEE Transaction on Information Technology in Biomedicine*, 13(5):721–733, 2009.

[13] R. J. Pariser and D. M. Pariser. Primary care physicians' errors in handling cutaneous disorders:: A prospective survey. *Journal of the American Academy of Dermatology*, 17:239–245, 1987.

[14] H. Pehamberger, A. Steiner, and K. Wolff. In vivo epiluminescence microscopy of pigmented skin lesions. I: Pattern analysis of pigmented skin lesions. *Journal of the American Academy of Dermatology*, 17:571–583, 1987.

[15] W. Stolz, A. Riemann, A. B. Cognetta, L. Pillet, W. Abmayr, D. Holzel, P. Bilek, F. Nachbar, M. Landthaler, and O. Braun-Falco. ABCD rule of dermatoscopy: A new practical method for early recognition of malignant melanoma. *European Journal of Dermatology*, 4:521–527, 1994.

[16] S. W. Menzies, C. Ingvar, and W. H. McCarthy. A sensitivity and specificity analysis of the surface microscopy features of invasive melanoma. *Melanoma Research*, 6(1):55–62, February 1996.

[17] G. Argenziano, G. Fabbrocini, P. Carli, V. De Giorgi, E. Sammarco, and M. Delfino. Epiluminescence microscopy for the diagnosis of doubtful melanocytic skin lesions: Comparison of the ABCD rule of dermatoscopy and a new 7-point checklist based on pattern analysis. *Archives of Dermatology*, 134:1563–1570, 1998.

[18] J. Henning, S. Dusza, S. Wang, A. Marghoob, H. Rabinovitz, D. Polsky, and A. Kopf. The CASH (colour, architecture, symmetry, and homogeneity) algorithm for dermoscopy. *Journal of the American Academy of Dermatology*, 56(1):45–52, 2007.

[19] M. Binder, M. Schwarz, A. Winkler, A. Steiner, A. Kaider, K. Wolff, and h. Pehamberger. Epiluminescence microscopy. A useful tool for the diagnosis of pigmented skin lesions for formally trained dermatologists. *Archives of Dermatology*, 131(3):286–291, 1995.

[20] G. Argenziano, H. P. Soyer, S. Chimenti, R. Talamini, R. Corona, F. Sera, M. Binder, L. Cerroni, G. De Rosa, G. Ferrara, R. Hofmann-Wellenhof, M. Landthaler, S. W. Menzies, H. Pehamberger, D. Piccolo, H. S. Rabinovitz, R. Schiffner, S. Staibano, W. Stolz, I. Bartenjev, A. Blum, R. Braun, H. Cabo, P. Carli, V. De Giorgi, M. G. Fleming, J. M. Grichnik, C. M. Grin, A. C. Halpern, R. Johr, B. Katz, R. O.

Kenet, H. Kittler, J. Kreusch, J. Malvehy, G. Mazzocchetti, M. Oliviero, F. Ozdemir, K. Peris, R. Perotti, A. Perusquia, M. A. Pizzichetta, S. Puig, B. Rao, P. Rubegni, T. Saida, M. Scalvenzi, S. Seidenari, I. Stanganelli, M. Tanaka, K. Westerhoff, I. H. Wolf, O. Braun-Falco, H. Kerl, T. Nishikawa, K. Wolff, and A. W. Kopf. Dermoscopy of pigmented skin lesions: Results of a consensus meeting via the Internet. *Journal of the American Academy of Dermatology*, 48:679–693, 2003.

[21] W. Stolz, O. Braun-Falco, P. Bilek, M. Landthaler, W. H. C. Burgdorf, and A. B. Cognetta. *colour Atlas of Dermatoscopy*, volume 1. Blackwell, Berlin, 2 edition, March 2002.

[22] S. Gilmore, R. Hofmann-Wellenhof, J. Muir, and H P. Soyer. Lacunarity analysis: A promising method for the automated assessment of melanocytic naevi and melanoma. *Public Library of Science*, 4:1–10, October 2009.

[23] W. C. Lambert, A. Lapidus, and B. K. Rao. Melanoma diagnosis by computerised analysis of clinical images. *Archives of Dermatology*, 3:377–378, 2001.

[24] J. Boldrick, C. Layton, J. Nguyen, and S. Swetter. Evaluation of digital dermoscopy in a pigmented lesion clinic: Clinician versus computer assessment of malignancy risk. *Journal of the American Academy of Dermatology*, 56(3):417–421, 2007.

[25] A. Perrinaud, O. Gaide, L. French, J. Saurat, A. Marghoob, and R. Braun. Can automated dermoscopy image analysis instruments provide added benefit for the dermatologist? A study comparing the results of three systems. *British Journal of Dermatology*, 157:926–933, 2007.

[26] H. Iyatomi, H. Oka, M. Saito, A. Miyake, M. Kimoto, J. Yamagami, S. Kobayashi, A. Tanikawa, M. Hagiwara, K. Ogawa, G. Argenziano, H. P. Soyer, and M. Tanaka. Quantitative assessment of tumour extraction from dermoscopy images and evaluation of computer-based extraction methods for an automatic melanoma diagnostic system. *Melanoma Research*, 16(2):183–190, 2006.

[27] G. R. Day and R. H. Barbour. Automated melanoma diagnosis: Where are we at? *Skin Research and Technology*, 6:1–5, 2000.

[28] B. Rosado, S. W. Menzies, A. Harbauer, H. Pehamberger, K. Wolff, M. Binder, and H. Kittler. Accuracy of computer diagnosis of melanoma: A quantitative meta-analysis. *Archives of Dermatology*, 139:361–367, 2003.

[29] G. Argenziano, H. P. Soyer, and V. De Giorgi. *Interactive Atlas of Dermoscopy*. EDRA Medical Publishing and New Media, Milan, Italy, 2002.

[30] S. W. Menzies. Cutaneous melanoma: Making a clinical diagnosis, present and future. *Dermatologic Therapy*, 19:32–39, 2006.

[31] W. Stoecker, R. Moss, J. Stanley, X. Chen, K. Gupta, R. Narayana, B. Shrestha, and P. Jella. Automatic detection of critical dermoscopy features for malignant melanoma diagnosis, 2006. Patent Number: 11/421031, United States.

[32] M. Marozas and R. Jurkonis. Review on skin lesion imaging, analysis and automatic classification. In *Biomedical Engineering*, pages 109–114, Biomedical Engineering Institute, Kaunas University of Technology, Lithuania.

[33] I. M. Freedberg, A. Z. Eisen, K. Wolff, K. F. Austen, L. A. Goldsmith, S. I. Katz, and T. B. Fitzpatrick. *Fitzpatrick's Dermatology in General Medicine*, volume 1. McGraw-Hill, 5 edition, January 1999.

[34] Seer's web-based training modules. Available at: http://training.seer.cancer.gov/ss_module14_melanoma/unit01_sec01_intro.html. Accessed August 31, 2008.

[35] J. F. Thompson, R. A. Scolyer, and R. F. Kefford. Cutaneous melanoma. *The Lancet*, 365:687–701, 2005.

[36] Automated melanoma recognition. Available at: http://www.advancedmcode.org/automated-melanoma-recognition.html.

[37] S. B. Forman, T. C. Ferringer, S. J. Peckham, S. R. Dalton, G. T. Sasaki, L. F. Libow, and D. M. Elston. Is superficial spreading melanoma still the most common form of malignant melanoma? *Journal of the American Academy of Dermatology*, 58:1013–1020, 2008.

[38] E. Claridge, S. Cotton, P. Hall, and M. Moncrieff. From colour to tissue histology: Physics based interpretation of images of pigmented skin lesions. In *Medical Image Computing and Computer-Assisted Intervention*, volume 2488 of *Lecture Notes in Computer Science*, pages 730–738. Springer Berlin, 2002.

[39] I. H. Wolf, J. S., H. P. Soyer, and H. Kerl. Sensitivity in the clinical diagnosis of malignant melanoma. *Melanoma Research*, 8:425–429, 1998.

[40] J. M. Grant-Kels, E. T. Bason, and C. M. Grin. The misdiagnosis of malignant melanoma. *Journal of the American Academy of Dermatology*, 40:539–548, 1999.

[41] H. S. Rabinovitz R. P. Braun, M. Oliviero, A. W. Kopf, J. Saurat, and L. Thomas. Dermoscopic examination. In *colour Atlas of Melanocytic Lesions of the Skin*, chapter 1.3, pages 7–2. Springer, 2007.

[42] Dermoscopy: A practical guide. DVD-ROM. American Academy of Dermatology.

[43] G. Annessi, R. Bono, F. Sampogna, T. Faraggiana, and D. Abeni. Sensitivity, specificity, and diagnostic accuracy of three dermoscopic algorithmic methods in the diagnosis of doubtful melanocytic lesions: The importance of light brown structureless areas in differentiating atypical melanocytic nevi from thin melanomas. *Journal of the American Academy of Dermatology*, 56(5):759–767, 2007.

[44] 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. *A Cancer Journal for Clinicians*, 35:130–151, 1985.

[45] E. M. Fernandez and K. F. Helm. The diameter of melanomas. *Dermatologic Surgery*, 30(9):1219–1222, September 2004.

[46] N. Abbasi, H. Shaw, D. Rigel, R. Friedman, W. McCarthy, I. Osman, A. Kopf, and D. Polsky. Early diagnosis of cutaneous melanoma. revisiting the ABCD criteria. *Journal of the Americal Medical Association (JAMA)*, 292:2771–2776, 2004.

[47] V. Skladnev, A. Gutenev, Scott W. Menzies, L. Bischof, G. Talbot, E. Breen, and M. Buckley. Diagnostic feature extraction in dermatological examination, 2004. Patent Number: 10/478078, United States.

[48] S. W. Menzies, L. Bischof, H. Talbot, A. Gutenev, M. Avramidis, L. Wong, S. K. Lo, G. Mackellar, V. Skladnev, W. McCarthy, J. Kelly, B. Cranney, P. Lye, H. Rabinovitz, M. Oliviero, A. Blum, A. Varol, B. De'Ambrosis, R. McCleod, H. Koga, C. Grin, R. Braun, and R. Johr. The performance of solarscan. An automated dermoscopy image analysis instrument for the diagnosis of primary melanoma. *Archives of Dermatology*, 141:1388–1397, 2005.

[49] M. J. Jamora, B. D. Wainwright, S. A. Meehan, and J. C. Bystryn. Improved identification of potentially dangerous pigmented skin lesions by computerised image analysis. *Archives of Dermatology*, 139:195–198, 2003.

[50] P. Rubegni, M. Burroni, G. Cevenini, R. Perotti, G. Dell'Eva, P. Barbini, M. Fimiani, and L. Andreassi. Digital dermoscopy analysis and artificial neural network for the differentiation of clinically atypical pigmented skin lesions: A retrospective study. *The journal of investigative dermatology*, 119:471–474, 2002.

[51] K. Hoffmann, T. Gambichler, and A. Rick. Diagnostic and neural analysis of skin cancer (danaos). A multicentre study for collection and computer-aided analysis of data from pigmented skin lesions using digital dermoscopy. *British Journal of Dermatology*, 149:801–809, 2003.

[52] M. Elbaum, A. Kopf, H. Rabinovitz, R. Langley, and H. Kamino. Automatic differentiation of melanoma from melanocytic nevi with multispectral digital dermoscopy: A feasibility study. *Journal of the American Academy of Dermatology*, 44:207–218, 2001.

[53] M. Elbaum. Computer-aided melanoma diagnosis. *Dermatologic clinics*, 20:735–747, 2002.

[54] D. Gutkowicz-Krusin, M. Elbaum M, A. Jacobs, S. Keem, A. W. Kopf, H. Kamino ad S. Wang, P. Rubin, H. Rabinovitz, and M. Oliviero. Precision of automatic measurements of pigmented skin lesion parameters with a melafind (tm) multispectral digital dermoscope. *Melanoma Research*, 10:563–570, 2000.

[55] M. Moncrieff, S. Cotton, E. Claridge, and P. Hall. Spectrophotometric intracutaneous analysis: A new technique for imaging pigmented skin lesions. *British Journal of Dermatology*, 146:448–457, 2002.

[56] S. Q. Wang, H. Rabinovitz, A. W. Kopf, and M. Oliviero. Current technologies for the in vivo diagnosis of cutaneous melanomas. *Clinics in dermatology*, 22:217–222, 2004.

[57] T. Schindewolf, R. Schiffner, W. Stolz, R. Albert, W. Abmayr, and H. Harms. Evaluation of different image acquisition techniques for a computer vision system in the diagnosis of malignant melanoma. *Clinics in dermatology*, 31:33–41, 1994.

[58] H. Ganster, A. Pinz, R. Rohrer, E. Wildling, M. Binder, and H. Kittler. Automated melanoma recognition. *IEEE Transactions on Medical Imaging*, 20:233–239, March 2001.

[59] S. Dreiseitl, L. Ohno-Machado, H. Kittler, S. Vinterbo, H. Billhardt, and M. Binder. A comparison of machine learning methods for the diagnosis of pigmented skin lesions. *Journal of Biomedical Informatics*, 34:28–36, 2001.

[60] M. Celebi, H. Kingravi, B. Uddin, H. Iyatomi, Y. A. Aslandogan, W. V. Stoecker, and R. H. Moss. A methodological approach to the classification of dermoscopy images. *computerised Medical Imaging and Graphics*, 31:362–373, 2007.

[61] G. Surowka and K. Grzesiak-Kopec. Different learning paradigms for the classification of melanoma skin lesions using wavelets. In *Proceedings of 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS 2007)*, pages 3136–3139, 2007.

[62] T. Hwanga, W. Leeb, S. Huad, and J. Fang. Cisplatin encapsulated in phosphatidylethanolamine liposomes enhances the in vitro cytotoxicity and in vivo intratumor drug accumulation against melanomas. *Journal of dermatological science*, 46:11–20, 2007.

[63] A. P. Dhawan. Early detection of cutaneous malignant melanoma by three dimensional nevoscopy. *Computer Methods and Programs in Biomedicine*, 21:59–68, 1985.

[64] S. V. Patwardhan, S. Dai, and A. P. Dhawan. Multi-spectral image analysis and classification of melanoma using fuzzy membership based partitions. *computerised Medical Imaging and Graphics*, 29:287–296, 2005.

[65] X. Yuan, Z. Yang, G. Zouridakis, and N. Mullani. Svm-based texture classification and application to early melanoma detection. In *Proceedings of 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS06)*, pages 4775–4778, 2006.

[66] R. G. Langley, M. Rajadhyaksha, and P. J. Dwyer et al. Confocal scanning laser microscopy of benign and malignant melanocytic skin lesions in vivo. *Journal of the American Academy of Dermatology*, 45:365–376, 2001.

[67] K. J. Busam, K. Hester, C. Charles, D. L. Sachs, C. R. Antonescu, S. Gonzalez, and A. C. Halpern. Detection of clinically amelanotic malignant melanoma and

assessment of its margins by in vivo confocal scanning laser microscopy. *Archives of Dermatology*, 137:923–929, 2001.

[68] N. Lassau, A. Spatz, M. F. Avril, A. Tardivon, A. Margulis, G. Mamelle, D. Vanel, and J. Leclere. Value of high-frequency us for preoperative assessment of skin tumors. *Radiographics*, 17:1559–1565, 1997.

[69] C. C. Harland, S. G. Kale, P. Jackson, P.S. Mortimer, and J. C. Bamber. Differentiation of common benign pigmented skin lesions from melanoma by high-resolution ultrasound. *British Journal of Dermatology*, 143:281–289, 2000.

[70] K. Hoffmann, J. Jung, S. Gammal, and P. Altmeyer. Malignant melanoma in 20-mhz b scan sonography. *Dermatology*, 185:40–55, 1992.

[71] J. Solomon, S. Mavinkurve, D. Cox, and R. M. Summers. Computer-assisted detection of subcutaneous melanomas. *Academic radiology*, 21:678–685, 2004.

[72] C. Pleiss, J. Risse, H. Biersack, and H. Bender. Role of fdg-pet in the assessment of survival prognosis in melanoma. *Cancer Biotherapy and Radiopharmaceuticals*, 22:740–747, 2007.

[73] I. K. Daftari, E. Aghaian, J. M. O'Brien, W.Dillon, and T. L. Phillips. 3d mri-based tumor delineation of ocular melanoma and its comparison with conventional techniques. *Medical physics*, 32:3355–3362, 2005.

[74] P. Aberg, I. Nicander, J. Hansson, P. Geladi, U. Holmgren, and S. Ollmar. Skin cancer identification using multifrequency electrical impedance-a potential screening tool. *IEEE transactions on biomedical engineering*, 51:2097–2102, 2004.

[75] S. Sigurdsson, P. A. Philipsen, L. K. Hansen, J. Larsen, M. Gniadecka, and H. C. Wulf. Detection of skin cancer by classification of raman spectra. *IEEE transactions on bio-medical engineering*, 51:1784–1793, 2004.

[76] M. Sonka, V. Hlavac, and R. Boyle. *Image processing, analysis, and machine vision.* Cengage-Engineering, 2007.

[77] R. M. Haralick and L. G. Shapiro. Image segmentation techniques. *Computer Vision, Graphics, and Image Processing*, 29(1):100–132, January 1985.

[78] M. E. Celebi, H. Iyatomi, G. Schaefer, and W. V. Stoecker. Lesion border detection in dermoscopy images. *Computerised Medical Imaging and Graphics*, 33(2):148–153, 2009.

[79] T. K. Lee, V. Ng, D. McLean, A. Coldman, R. Gallagherand, and J. Sale. A multi-stage segmentation method for images of skin lesions. In *IEEE Pacific Rim Conference on Communications, Computers, and Signal Processing*, pages 602–605, 1995.

[80] M. Hintz-Madsen, L. K. Hansen, J. Larsen, and K. Drzewiecki. A probabilistic neural network framework for detection of malignant melanoma. In *Artificial Neural Networks in Cancer Diagnosis, Prognosis and Patient Management*, pages 141–183. CRC Press, 2001.

[81] A. Sbonera, E. Blanzieria, C. Ecchera, P. Bauerb, M. Cristofolinib, G. Zumianib, and S. Fortia. A knowledge based system for early melanoma diagnosis support. In *Proceedings of the 6th Intelligent Data Analysis in Medicine and Pharmacology (IDAMAP) Workshop*, 2001.

[82] G. A. Hance, S. E. Umbaugh, R. H. Moss, and W. V. Stoecker. Unsupervised colour image segmentation: With application to skin tumor borders. *IEEE Engineering in Medicine and Biology Magazine*, 15:104–111, 1996.

[83] P. Schmid. Segmentation of digitised dermatoscopic images by two-dimensional colour clustering. *IEEE Transactions on Medical Imaging*, 18(2):164–171, 1999.

[84] R. Cucchiara, C. Grana, S. Seidenari, and G. Pellacani. Exploiting colour and topological features for region segmentation with recursive fuzzy c-means. *Machine Graphics and Vision*, 11:169–182, 2002.

[85] H. Galda, H. Murao, H. Tamaki, and S. Kitamura. Skin image segmentation using a self-organising map and genetic algorithms. *Transactions of the Institute of Electrical Engineers of Japan-Part C*, 123:2056–2062, 2002.

[86] P. Schmid-Saugeon, J. Guillod, and J. P. Thiran. Towards a computer-aided diagnosis system for pigmented skin lesions. *computerised Medical Imaging and Graphics*, 27(2):69–72, January 2003.

[87] R. Melli, C. Grana, and R. Cucchiara. Comparison of colour clustering algorithms for segmentation of dermatological images. In *SPIE Medical Imaging*, volume 6144, pages 3S1–3S9, 2006.

[88] D. Delgado, C. Butakoff, B. K. Ersboll, and W. V. Stoecker. Independent histogram pursuit for segmentation of skin lesions. *IEEE transactions on biomedical engineering*, 55:157–161, 2008.

[89] H. Zhou, M. Chen, L. Zou, R. Gass, L. Ferris, L. Drogowski, and J. M. Rehg. Spatially constrained segmentation of dermoscopy images. In *5th IEEE International Symposium on Biomedical Imaging*, pages 800–803, 2008.

[90] W. E. Denton, A. W. G Duller, and P. J. Fish. Boundary detection for skin lesions: an edge focusing algorithm. In *5th IEEE International Conference on Image Processing and its Applications*, pages 399–403, 1995.

[91] A. J. Round, A. W. Duller, and P.J. Fish. Colour segmentation for lesion classification. In *19th International Conference - IEEE/EMBS*, pages 582–585, 1997.

[92] J. Gao, J. Zhang, and M. G. Fleming. Segmentation of dermatoscopic images by stabilised inverse diffusion equations. In *IEEE International Conference on Image Processing (ICIP 1998)*, volume 3, pages 823–827, 1998.

[93] E. Zagrouba and W. Barhoumi. A preliminary approach for the automated recognition of malignant melanoma. *Image Analysis and Stereology Journal*, 23:121–135, 2004.

[94] E. Zagrouba and W. Barhoumi. An accelerated system for melanoma diagnosis based on subset feature selection. *Journal of Computing and Information Technology*, 1:69–82, 2005.

[95] M. E. Celebi, Y. A. Aslandogan, W. V. Stoecker, H. Iyatomi, H. Oka, and X. Chen. Unsupervised border detection in dermoscopy images. *Skin Research and Technology*, 13:454–462, 2007.

[96] M. E. Celebi, H. A. Kingravi, H. Iyatomi, Y. A. Aslandogan, W. V. Stoecker, R. H. Moss, J. M. Malters, J. M. Grichnik, A. A. Marghoob, H. S. Rabinovitz, and S. W. Menzies. Border detection in dermoscopy images using statistical region merging. *Skin Research and Technology*, 14:347–353, 2008.

[97] P. Schmid. Lesion detection in dermatoscopic images using anisotropic diffusion and morphological flooding. In *IEEE International Conference on Image Processing (ICIP 1999)*, volume 3, pages 449–453, 1999.

[98] Y. V. Haeghen, J. M. Naeyaert, and I. Lemahieu. Development of a dermatological workstation: preliminary results on lesion segmentation in cie l*a*b* colour space. In *International Conference on colour in Graphics and Image Processing*, volume 1, pages 6572–6575, 2000.

[99] B. Erkol, R. Moss, R. Stanley, W. Stoecker, and E. Hvatum. Automatic lesion boundary detection in dermoscopy images using gradient vector flow snakes. *Skin Research and Technology*, 11:17–26, 2005.

[100] T. Mendonca, A. Marcal, A. Vieira, J. C. Nascimento, M. Silveira, J. S. Marques, and J. Rozeira. Comparison of segmentation methods for automatic diagnosis of dermoscopy images. In *29th IEEE EMBS Annual International Conference*, volume 1, pages 6572–6575, 2007.

[101] T. Donadey, C. Serruys, and A. Giron. Boundary detection of black skin tumors using an adaptive radial-based approach. In *SPIE Medical Imaging*, volume 3379, pages 810–6, 2000.

[102] P. Pagadala. Tumor border detection in epiluminescence microscopy images. Master's thesis, Department of Electrical and Computer Engineering, University of Missouri- Rolla, Sweden, August 1998.

[103] T. Donadey, C. Serruys, A. Giron, and W. Woelker. Image segmentation based on adaptive 3-d-analysis of the cie-l*a*b colour space. In *SPIE Visual Communications and Image Processing*, volume 2727, pages 1197–1203, 1996.

[104] P. Heckbert. colour image quantisation for frame buffer display. *Computer Graphics*, 16, July 1982.

[105] J. MacQueen. Some methods for classification and analysis of multivariate observations. In *Proceedings of Fifth Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pages 281–297. University of California, 1967.

[106] J. C. Dunn. A fuzzy relative of the isodata process and its use in detecting compact well-separated clusters. *Cybernetics and Systems*, 3:32–57, 1973.

[107] K. Fukunaga and L. Hostetler. The estimation of the gradient of a density function, with applications in pattern recognition. *IEEE Transactions on Information Theory*, 21:32–40, 1975.

[108] D. Comaniciu and P. Meer. Robust analysis of feature spaces: colour image segmentation. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 750–755, 1997.

[109] N. Otsu. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1):62–66, 1979.

[110] Y. Deng Y and B. S. Manjunath. Unsupervised segmentation of colour-texture regions in images and video. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23:800–810, 2001.

[111] T. Lee, , V. Ng, R. Gallagher, A. Coldman, and D. McLean. Dullrazor: A software approach to hair removal from images. *Computers in Biology and Medicine*, 27:533–543, 1997.

[112] W. V. Stoecker, W. Weiling Li, and R. H. Moss. Automatic detection of asymmetry in skin tumors. *computerised Medical Imaging and Graphics*, 16(3):191–197, 1992.

[113] I. Maglogiannis and D. I. Kosmopoulos. Computational vision systems for the detection of malignant melanoma. *Oncology Reports*, 15:1027–1032, 2006.

[114] S. Seidenari, G. Pellacani, and C. Grana. Asymmetry in dermoscopic melanocytic lesion images: A computer description based on colour distribution. *Acta Dermato-Venereologica*, 86(2):123–128, March 2006.

[115] Y. Chang, R. Joe Stanley, R. H. Moss, and W. V. Stoecker. A systematic heuristic approach for feature selection for melanoma discrimination using clinical images. *Skin Research and Technology*, 11:165–178, 2005.

[116] Z. She, Y. Liu, and A. Damatoa. Combination of features from skin pattern and ABCD analysis for lesion classification. *Skin Research and Technology*, 13:25–33, 2007.

[117] K. M. Clawson, P. J. Morrow, B. W. Scotney, D. J. McKenna, and O. M. Dolan. Determination of optimal axes for skin lesion asymmetry quantification. In *IEEE International Conference on Image Processing (ICIP 2007)*, volume 2, pages 453–456, 2007.

[118] V. Ng, B. Fung, and T. K. Lee. Determining the asymmetry of skin lesion with fuzzy borders. *Computers in Biology and Medicine*, 35:103–120, December 2005.

[119] L. Andreassi, R. Perotti, M. Burroni, G. Dell 'Eva, and M. Biagioli. computerised image analysis of pigmented lesions. *Chronica Dermatologica*, 5:11–24, 1995.

[120] V. Ng and D. Cheung. Measuring asymmetries of skin lesions. In *IEEE International Conference on Systems, Man, and Cybernetics*, volume 5, pages 4211–4216, October 1997.

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

[122] A. G. Manousaki, A. G. Manios, E. I. Tsompanaki, J. G. Panayiotides, D. D. Tsiftsis, A. K. Kostaki, and A. D. Tosca. A simple digital image processing system to aid in melanoma diagnosis in an everyday melanocytic skin lesion unit. A preliminary report. *International Journal of Dermatology*, 45:402–410, April 2006.

[123] T. Holmstrom. A survey and evaluation of features for the diagnosis of malignant melanoma. Master's thesis, Department of Computing Science, Umea University, Sweden, August 2005.

[124] A. Bono, S. Tomatis, C. Bartoli, G. Tragni, G. Radaelli, A. Maurichi, and R. Marchesini. The ABCD system of melanoma detection: A spectrophotometric analysis of the asymmetry, border, colour, and dimension. *Cancer*, 85(1):72–77, January 1999.

[125] H. Zabrodsky, S. Peleg, and D. Avnir. Continuous symmetry measures. *Journal of the American Chemical Society*, 114(20):7843–7851, November 1992.

[126] H. Zabrodsky, S. Peleg, and D. Avnir. A measure of symmetry based on shape similarity. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 92)*, pages 703–706, June 1992.

[127] H. Zabrodsky, S. Peleg, and D. Avnir. Symmetry as a continuous feature. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(12):1154–1166, December 1995.

[128] E. Claridge, P. Hall, M. Keefe, and J. P. Allen. Shape analysis for classification of malignant melanoma. *Journal of biomedical engineering*, 14(3):229–234, May 1992.

[129] S. M. Chung and Q. Wang. Content-based retrieval and data mining of a skin cancer image database. In *International Conference on Information Technology - Coding and Computing (ITCC2001)*, pages 611–615, 2001.

[130] S. Seidenari, G. Pellacani, and C. Grana. Pigment distribution in melanocytic lesion images: A digital parameter to be employed for computer-aided diagnosis. *Skin Research and Technology*, 11:236–241, 2005.

[131] G. R. Day. How blurry is that border? An investigation into algorithmic reproduction of skin lesion border cut-off. *computerised Medical Imaging and Graphics*, 24(2):69–72, January 2000.

[132] T. K. Lee, M. S. Atkins, R. Gallagher, C. E. Macaulay, A. Coldman, and D. I. McLean. Describing the structural shape of melanocytic lesions. In *SPIE Medical Imaging*, volume 3661, pages 1170–1179, 1999.

[133] T. K. Lee and M. S. Atkins. New approach to measure border irregularity for melanocytic lesions. In *SPIE Medical Imaging*, volume 3979, pages 668–675, 2000.

[134] T. K. Lee, D. I. McLeanc, and M. S. Atkins. Irregularity index: A new border irregularity measure for cutaneous melanocytic lesions. *Medical Image Analysis*, 7:47–64, 2003.

[135] H. P. Greenspan and D. J. Benney. *Calculus-An Introduction to Applied Mathematics*. McGraw Hill, 1973.

[136] T. K. Lee and E. Claridge. Predictive power of irregular border shapes for malignant melanomas. *Skin Research and Technology*, 11:1–8, February 2005.

[137] Y. Cheng, R. Swamisai, S. E. Umbaugh, R. H. Moss, W. V. Stoecker, S. Teegala, and S. K. Srinivasan. Skin lesion classification using relative colour features. *Skin Research and Technology*, 14:53–64, 2008.

[138] T. K. Lee. *Measuring Border Irregularity and Shape of Cutaneous Melanocytic Lesions*. PhD thesis, School of Computing Science - Simon Fraser University, Canada, January 2001.

[139] I. Maglogiannis, S. Pavlopoulos, and D. Koutsouris. An integrated computer supported acquisition, handling, and characterisation system for pigmented skin lesions in dermatological images. *IEEE Transactions on Information Technology in Biomedicine*, 9:86–98, March 2005.

[140] V. Ng and T. K. Lee. Measuring border irregularities of skin lesions using fractal dimensions. In *SPIE Electronic Imaging and Multimedia Systems*, volume 2898, pages 64–72, November 1996.

[141] B. S. Aribisala and E. Claridge. A border irregularity measure using a modified conditional entropy method as a malignant melanoma predictor. In *Image Analysis and Recognition*, volume 3656 of *Lecture Notes in Computer Science*, pages 914–921. Springer Berlin, 2005.

[142] S. Seidenari, G. Pellacani, and C. Grana. Computer description of colours in dermoscopic melanocytic lesion images reproducing clinical assessment. *British Journal of Dermatology*, 149:523–529, 2003.

[143] Y. Faziloglu, R. J. Stanley, R. H. Moss, W. V. Stoecker, and R. P. McLean. Colour histogram analysis for melanoma discrimination in clinical images. *Skin Research and Technology*, 9:147–155, 2003.

[144] S. E. Umbaugh, R. H. Moss, and W. V. Stoecker. Automatic colour segmentation of images with application to detection of variegated colouring in skin tumours. *IEEE Engineering in Medicine and Biology*, 8:43–52, 1989.

[145] J. Chen, R. J. Stanley, R. H. Moss, and W. V. Stoecker. Colour analysis of skin lesion regions for melanoma discrimination in clinical images. *Skin Research and Technology*, 9:94–104, 2003.

[146] R. J. Stanley, R. H. Moss, W. V. Stoecker, and C. Aggarwal. A fuzzy-based histogram analysis technique for skin lesion discrimination in dermatology clinical images. *computerised Medical Imaging and Graphics*, 27(27):387–396, 2003.

[147] R. Joe Stanley, W. V. Stoecker, and R. H. Moss. A relative colour approach to colour discrimination for malignant melanoma detection in dermoscopy images. *Skin Research and Technology*, 13:62–72, 2007.

[148] I. Maglogiannis, E. Zafiropoulos, and C. Kyranoudis. Intelligent segmentation and classification of pigmented skin lesions in dermatological images. In *Advances in Artificial Intelligence*, volume 3955 of *Lecture Notes in Computer Science*, pages 214–223. Springer Berlin, 2006.

[149] C. Grana, R. Cucchiara, G. Pellacani, and S. Seidenari. Line detection and texture characterisation of network patterns. In *Proceedings of the 18th International Conference on Pattern Recognition (ICPR 2006)*, volume 2, pages 275–278, 2006.

[150] S. Fischer, P. Schmid, and G. Joe. Analysis of skin lesions with pigmented networks. In *Proceedings of International Conference on Image Processing*, volume 1, pages 323–326, 1996.

[151] R. C. Gonzalez and P. Wintz. *Digital Image Processing*. Addison-Wesley, 1987.

[152] M. G. Fleming, C. Steger, A. B. Cognetta, and J. Zhang. Analysis of the network pattern in dermatoscopic images. *Skin Research and Technology*, 5:42–48, 1999.

[153] A. Murali, W. V. Stoecker, and R. H. Moss. Detection of solid pigment in dermatoscopy images using texture analysis. *Skin Research and Technology*, 6:193–198, 2000.

[154] M. E. Celebi, H. A. Kingravi, Y. A. Aslandogan, and W. V. Stoecker. Detection of blue-white veil areas in dermoscopy images using machine learning techniques. In *SPIE Medical Imaging*, volume 6144, pages 5T1–5T8, 2006.

[155] W. V. Stoecker, M. Wronkiewiecz, R. Chowdhury, R. Joe Stanley, J. Xua, A. Bangert, B. Shrestha, D. A. Calcara, H. S. Rabinovitz, M. Oliviero, F. Ahmed, L. A. Perry, and R. Drugge. Detection of granularity in dermoscopy images of malignant melanoma using colour and texture features. *Computerised Medical Imaging and Graphics*, 35:144–147, 2011.

[156] M. Dash and H. Liu. Feature selection for classification. *Intelligent Data Analysis*, 1:679–693, 1997.

[157] H. Liu and H. Motoda. *Feature Selection for Knowledge Discovery and Data Mining*. Kluwer Academic, Boston, 1998.

[158] P. Langley. Selection of relevant features in machine learning. In *AAAI Fall Symposium on Relevance*, pages 140–144, 1994.

[159] M. Dash, K. Choi, P. Scheuermann, and H. Liu. Feature selection for clustering-a filter solution. In *Second IEEE International Conference on Data Mining (ICDM'02)*, pages 115–122, 2002.

[160] M. A. Hall. Correlation-based feature selection for discrete and numeric class machine learning. In *17th International Conference on Machine Learning*, pages 359–366, 2000.

[161] L. Yu and H. Liu. Feature selection for high-dimensional data: A fast correlation-based filter solution. In *20th International Conference on Machine Learning*, pages 856–863, 2003.

[162] R. Kohavi and G. H. John. Artificial intelligence. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 97:273–324, 1997.

[163] R. Caruana and D. Freitag. Greedy attribute selection. In *11th International Conference on Machine Learning*, pages 28–36, 1994.

[164] J. G. Dy and C. E. Brodley. Feature subset selection and order identification for unsupervised learning. In *17th International Conference on Machine Learning*, pages 247–254, 2000.

[165] Y. Kim, W. Street, and F. Menczer. Feature selection for unsupervised learning via evolutionary search. In *sixth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 365–369, 2000.

[166] S. Das. Filters, wrappers and a boosting-based hybrid for feature selection. In *18th International Conference on Machine Learning*, pages 74–81, 2001.

[167] A. Y. Ng. On feature selection: Learning with exponentially many irrelevant features as training examples. In *15th International Conference on Machine Learning*, pages 404–412, 1998.

[168] E. Xing, M. Jordan, and R. Karp. Feature selection for high-dimensional genomic microarray data. In *18th International Conference on Machine Learning*, pages 601–608, 2001.

[169] R. Walvick, K. Patel, S. Patwardhan, and A. Dhawan. Classification of melanoma using wavelet-transform-based optimal feature set. In *SPIE Medical Imaging 2002: Image Processing*, volume 5370, pages 944–951, 2004.

[170] T. Roβ, H. Handels, J. Kreusch, H. Busche, H. Wolf, and S. Poppl. Automatic classification of skin tumours with high resolution surface profiles. In V. Hlavac and R. Sara, editors, *Computer Analysis of Images and Patterns*, volume 970 of *Lecture Notes in Computer Science*, pages 368–375. Springer Berlin / Heidelberg, 1995.

[171] A. Nimunkar, A. Dhawan, P. Relue, and S. Patwardhan. Wavelet and statistical analysis for melanoma. In *SPIE Medical Imaging 2002: Image Processing*, volume 4684, pages 1346–1352, 2002.

[172] H. Iyatomi, H. Oka, M. Emre Celebi, M. Tanaka, and K. Ogawa. Parameterisation of dermoscopic findings for the internet-based melanoma screening system. In *IEEE Symposium on Computational Intelligence in Image and Signal Processing (CIISP 2007)*, pages 189–193, 2007.

[173] A. Green, N. Martin, J. Pfitzner, MO'Rourke, and N. Knight. Computer image analysis in the diagnosis of melanoma. *Journal of the American Academy of Dermatology*, 31:958–964, 1994.

[174] G. Pellacani, M. Martini, and S. Seidenari. Digital videomicroscopy with image analysis and automatic classification as an aid for diagnosis of spitz nevus. *Skin Research and Technology*, 5:266–272, 1999.

[175] M. Burroni, R. Corona, G. Dell'Eva, F. Sera, R. Bono, P. Puddu, R. Perotti, F. Nobile, L. Andreassi, and P. Rubegni. Melanoma computer-aided diagnosis: Reliability and feasibility study. *Clinical Cancer Research*, 10:1881–1886, March 2004.

[176] K. V. Mardia, J. T. Kent, and J. M. Bibby. *Multivariate analysis*. London Academic Press, London, U.K., 1979.

[177] S. E. Umbaugh, R. H. Moss, and W. V. Stoecker. Applying artificial intelligence to the identification of variegated colouring in skin tumors. *IEEE engineering in medicine and biology magazine*, 10:57–62, 1991.

[178] B. D. Ripley. Neural networks and related methods for classification. *Journal of the Royal Statistical Society B*, 56(3):409–456, 1994.

[179] C. M. Bishop. *Neural Networks for Pattern Recognition*. Oxford University Press, 1995.

[180] B. V. Dasarathy. *Nearest neighbor (NN) norms: NN pattern classification techniques*. IEEE Computer Society Press, Los Alamitos, CA, 1991.

[181] T. Tommasi, E. La Torre, and B. Caputo. Melanoma recognition using representative and discriminative kernel classifiers. In *Computer Vision Approaches to Medical Image Analysis*, volume 4241 of *Lecture Notes in Computer Science*, pages 1–12. Springer Berlin, 2006.

[182] M. Amico, M. Ferri, and I. Stanganelli. Qualitative asymmetry measure for melanoma detection. In *IEEE International Symposium on Biomedical Imaging: Nano to Macro*, volume 2, pages 1155–1158, 2004.

[183] C. J. C. Burges. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 2(2):121–167, 1998.

[184] A. Kjoelen, M. J. Thompson, S. E. Umbaugh, R. H. Moss, and W. V. Stoecker. Performance of ai methods in detecting melanoma. *IEEE Engineering in Medicine and Biology Magazine*, 14:411–416, 1995.

[185] M. Wiltgen, A. Gergerb, and J. Smolle. Tissue counter analysis of benign common nevi and malignant melanoma. *International Journal of Medical Informatics*, 69:17–28, 2003.

[186] M. Zorman, M. M. Stiglic, Peter K., and I. Maltic. The limitations of decision trees and automatic learning in real world medical decision making. *Journal of Medical Systems*, 21(6):403–415, 1997.

[187] D. E. Clark. Computational methods for probabilistic decision trees. *Computers and Biomedical Research*, 30:19–33, 1997.

[188] L. Breiman. Random forests. *Machine learning*, 45:5–32, 2001.

[189] N. Landwehr, M. Hall, and E. Frank. Logistic model trees. *Machine Learning*, 95:161–205, 2005.

[190] H. Iyatomi, H. Oka, M. Emre Celebi, M. Hashimoto, M. Hagiwara, M. Tanaka, and K. Ogawa. An improved internet-based melanoma screening system with dermatologist-like tumor area extraction algorithm. *computerised Medical Imaging and Graphics*, 32(7):566–579, 2008.

[191] L. Lucchese and S.K. Mitra. colour image segmentation: A state-of-the-art survey. In *Proceedings of Indian National Science Academy Part A, PINSA2001*, pages 207–221, 2001.

[192] K. N. Plataniotis and A. N. Venetsanopoulos. *colour Image Processing and Applications*. Springer-Verlag, 2000.

[193] International commission on illumination. Available at: http://www.cie.co.at/.

[194] R. M. Haralick and L. G. Shapiro. *Computer and Robot Vision*, volume 1. Addison-Wesley, 1992.

[195] T. Sorensen. A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on danish commons. *Royal Danish Academy of Sciences and Letters*, 5:1–34, 1948.

[196] S. Finch and G. Cumming. Putting research in context: Understanding confidence intervals from one or more studies. *Journal of Pediatric Psychology*, 34:903–916, 2009.

[197] L. Xu, M. Jackowski, A. Goshtasby, D. Rosemanb, S. Binesb, C. Yu, A. Dhawand, and A. Huntleye. Segmentation of skin cancer images. *Image and Vision Computing*, 17:65–74, 1999.

[198] L. G. Shapiro and G. C. Stockman. *Computer Vision*. Prentence Hall, 2001.

[199] R. Kohavi. A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*, pages 1137–1143, Montreal, Canada, 1995. San Francisco: Morgan Kaufmann.

[200] G. Joel, P. Schmid-Saugeon, D. Guggisberg, J. P. Cerottini, R. Braun, J. Krischer, J. Saurat, and M. Kunt. Validation of segmentation techniques for digital dermoscopy. *Skin Research and Technology*, 8:240–249, 2002.

[201] M. E. Celebi, G. Schaefer, H. Iyatomi, W. V. Stoecker, J. M. Malters, and J. M. Grichnik. An improved objective evaluation measure for border detection in dermoscopy images. *Skin Research and Technology*, 15:444–450, 2009.

[202] R. H. Byrd, R. B. Schnabel, and G. A. Shultz. A trust region algorithm for nonlinearly constrained optimisation. *SIAM Journal on Numerical Analysis*, 24:1152–1170, 1987.

[203] R. H. Byrd, M. E. Hribar, and J. Nocedal. An interior point algorithm for large-scale nonlinear programming. *SIAM Journal on optimisation*, 9(4):877–900, 1999.

[204] R. Garnavi, M. Aldeen, M. E. Celebi, S. Finch, and G. Varigos. Border detection in dermoscopy images using hybrid thresholding on optimised colour channels. *computerised Medical Imaging and Graphics, Special Issue: Advances in Skin Cancer Image Analysis*, 35:105–115, 2011.

[205] R. Garnavi, M. Aldeen, M. E. Celebi, A. Bhuiyan, C. Dolianitis, and G. Varigos. Automatic segmentation of dermoscopy images using histogram thresholding on optimal colour channels. *International Journal of Medicine and Medical Sciences*, 1(2):126–134, 2010.

[206] R. Garnavi, M. Aldeen, M. E. Celebi, A. Bhuiyan, C., and G. Varigos. Skin lesion segmentation using colour channel optimisation and clustering-based histogram thresholding. In *International Conference on Machine Vision, Image Processing, and Pattern Analysis (MVIPPA09)*, volume 60, pages 549–557, Bangkok, Thailand, 2009. World Academy of Science, Engineering and Technology.

[207] R. Garnavi, M. Aldeen, S. Finch, and G. Varigos. Global versus hybrid thresholding for border detection in dermoscopy images. In A. Elmoataz, O. Lezoray, F. Nouboud, D. Mammass, and J. Meunier, editors, *Image and Signal Processing*, volume 6134 of *Lecture Notes in Computer Science*, pages 531–540. Springer Berlin/Heidelberg, 2010.

[208] R. Garnavi, M. Aldeen, and M. E. Celebi. Weighted performance index for objective evaluation of border detection methods in dermoscopy images. *Skin Research and Technology*, 17:35–44, 2011.

[209] R. Garnavi and M. Aldeen. Optimised weighted performance index for evaluation of border detection methods. *submitted to IEEE Transactions on Information Technology in Biomedicine*, submission: July 2010, revision: June 2011.

[210] S. Patwardhan, A. Dhawan, and P. Relue. Classification of melanoma using tree structured wavelet transforms. *Computer Methods and Programs in Biomedicine*, 72:223–239, 2003.

[211] H. Handels, T. Roβ, J. Kreuschb, H. H. Wolffb, and S. J. Poppla. Feature selection for optimised skin tumor recognition using genetic algorithms. *Artificial Intelligence in Medicine*, 16:283–297, 1999.

[212] I. Kononenko. Estimating attributes: Analysis and extensions of relief. In *European Conference on Machine Learning*, pages 171–182, 1994.

[213] M. Hall. *Correlation-based Feature Subset Selection for Machine Learning*. PhD thesis, University of Waikato, Hamilton, New Zealand, 1998.

[214] S. G. Mallat. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11:674–693, 1989.

[215] T. Chang and C. Jay Kuo. Texture analysis and classification with tree structured wavelet transform. *IEEE Transaction on Information Technology in Biomedicine*, 2:429–441, 1993.

[216] P. J. Brockwell and R. A. Davis. *Time Series: Theory and Methods*. Springer, 2 edition, 2006.

[217] I. H. Witten and E. Frank. *Practical machine learning tools and techniques*. Morgan Kaufmann, San Francisco, 2 edition, 2005.

[218] S. Russell and P. Norvig. *Artificial Intelligence: A Modern Approach*. Prentice Hall, 1 edition, 2003.

[219] V. N. Vapnik. *Statistical learning theory*. Wiely, New York, 1998.

[220] J. Platt. Fast training of support vector machines using sequential minimal optimisation. pages 185–208, 1999.

[221] P. Sajda. Machine learning for detection and diagnosis of disease. *Biomedical Engineering*, 8:537–565, 2006.

[222] B. E. Boser, I. M. Guyon, and V. N. Vapnik. A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on Computational Learning Theory*, pages 144–152, 1992.

[223] S. S. Keerthi and C-J. Lin. Asymptotic behaviors of support vector machines with gaussian kernel. *Neural Computtation*, 15:1667–1689, 2003.

[224] C. J. C. Burges. A tutorial on support vector machines for pattern recognition. *Data mining and knowledge discovery*, 2:121–167, 1998.

[225] F. Yang, H. Wang, H. Mi, C. Lin, and W. Cai. Using random forest for reliable classification and cost-sensitive learning for medical diagnosis. *BMC Bioinformatics*, 10, 2009.

[226] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone. *Classification and Regression Trees*. Chapman and Hall, Belmont, CA, 1984.

[227] J. R. Quinlan. *C4.5: programs for machine learning*. Morgan Kaufmann, 1993.

[228] D. M. Chickering. Learning bayesian networks is np-complete. In D. Fisher and H. Lenz, editors, *Artificial Intelligence and Statistics V.*, chapter 12, pages 121–130. Springer-Verlag, 1996.

[229] N. Friedman, D. Geiger, and M. Goldszmidt. Bayesian network classifiers. *Machine Learning*, 29:131–163, 1997.

[230] Chao Sima and Edward R. Dougherty. The peaking phenomenon in the presence of feature-selection. *Pattern Recognition Letters*, 29:1667–1674, 2008.

[231] M. Hall, E. Frank, and G. Holmes. The WEKA data mining software: An update. *SIGKDD Explorations*, 11:10–18, 2009.

[232] R. Garnavi, M. Aldeen, and J. Bailey. Classification of melanoma lesions using wavelet-based texture analysis. In *Digital Image Computing: Techniques and Applications (DICTA2010)*, pages 75–81, Syndey, Australia, 2010. IEEE press.

[233] R. Garnavi, M. Aldeen, and J. Bailey. Computer-aided diagnosis of melanoma using geometry, time series and wavelet-based texture analysis. *Submitted to IEEE Transactions on Information Technology in Biomedicine*, submission: June 2011.

University Library

A gateway to Melbourne's research publications

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:
GARNAVI, RAHIL

Title:
Computer-aided diagnosis of melanoma

Date:
2011

Citation:
Garnavi, R. (2011). Computer-aided diagnosis of melanoma. PhD thesis, Engineering, Department of Electrical and Electronic Engineering, The University of Melbourne.

Persistent Link:
http://hdl.handle.net/11343/36426

File Description:
Computer-aided diagnosis of melanoma

Terms and Conditions:
Terms and Conditions: Copyright in works deposited in Minerva Access is retained by the copyright owner. The work may not be altered without permission from the copyright owner. Readers may only download, print and save electronic copies of whole works for their own personal non-commercial use. Any use that exceeds these limits requires permission from the copyright owner. Attribution is essential when quoting or paraphrasing from these works.