Developing Improved Algorithms for Detection and Analysis of Skin Cancer

By

Ammara Masood

Submitted in partial fulfilment of the requirements for the

Doctor of Philosophy

Faculty of Engineering and Information Technology

UNIVERSITY OF TECHNOLOGY, SYDNEY

July, 2016

CERTIFICATE OF AUTHORSHIP/ORIGINALITY

I certify that the work in this thesis has not previously been submitted for a degree nor has it been submitted as part of requirements for a degree except as fully acknowledged within the text.

I also certify that the thesis has been written by me. Any help that I have received in my research work and the preparation of the thesis itself has been acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

Signature of Candidate

Acknowledgment

Over the last four years I have had the privilege of working with many people who have made my time at University of Technology, Sydney an enjoyable and intellectually stimulating experience.

I would like to thank all the people who have helped me along the way and contributed to this thesis. I am especially grateful to my Supervisor Associate Professor Dr. Adel Al-Jumaily. He has been very motivating throughout these years. While working with Dr. Adel, I have learned how to pursue my research with an intellectual rigor and how to critically evaluate my work and improve my research skills.

I would like to take this opportunity to acknowledge Scope Global for sponsoring my PhD through the Endeavour Post Graduate Award.

I would like to acknowledge Engr. Omama Masood for doing the proof reading and Engr. Tariq Adnan and Engr. Bani Kaur for helping in formatting corrections.

I would like to thank my friends and colleges who have encouraged and supported me to accomplish my degree.

Finally, I would like to acknowledge the endless love of my family who have been a constant source of support. Their faith in me and encouragement has helped me throughout my life.

Dedication

This thesis is dedicated to the five most important persons of my life without their support this journey would have been just a dream. Special thanks to my Grandfather M.R.A. Khan, my father M. Masood Alam Khan, my mother Azra Muzammil, my aunt Bushra Muzammil and my very best friend Tariq Adnan for being my inspiration and encouraging me throughout my career.

.

Abstract

Malignant melanoma is one of the deadliest forms of skin cancer and number of cases showed rapid increase in Europe, America, and Australia over the last few decades. Australia has one of the highest rates of skin cancer in the world, at nearly four times the rates in Canada, the US and the UK. Cancer treatment costs constitute more 7.2% of health system costs. However, a recovery rate of around 95% can be achieved if melanoma is detected at an early stage. Early diagnosis is obviously dependent upon accurate assessment by a medical practitioner. The variations of diagnosis are sufficiency large and there is a lack of detail of the test methods. This thesis investigates the methods for automated analysis of skin images to develop improved algorithms and to extend the functionality of the existing methods used in various stages of the automated diagnostic system. This in the long run can provide an alternative basis for researchers to experiment new and existing methodologies for skin cancer detection and diagnosis to help the medical practitioners.

The objective is to have a detailed investigation for the requirements of automated skin cancer diagnostic systems, improve and develop relevant segmentation, feature

selection and classification methods to deal with complex structures present in both dermoscopic/digital images and histopathological images.

During the course of this thesis, several algorithms were developed. These algorithms were used in skin cancer diagnosis studies and some of them can also be applied in wider machine learning areas. The most important contributions of this thesis can be summarized as below:

-Developing new segmentation algorithms designed specifically for skin cancer images including digital images of lesions and histopathalogical images with attention to their respective properties. The proposed algorithm uses a two-stage approach. Initially coarse segmentation of lesion area is done based on histogram analysis based orientation sensitive fuzzy C Mean clustering algorithm. The result of stage 1 is used for the initialization of a level set based algorithm developed for detecting finer differentiating details. The proposed algorithms achieved true detection rate of around 93% for external skin lesion images and around 88% for histopathological images.

-Developing adaptive differential evolution based feature selection and parameter optimization algorithm. The proposed method is aimed to come up with an efficient approach to provide good accuracy for the skin cancer detection, while taking care of number of features and parameter tuning of feature selection and classification algorithm, as they all play important role in the overall analysis phase. The proposed method was also tested on 10 standard datasets for different kind of cancers and results shows improved performance for all the datasets compared to various state-of the art methods.

- Proposing a parallelized knowledge based learning model which can make better use of the differentiating features along with increasing the generalization capability of the classification phase using advised support vector machine. Two classification algorithms were also developed for skin cancer data analysis, which can make use of both labelled and unlabelled data for training. First one is based on semi advised support vector machine. While the second one based on Deep Learning approach. The method of integrating the results of these two methods is also proposed. The experimental analysis showed very promising results for the appropriate diagnosis of melanoma. The classification accuracy achieved with the help of proposed algorithms was around 95% for external skin lesion classification and around 92 % for histopathalogical image analysis.

Skin cancer dataset used in this thesis is obtained mainly from Sydney Melanoma Diagnostic Centre, Royal Prince Alfred Hospital. While for comparative analysis and benchmarking of the few algorithms some standard online cancer datasets were also used. Obtained result shows a good performance in segmentation and classification and can form the basis of more advanced computer aided diagnostic systems. While in future, the developed algorithms can also be extended for other kind of image analysis applications.

Table of Contents

List of Abbreviations

Adaptive Differential evolution (ADE)

Adaptive Sequential Forward Floating Selection (ASFFS)

Adaptive Snake (AS)

Adaptive Thresholding based LS (AT-LS)

Advised Weighted Support Vector Machine (A-W-SVM)

Anisotropic Diffusion Filters (ADF)

Ant Colony Optimization (ACO)

Artificial Neural Network (ANN)

Atypical Pigment Network (APN)

Basal Cell Carcinoma (BCC)

Binary Particle Swarm Optimization (BPSO)

Circularity Index (CI)

Computer Aided Diagnosis (CAD)

Confocal Scanning Laser Microscopy (CSLM)

Derivatives of Gaussian (DOG)

Dice Similarity Coefficient (DSC)

Differential Evolution (DE)

Digital Dermoscopy Analysis (DDA)

Discrete Wavelet Transform (DWT)

Epiluminescence Microscopy (ELM)

Expectation Maximization (EM)

Expectation Maximization based level set (EM-LS)

Expertise Weight (EW)

False Negative Error (FNE)

False Positive Error (FPE)

Fast Fourier Transform (FFT)

Genetic Algorithm (GA)

Geodesic active Contour (GAC)

Gradient Vector Flow (GVF)

Gray Level Run length Matrix (GLRLM)

Gray Tone difference Matrix (GTDM)

Grey Level Co-Occurrence Matrix (GLCM)

Ground Truth (GT)

Hammoude Distance (HM)

Histogram analysis based Fuzzy C Mean thresholding (H-FCM)

Hue Saturation Lightness (HSL)

Hue-Saturation- Intensity (HSI)

Hybrid Genetic Algorithm (HGA)

Improved Binary Particle Swarm Optimization (IBPSO)

Independent Histogram Pursuit (IHP)

K Nearest Neighbour (KNN)

K mean clustering based LS (K-LS)

Karhunen-Loéve (KL)

Laplacian-of-Gaussian (LoG)

Level Set (LS)

Levenberg–Marquardt (LM)

Logistic regression (LogR)

Magnetic Resonance Imaging (MRI)

Malignant Melanoma (MM)

Multiple Layer Perceptron (MLP)

Optical Coherence Tomography (OCT)

Particle Swarm Optimization (PSO)

Peer Group Filter (PGF)

Pigmented Skin Lesions (PSLs)

Pigmented Spitz Naevi (PSNs)

Principal Component Analysis (PCA)

Radial Basis Function (RBF)

Receiver Operating Characteristic (ROC)

Restricted Boltzmann Machines (RBM)

Scaled Conjugate Gradient (SCG)

Segmented Region (SR)

Sequential Backward Elimination (SBE)

Sequential Floating Forward Selection (SFFS)

Sequential Forward Selection (SFS)

Spatial Fuzzy Clustering based LS (SF-LS)

Squamous Cell Carcinoma (SCC)

Statistical Region Merging (SRM)

Support Vector Machine (SVM)

Support Vectors (SV)

Total Body Photograph (TBP)

Total Dermatoscopy Score (TDS)

Transductive Support Vector Machines (TSVM)

True Positive Rate (TPR)

Typical Pigment Network (TPN)

Wavelet Packet Transform (WPT)

Wavelet Packet Transform (WPT)

List of Figures

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

List of Tables

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

Chapter 1

Introduction

This thesis aims to contribute in the development of automated diagnosis systems for skin cancer diagnosis. In this chapter motivations for the thesis are presented, including an outline of the problems to be addressed that will be expanded upon later in the thesis. After that the thesis questions will be stated, including the reasons for their importance, followed by a chapter-by-chapter synopsis of the thesis contents.

1.1 Problem & Motivation

Skin cancer represents a major health problem. Malignant melanoma is one of the deadliest forms of skin cancer which accounts for 75% of all cancer deaths [1] [2] [3]. The number of melanoma cases showed rapid increase in Europe, America, and Australia over the last few decades. In US, one person dies of melanoma every 57 minutes [4]. Australia has one of the highest rates of skin cancer in the world, at nearly four times the rates in Canada, the US and the UK [5]. In 2014, an estimated 128,000 new cases of cancer were diagnosed in Australia, with that number expected to rise to 150,000 by 2020 [6]. Two in three Australians will be diagnosed with skin cancer by the time they are 70. The Chair of Public Health Committee pronounced that more than 430,000 cases are treated for non-melanoma, and more than 10,300 people are treated for melanoma, with 1430 people dying each year [7]. Cancer treatment costs constitute more than $3.8 billion (7.2%) of health system costs. In 2001, it was estimated the treatment of non-melanoma skin cancer cost $264 million and that of melanoma cost $30 million [1].

If the melanoma gets deeper more than 3 millimetres[8], the chance of survival is 59% [9]. However, melanoma has more than 95% cure rate if diagnosed (detected) and treated in early stages. It can be removed by simple surgery if it has not entered the blood stream. Early diagnosis is obviously dependent upon patient attention and accurate assessment by a medical practitioner. The variations of diagnosis are sufficiency large and there is a lack of detail of the test methods. The diagnostic is mostly based on visual

inspection. The traditional imaging is just a recording of what the human eye can see using a digital camera while the Dermoscopy known as Epiluminescence Light Microscopy (ELM) images need a professional with experience to get the required image.

Another way of diagnosis is thorough histopathological image analysis by pathologists. Pathologists use histopathological images of biopsy samples removed from patients and examine them under a microscope. A pathologist typically examines the image to observe the deviations in the cell structures and/or the change in the distribution of the cells across the tissue under examination. However, these judgments depend on their personal experience and expertise and often lead to considerable variability [10, 11].

To overcome these problems and improve the reliability of diagnosis process, it is important to develop computational tools for automated diagnosis that operate on quantitative measures. Such tools can facilitate objective mathematical judgment complementary to that of medical experts, and help them in identifying the affected areas more efficiently.

1.2 Thesis Aims and Contributions

In this section objectives and goals together with contributions of this thesis are highlighted.

1.2.1 Thesis aim and objectives

As most general practitioners have less experience in the full range of melanoma forms, which make many cases not diagnosed properly. Last decade researchers developed automatic early diagnostic aided systems to help in improving the diagnostic accuracy for skin cancer. The common stages of computer aided diagnostic systems based on image analysis include pre-processing, segmentation, feature extraction/selection and classification. However, variations of diagnosis are sufficiency large and there are many challenges that need to be addressed.

The main motivation behind this thesis is to develop improved algorithms and to extend the functionality of the existing methods used in various steps of the automated diagnostic system and in the long run provide an alternative basis for researchers to experiment new and existing methodologies for skin cancer detection and diagnosis. Some of the research questions addressed in this thesis include the following.

How to improve the segmentation process of skin lesion so that it can deal with various different types of complexities present in skin images that can affect the segmentation process, both for dermoscopic/digital images and histopathalogical images?

What features extraction methods can be used to get a feature set that can provide the most differentiating information for melanoma and benign to help the classification process?

How to do proper features selection and make efficient use of extracted features for improving the classifier performance?

How to efficiently train the classifier using both labelled and unlabelled training data to improve its generalization performance if a limited amount of labelled training data is available, especially for the case of histopathalogical image analysis.

This PhD research targets the above-mentioned questions in an attempt to solve them by developing new algorithms or improving the performance of existing algorithms used in the area. The developed methods target different parts of the automated recognition system like segmentation, features extraction, feature selection/dimensionality reduction techniques, and pattern recognizers or classifiers. The results of the proposed methods for various stages are compared with other algorithms and quite promising results are observed.

1.2.2 Contributions of the Thesis

The main contribution of the thesis is to develop a better understanding of the requirements of automated skin cancer diagnostic systems, provide improved algorithms for segmentation and analysis of skin images and detection of melanoma; furthermore some parts of this thesis can be used in a wider range of machine learning and pattern recognition studies. All of these will be discussed in more detail in the related chapters.

A. Contributions to Diagnosis of Skin Cancer

- Proposing novel segmentation method for skin cancer detection and demonstrating the importance of segmentation both for normal skin images and histopathological images (Chapter 4).
- Finding best differentiating feature sets that can help in better classification performance in the analysis phase (Chapter 5).

- Proposing knowledge based learning model that provided good results for automated diagnosis of skin cancer. In addition to that a parallelized version based on semi advised SVM and deep learning machine is also proposed for pattern recognition in the cases where less labelled and more unlabelled data is available (chapter 6).

B. Contributions beyond Skin Cancer Case Study

- Presenting a review on automated detection methods and providing a comprehensive research framework for developing efficient automated diagnostic models. (Chapter 2).
- Proposing an adaptive Differential evolution based feature selection method (Chapter 5)
- Proposing learning model that can be trained using unlabelled data by making efficient use of limited labelled training data for complex pattern recognition applications like histopathological image analysis. (chapter 6)

1.2.3 Structure of the Thesis

In **chapter 2** some general aspects of skin cancer are introduced. Of course, every related medical detail cannot be covered, but this chapter gives some essential medical knowledge of skin cancer for researchers working in the area of pattern recognition. Different types of skin cancer are introduced followed by introduction to skin lesion imaging methods. Symptoms of skin cancer and different diagnosis approaches are also considered and finally histopathological image analysis and its importance in skin cancer diagnosis are discussed.

In **chapter 3** the literature relating to the development of image based automated diagnostic systems for skin cancer are reviewed. The aim of this review is to summarize and compare the state of the art in these systems and current practices, problems, and prospects of image acquisition, pre-processing, segmentation, feature extraction and selection, and classification stages. A general scheme for computer aided diagnosis system is provided. This chapter reports statistics and results from some of the important implementations present in literature.

Analysis of related works is performed with respect to several criteria, such as lesion segmentation methods, feature extraction and selection process, size of data sets, classification techniques, and performance measures used in reporting the diagnosis results. Whenever available, indication of various conditions that may affect the technique's performance is reported. The deficiencies in some of the existing studies are highlighted and suggestions for further research are provided. This chapter provide a framework that represents a comprehensive guideline for selecting suitable algorithms needed for different stages of automatic diagnostic procedure for ensuring timely diagnosis of skin cancer. It also provides framework for comparative assessment of skin cancer diagnostic models.

Chapter 4 provides details of the proposed methods for pre-processing and segmentation stages. The chapter is divided into two parts, part 1 covers the work done based on the digital skin cancer images. The theoretical details of the proposed segmentation method are provided, followed by experimental evaluation and comparative analysis. While part 2 provides details for the research carried out based on histopathological images. The details of the proposed pre-processing and segmentation methods are provided followed by the experimental results, performance comparisons and conclusion.

Feature extraction methods for skin image analysis are studied in **Chapter 5**. This chapter provides details of various differentiating features for melanoma detection including physical features and statistical and spectral analysis based features.

Feature selection phase is also presented in **Chapter 5**. The details of proposed adaptive differential evolution based feature selection and parameter optimization algorithm are provided. Afterwards, different experimental results are examined based on skin cancer database as well as general benchmarks cancer datasets to show the performance of the proposed algorithm.

Chapter 6 provides details of the classification stage. The two proposed classification algorithms are explained in detail. First one provides the classification process based on labelled training data using advised support vector machine. A parallelized expertise weight based learning model is also proposed which can make better use of the differentiating features along with increasing the generalization capability of the classification phase. Second part provides the method proposed for using

unlabelled data for training a classifier by making efficient use of limited labelled data. A semi advised SVM algorithm is proposed that is used in parallel with a deep learning based algorithm for the analysis and classification of skin images. The related experimental work and comparative analysis is provided.

Chapter 7 provides a brief summary of the thesis and suggestions for future work are provided.

1.2.4 Publications Resulting from the Thesis

Results of these investigations have been published in a number of papers, these peer reviewed publications includes 3 Journals, 4 Book Chapters in Lecture Notes in Computer Science (LNCS), and 6 International Conferences. In addition to that 2 research papers are submitted to the Journals as mentioned in the following list and are in review process.

Journal

1. Ammara Masood, Adel Al-Jumaily, "Semi Advised SVM with Adaptive Differential Evolution Based Feature Selection for Skin Cancer Diagnosis", Journal of Computer and Communications, Vol.3 No.11, 2015.

2. Ammara Masood, Adel Ali Al-Jumaily, "Computer Aided Diagnostic Support System for Skin cancer: Review of techniques and algorithms", International Journal of Biomedical Imaging, pp 22, 2013.

3. Ammara Masood, Adel Ali Al-Jumaily, "Fuzzy C mean Thresholding based Level Set for Automated Segmentation of Skin Lesions", Journal of Signal and Information Processing, Vol. 4, No. 3B, ISSN: 2159-4465, August 2013, pp. 66-71.

4. Ammara Masood, Adel Al-Jumaily, "Orientation Sensitive Fuzzy C Means Based Fast Level Set Evolution for skin Histopathological Images Segmentation", EURASIP Journal on Image and Video Processing. (Submitted to the Journal)

5. Ammara Masood, Adel Al-Jumaily, "Adaptive Differential Evolution based feature Selection with Semi Advised SVM for Partially Labeled Medical Data Analysis", Expert Systems with Applications. (Submitted to the Journal)

Book chapter

6. Ammara Masood, Adel Al-Jumaily, "Adaptive Differential Evolution based Feature Selection and Parameter Optimization for Advised SVM Classifier ," ICONIP 2015, Part I, LNCS 9489, pp. 401–410, 2015.

7. Masood, Ammara Al-Jumaily, Adel, Adnan, Tariq, "Development of Automated Diagnostic System for Skin Cancer: Performance Analysis of Neural Network Learning Algorithms for Classification, S. Wermter et al. (Eds.): ICANN 2014, Lecture notes in Computer Science LNCS 8681, pp. 837–844, 2014, Springer International Publishing Switzerland 2014.

8. Ammara Masood, Adel Al-Jumaily, and Khairul Anam, "Texture Analysis Based Automated Decision Support System for Classification of Skin Cancer Using SA-SVM", C.K. Loo et al. (Eds.): ICONIP 2014, Part II, Lecture notes in Computer Science LNCS 8835, pp. 101–109, 2014, Springer International Publishing Switzerland 2014.

9. Ammara Masood, Adel Ali Al-Jumaily, and Yashar Maali, "Level Set Initialization Based on Modified Fuzzy C Means Thresholding for Automated Segmentation of Skin Lesions", M. Lee et al. (Eds.): ICONIP 2013, Part III, Lecture notes in Computer Science (LNCS) 8228, pp. 341–351, 2013.

Conferences

10. Ammara Masood, Adel Al- Jumaily, **"**Differential evolution based advised SVM for histopathalogical image analysis for skin cancer detection", 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2015 Milan, Italy, 25-29 Aug. 2015, pp.781 – 784.

11. Ammara Masood, Adel Al- Jumaily, Khairul Anam, "Self-Supervised Learning Model for Skin Cancer Diagnosis", 7th Annual International IEEE EMBS Conference on Neural Engineering Montpellier, France, 22 - 24 April, 2015, pp.1012-1014**.**

12. Ammara Masood, Adel Al-Jumaily, and Yee Mon Aung , "Scaled Conjugate Gradient based Decision Support System for Automated Diagnosis of Skin Cancer", 11th IASTED

International Conference on Biomedical Engineering, paper ID 818-020, June 23 – 25, 2014, Zurich, Switzerland.

13. Ammara MAsood, Adel Al Jumaily, "SA-SVM based automated diagnostic System for Skin Cancer", 3rd International Conference on Image, Vision and Computing, Paris 2014.

14. Ammara Masood, Adel Ali Al-Jumaily, "Integrating soft and hard threshold selection algorithms for accurate segmentation of skin lesion", Middle East Conference on Biomedical Engineering (MECBME), February 2014, pp. 83-86.

15. Ammara Masood, Adel Ali Al Jumaily, Azadeh Noori Hoshyar, Omama Masood, "Automated Segmentation of Skin Lesions: Modified Fuzzy C mean Thresholding Based Level Set Method", Proceeding of 16th International Multitopic Conference INMIC 2013, December 2013, pp. 201-206.

1.2.5 Summary

This chapter was aimed to provide an introduction to the thesis aims and motivations and brief overview of the research contributions of this thesis along with details of the resulting publications from the work presented in this thesis.

Chapter 2

Overview on Skin Cancer and Diagnosis Methods

2.1 Introduction

This chapter reviews some medical aspects of skin cancer. It also presents general information about the structure of skin and types of skin cancer as well as some information on the diagnosing methods used by medical experts. This review does not aim at listing what have already been published, however it aims at providing general details and concepts about the definitions and related issues of the skin cancer from medical point of view.

2.2 Human Skin

The skin is the largest organ of the body. It provides protection against heat, sunlight, injury, and infection. Skin also helps in controlling body temperature and stores water, fat, and vitamin D.

2.2.1 Types of Skin

There are two main kinds of human skin as shown in Figure 2.1. One is glabrous skin (non-hairy skin), which is found on the palms and soles, it is grooved on its surface by continuously alternating ridges and sulci, which has unique configurations for individuals, known as dermatoglyphics. It is characterized by a thick epidermis which is divided into several well-marked layers, including a compact stratum corneum, by the presence of encapsulated sense organs within the dermis, and by a lack of hair follicles and sebaceous glands.

Second type is the hair-bearing skin which has both hair follicles and sebaceous glands but lacks encapsulated sense organs. There is also wide variation in skin types between different body sites.

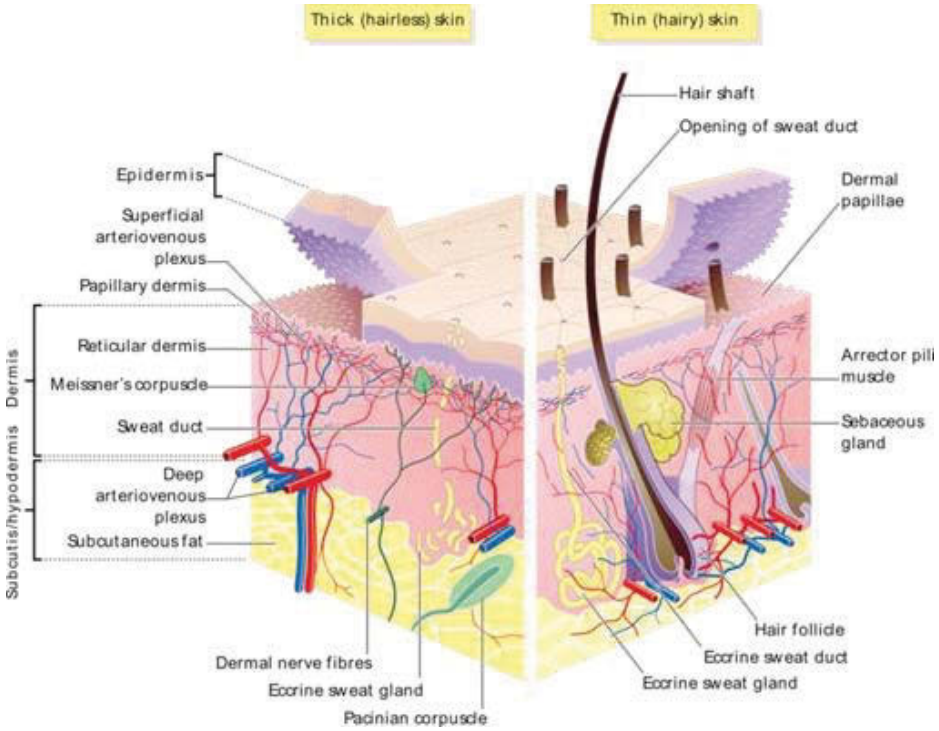


Figure 2.1 Types of Human Skin (Source en.wikipedia.org)

2.2.2 Layers of Human Skin

Skin is majorly composed of three primary layers as shown in Figure 2.2, 2.3.

- The epidermis, which provides waterproofing and serves as a barrier to infection
- The dermis, which serves as a location for the appendages of skin including hair, sweat glands and sebaceous glands
- The hypodermis (subcutaneous adipose layer)

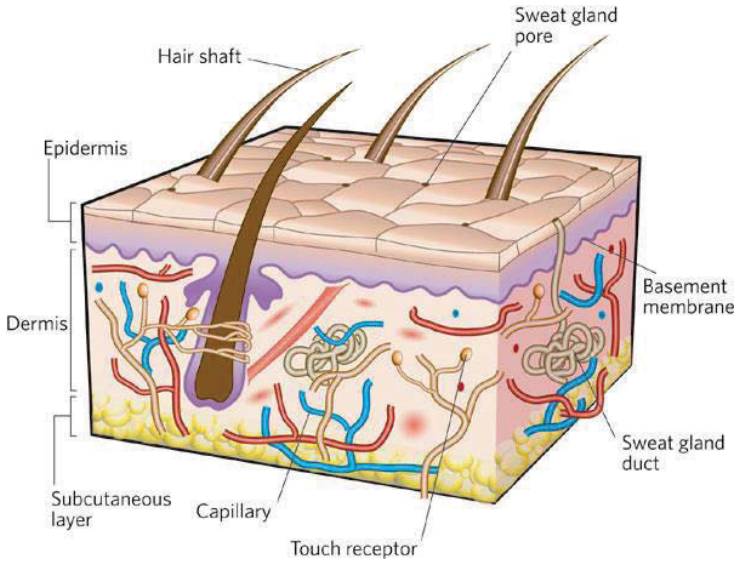


Figure 2.2 Layers of Human Skin (Source: silver-botanicals.com)

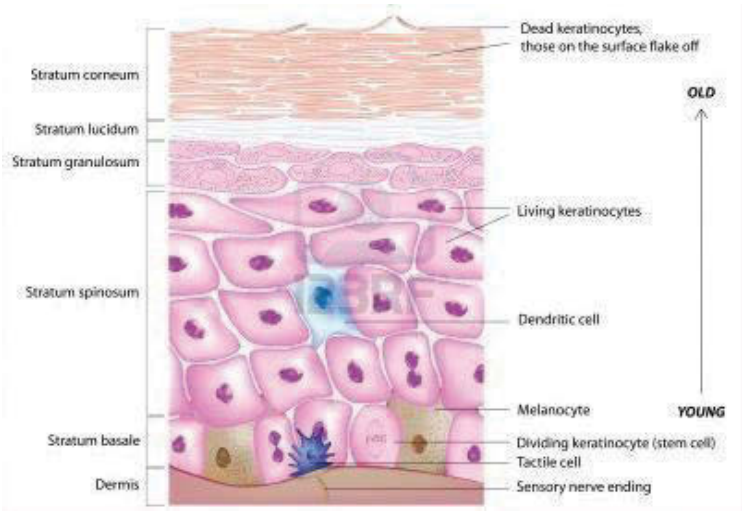


Figure 2.3 Structure of the Epidermis (Source: silver-botanicals.com)

2.3 Cancer

It is the term normally given to a large group of diseases that vary in type and location. Cancer can be defined as a class of diseases or disorders characterized by uncontrolled division of cells and the ability of these cells to invade other tissues, either by direct growth into adjacent tissue through invasion or by implantation into distant sites by metastasis in which cancer cells are transported through the bloodstream or lymphatic system.

Cancer is the term used to describe a lot of different diseases which include malignant tumours, leukaemia (a disorder of white blood cells), sarcoma of the bones, Hodgkin's disease and non-Hodgkin's lymphoma (affecting the lymph nodes) in which uncontrolled cell growth threatens the rest of the body.

There are more than 100 different types of cancer, but the most common five types are: bowel cancer, breast and prostate, melanoma and lung cancer (excluding non-melanoma skin cancer) and it forms 60% of all the cases. In Australia, the most common cancers in men are prostate, colorectal, and lung cancers and melanoma, but in women breast cancer, colorectal cancer, melanoma and lung cancer are more common.

2.3.1. Development stages of Cancer disease

Cancer cells are very similar to cells of the organism from which they originated and have similar (not identical) DNA and RNA. This is the reason that they are not very often detected by the immune system, in particular if it is weakened.

These cells usually have an increased ability to divide rapidly and their number of divisions is not limited by telomeres on DNA (a counter system to limit number of divisions to 40-60). Thus, rather than dividing in a controlled and programmed manner, the cell continues to divide and multiply abnormally, until a detectable lump or tumor develops. This can lead to the formation of large masses of tissue and in turn may lead to disruption of bodily functions due to destruction of organs or vital structures. It is one of the principal causes of death in developed countries. Stages of cancer development and metastasis are illustrated in Figure 2.4.

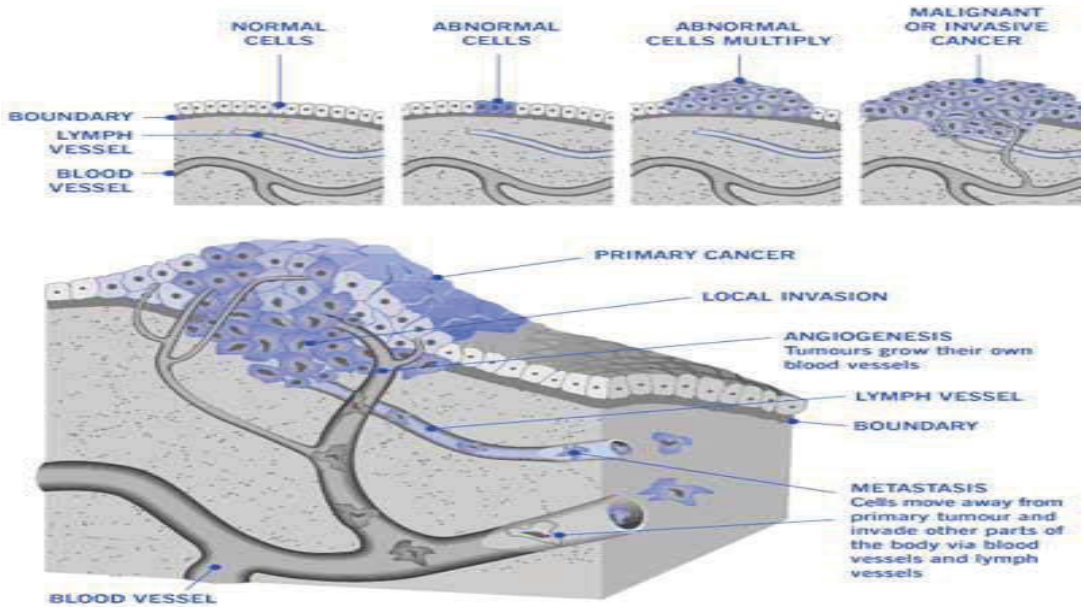


Figure 2.4 Stages of Cancer Development and Metastasis (Source en.wikipedia.org)

2.4 Skin Cancer

Like all body tissues, the skin is also made of tiny building blocks called cells. These cells can sometimes become cancerous, from exposure to ultraviolet (UV) radiation or certain other reasons like family history etc.

The abnormal growths on the skin can be categorized as benign (not cancer) or malignant (cancer). Benign growths are not as harmful as malignant growths.

Benign growths (such as moles) are rarely a threat to life and generally can be removed and usually do not grow back again. Benign tumors do not invade the tissues around them and thus do not spread to other parts of the body.

On the other hand, Malignant growths (such as melanoma, basal cell cancer, or squamous cell cancer), may be a threat to life. They can grow back even after removal. They have the tendency to invade and damage nearby organs and tissues and thus may spread to other parts of the body.

The top layer of skin contains three different types of cells: basal cell cancer, squamous cell cancer and melanocyte. These cells can become cancerous so the skin cancer can be mainly divided into three categories:

• Basal Cell Skin Cancer

• Squamous Cell Skin Cancer

• Melanoma

Figure 2.5 shows layers of Skin, i.e. the Epidermis and Dermis (At the top, the close up shows a squamous cell, basal cell, and melanocyte)

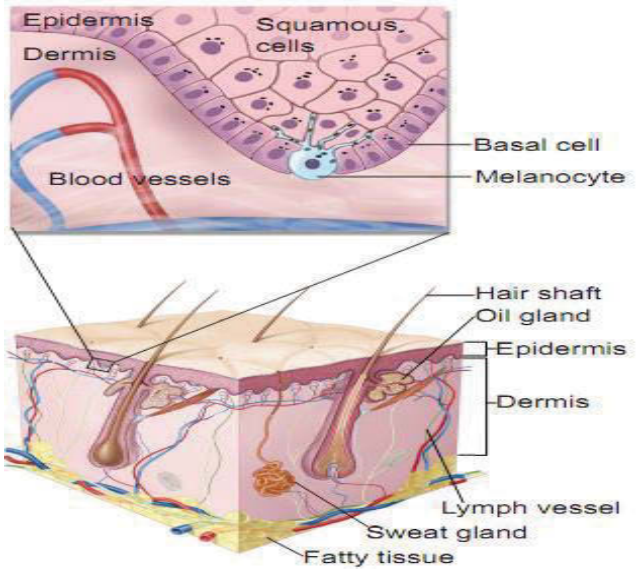


Figure 2.5 Details in Skin Layers (Source en.wikipedia.org)

Skin cancers are named for the type of cells that become malignant (cancer). A brief description of the first two types of cancer is provided below, while the details about melanoma is in next section, as that is the major type of cancer related to this research.

2.4.1 Basal Cell Carcinoma (BCC)

This is the most common and least dangerous skin cancer (BCC). It appears as a lump or scaly area and can be red, pale or pearly in colour. It grows slowly – usually on the head, neck, near eye or upper torso and can become ulcerated as it grows. One form of BCC is shown in Figure 2.6.

Figure 2.6 Basal Cell Carcinoma (BCC) (Source: Skinsite.com)

2.4.2 Squamous Cell Carcinoma (SCC)

The (SCC) cancer grows over a period of weeks or months and may spread to other parts of the body if not treated promptly. It occurs most often (but not only) on areas exposed to the sun. This can include the head, neck, handstand forearms. This cancer looks like thickened, red, scaly pots. One form of BCC is shown in Figure 2.7.

Figure 2.7 Squamous Cell Carcinoma (SCC) (Source: medicalook.com)

2.5 Melanoma

Melanoma begins in melanocytes (pigment cells). Most melanocytes are in the skin so Melanoma can often occur on any skin surface. In men, it is often found on the skin of the head, neck or between the shoulders and the hips. In women, it is often found on the skin on the lower legs or between the shoulders and the hips. Melanoma is rare in people with dark skin, and for them it is usually found under the fingernails, under the toenails, on the palms of the hands, or on the soles of the feet. Melanoma appears as a new spot or as an existing spot, freckle or mole that changes colour, size or shape with time. It usually has an irregular, smudgy outline and often has more than one colour. Some form of melanoma are depicted in Figure 2.8.

Figure 2.8 Melanoma (Source: melanomaknowmore.com)

Melanocytes in the eye, respiratory system, nervous system, even on cortex of brain and mucous membranes (e.g. lining of the mouth and nasal passages) can also become cancerous. It is the most dangerous type of skin cancer and can grow quickly and develops over weeks to months. If caught early, it is usually curable, but if it spreads to other parts of the body, it can become very difficult to cure. Even small melanomas can spread through the body via blood and lymphatic vessels. The cancer can spread to organs such as the lymph glands, lungs, liver and brain. Infact it can spread to almost any organ in the body. This is why melanoma is so dangerous. Once spread, it can invade and destroy the function in these organs which consequently leads to death in most cases.

The most important feature of a melanoma is its thickness (stage 0 is less than 0.1mm, stage I less than 2mm, stage II greater than 2mm, stage III spread to lymph nodes and stage IV distant spread). If distant spread is suspected, CT scans of the chest abdomen and pelvis are performed. The blood test LDH can sometimes be useful to assess

metastatic disease. Surgery can be curative for thin melanomas and requires that the melanoma be removed with at least 1-2cm of normal skin around it.

The main causes of Melanoma include sunburn[12], increased numbers of unusual moles[13]; depressed immune systems; a family history (in 10%, some having mutations in genes); fair skin and a previous melanoma history. [14-16], female has better survival rate as compared to male [17, 18]. Some more details for different types of skin melanoma [19-21].

2.6 Detection & Diagnosis Methods by Medical Experts

There are a number of methods that are used by the doctors for diagnosis of skin cancer that can depend on their expertise as well as availability of resources. In this section a brief overview of most commonly used skin imaging methods is provided with more emphasis on dermoscopic image analysis. A number of diagnosis algorithms used by the medical experts for skin surface image analysis are also described followed by the overview of the diagnosis based on histopathological analysis.

2.6.1 Skin Lesion Imaging Methods

It was found that there are number of various imaging methods of skin lesions.

- **2.6.1.1 Traditional Photography**

The simplest visualization method is photography. This method gives only top layer skin image.

- **2.6.1.2 Dermoscopy Imaging Techniques**

In order to get deeper layer image oil immersion is used. It reduces reflections of surface and brightens the image of epidermis – the second skin layer. The better results are reached when photos are made with polarized light source from top layer of skin. Then it is easier to estimate the lesion structures like dots, globules, nets that are the major indicators to melanoma diagnosis. Figure 2.12 shows example of Pigmented Skin Lesion Left: Traditional Imaging Technique. Right: Dermoscopy Imaging Technique

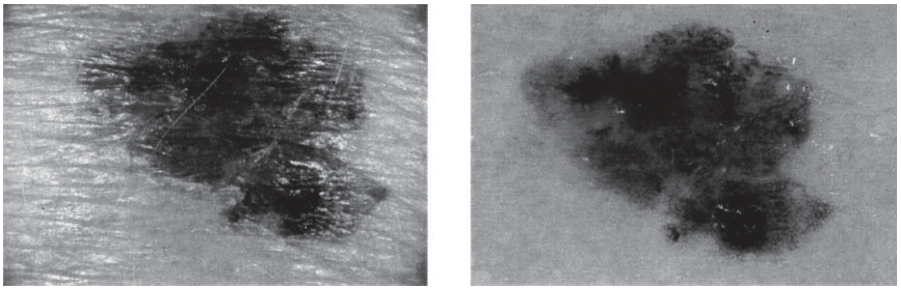


Figure 2.9 Traditional Imaging Vs. Dermoscopy

- **2.6.1.3 Multi Spectral Photography**

Other interesting solution of getting more information from skin is using multi spectral photography. It uses narrow frequency band of light illumination. Those images give information about consistence and concentration of absorbers and reflectors in the skin. The idea is that different pigment of skin absorbs different light wave, determining the colour of our skin. When those photos are made with range of light waves, the reflectance frequency characteristics of skin can be calculated. Comparing skin pigment consistency to normal skin characteristic can be help to make the diagnostic decisions.

- **2.6.1.4 Ultrasound Visualization**

Ultrasound visualization is usually used to measure depth of melanoma and if there is no melanoma practically there are no differences. When doctor diagnoses melanoma, then he uses high frequency ultrasound (over 30 MHz) to measure penetration depth in order to make correct cut during surgery.

2.6.2 Dermosopic Diagnosis of Skin Lesions

Dermoscopy is a non-invasive diagnostic technique used in dermatology for the in vivo observation of pigmented skin lesions. Dermoscopy is an imaging technique that provides a more direct link between biology and distinct visual characteristics. It is known as epiluminescence microscopy (ELM).

This diagnostic tool allows for a better visualization of surface and subsurface structures and thus permits the recognition of morphologic structures which are not visible by the naked eye. This technique has opened a new dimension of analysing the clinical morphologic features of pigmented skin lesions.

Optical magnification and either liquid immersion and low angle-of-incidence lighting or cross-polarized lighting is used to make the contact area translucent. This makes the subsurface structures more easily visible when compared to conventional macroscopic (clinical) images. Dermoscopy helps in the identification of dozens of morphological features such as pigment networks, dots/globules, streaks, blue-white areas, and blotches. This provides greater differentiation between difficult lesions such as pigmented Spitz nevi and small, clinically equivocal lesions and thus reduces the screening errors.

Dermatoscopy images are much more detailed than conventional macroscopic (clinical) images. Macroscopic images are merely images of skin lesions seen under a magnifying glass. Figure 2.13 shows a comparison of a macroscopic (Left) and a dermatoscopic image (Right) of the same skin lesion.

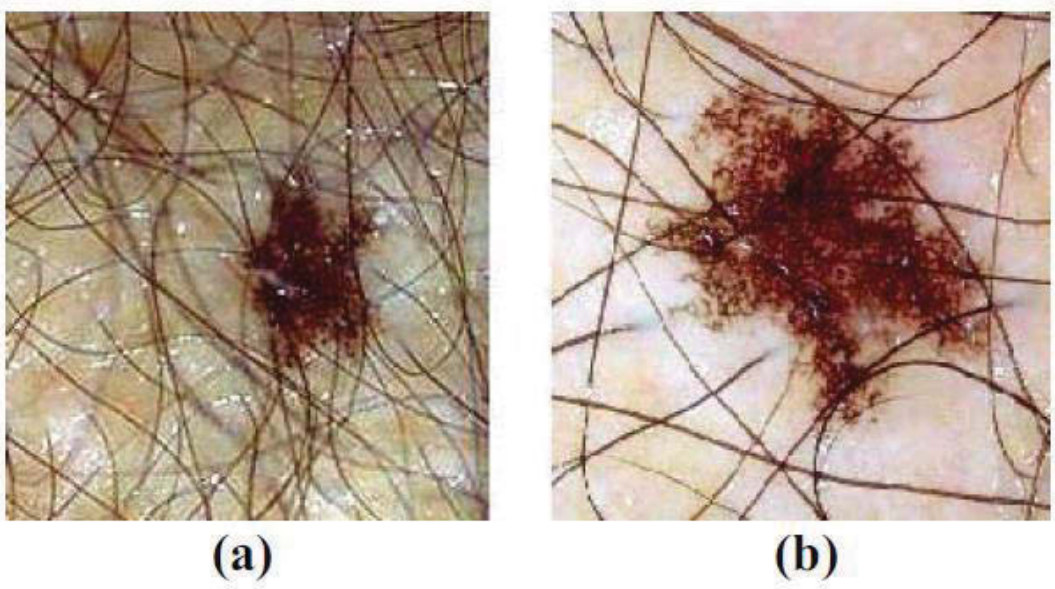


Figure 2.10 Macroscopic Image (a) Vs. Dermatoscopic Image (b)

Studies have shown that dermatoscopic diagnosis has an accuracy of 75% to 97% while for macroscopic diagnosis this range is 65% to 80% [22, 23]. However, it has also been noticed that dermatoscopy may actually lower the diagnostic accuracy in the hands of inexperienced dermatologists [23]. Therefore, due to the lack of reproducibility and subjectivity of human interpretation, the development of computerized techniques is of utmost importance [24].

Successful treatment of melanoma, by surgical removal depends upon the early detection of the lesion. The challenge is to diagnose and remove all malignant lesions at

an early stage and at the same time minimising the unnecessary removal of benign lesions. However, the differentiation of early melanoma from other non-malignant pigmented skin lesions is not trivial even for the experienced dermatologists. In several cases primary care physicians underestimate melanoma in its early stage [25].

Visual inspection with the naked eye has a relatively low sensitivity in detecting early melanoma In this context, several non-invasive diagnostic modalities, such as dermoscopy, total body photography, and reflectance confocal microscopy (RCM), have emerged in recent years that are aimed at increasing diagnostic accuracy. The main developments in this field are the integration of dermoscopy and digital photography into clinical practice.

Dermoscopy (also known as dermatoscopy or epiluminescence microscopy) helps in the visualization of anatomical structures within the epidermis and papillary dermis that cannot be visualised by the naked eye. This is achieved with the use of a hand held dermatoscope to magnify the skin surface and reduce the refraction of light by the corneal layer. It has been reported that dermoscopic training for primary care physicians can improve their ability to correctly refer individuals with suspicious lesions and decrease the rate of excision or referral in benign skin lesions.

2.6.3 Diagnosis Techniques

Various diagnostic algorithms (ABCD rule, Menzies method, seven-point checklist, and three-point check list) have been proposed in this regard to help assess the structures and patterns seen in dermoscopy.

- **2.6.3.1 ABCD-E Rule**

The ABCD rule of dermoscopy was the first method of the so-called melanoma algorithms used in dermoscopy. This model was described by Stolz and colleagues in 1994 [26]. For the calculation of ABCD score the criteria of asymmetry (A), abrupt cut-off of the pigment pattern at the border (B), different colors (C), and different structural components (D) are assessed to yield a semi-quantitative score. Calculation of ABCD score is illustrated in Table 2.1 and application of this rule is shown in Figure 2.14 and 2.15. The ABCD rule has been proved to be useful in:

(1) Early detection of malignant melanoma

(2) Discrimination between benign and malignant lesions

(3) Selection of lesions for excision in patients with numerous typically appearing nevi

(4) Monitoring of certain malignant nevi that for cosmetic or patient preference were not removed

Table 2.1 Calculation of ABCD score of Dermatoscopy

Criteria	Description	Possible Scores	Weight	Min/Max
Asymmetry	In 0, 1, or 2 axes; assess not only contour, but also colors and structures	0-2	x 1.3	0 - 2.6
Border	Abrupt ending of pigment pattern at the periphery in 0-8 segments	0-8	x 0.1	0 - 0.8
Colour	Presence of up to six colors 1-6 White, Red, light-brown, dark-brown, blue-gray, black	1-6	x 0.5	0.5 - 3
Differential Structure	Presence of Network, Structureless or homogeneous areas, Streaks, Dots, Globules	1-5	x 0.5	0.5 - 2.5
Total Dermatoscopy Score				1 - 8.9

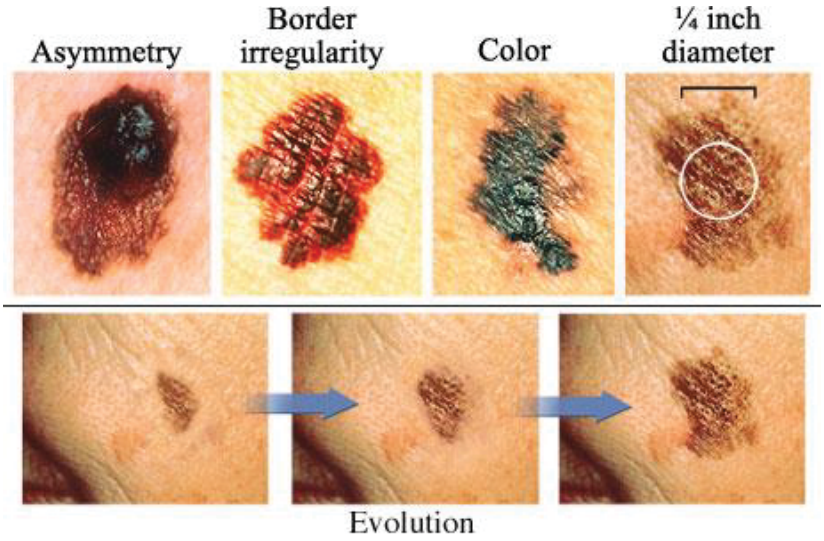


Figure 2.11 ABCD-E rule for Diagnosis of Melanoma

Total Dermatoscopy Score

For calculating TDS the individual scores are multiplied by different weight factors obtained by multivariate analysis. The formula for TDS and the resultant interpretations are provided as follows.

Formula for Calculating TDS

[(A score x 1.3) + (B score x 0.1) + (C score x 0.5) + (D score x 0.5)]

The interpretation of TDS results is illustrated in Table 2.2.

Table 2.2 Interpretation of TDS Results

Total Dermoscopy Score	Interpretation
< 4.75	Benign melanocytic lesion
4.8-5.45	Suspicious lesion; close follow-up or excision is recommended
>5.45	Lesion is highly suspicious for melanoma
False Positive Score (>5.45) Sometimes observed in	Reed and Spitz nevi Clark nevus with globular pattern Congenital melanocytic nevus Melanocytic nevus with exophytic papillary structure

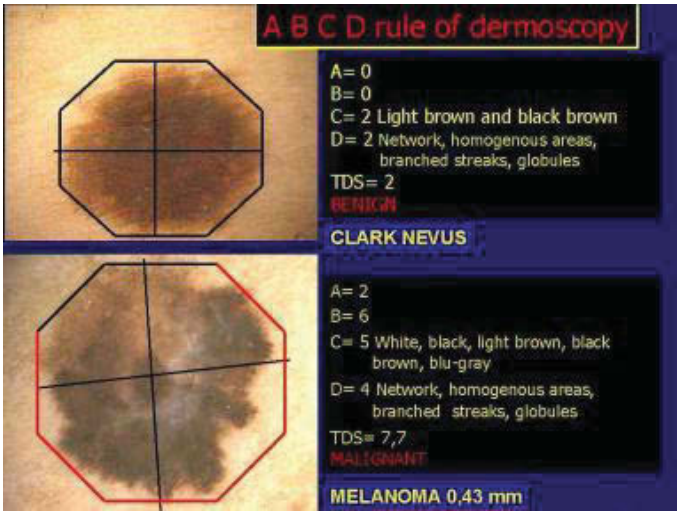


Figure 2.12 Application of ABCD rule [27]

- **2.6.3.2 Pattern Analysis**

The pattern analysis method seeks to identify specific patterns, which may be global (reticular, globular, cobblestone, homogeneous, starburst, parallel, multicomponent, nonspecific) or local (pigment network, dots/globules/ moles, streaks, blue-whitish veil, regression structures, hypopigmentation, blotches, vascular structures). Some of them are explained below.

Blue-White Veil: One of the useful features in dermoscopic diagnosis is the blue-white veil (irregular, structure less areas of confluent blue pigmentation with an overlying white "ground-glass" film) which is mostly associated with invasive melanoma [28]. Image showing melanoma with blue-white veil is shown in Figure 2.16.

Figure 2.13 Melanoma with a blue-white veil (a) clinical image (b) dermoscopy image

Pigment Network: A pigment network can be classified as either Typical or Atypical, where a working definition of a typical pigment network (TPN) is "a light-to-dark-brown network with small, uniformly spaced network holes and thin network lines distributed more or less regularly throughout the lesion and usually thinning out at the periphery". For an atypical pigment network (APN) we use the working definition "a black, brown or grey network with irregular holes and thick lines". The goal is to automatically classify a given image to one of three classes: Absent, Typical, or Atypical. Following Figure 2.17 exemplifies these 3 classes. These definitions can be used to subdivide the structure into the darker mesh of the pigment network (which is referred as the 'net') and the lighter colored areas the net surrounds (which is referred as the 'holes'). After identifying these substructures one can use the definitions above to derive a several structural, geometric, chromatic and textual features suitable for classification.

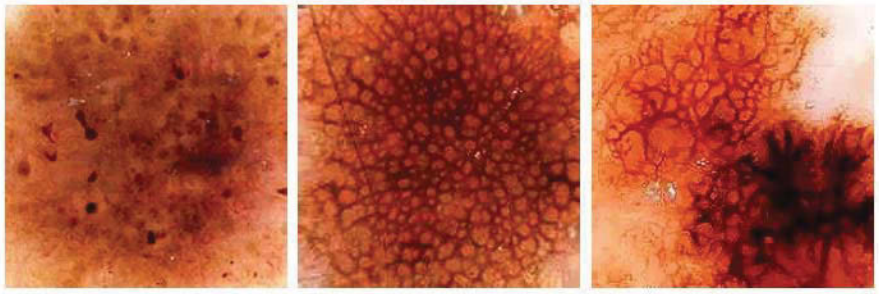


Figure 2.14 3 classes of pigment network: Absent, Typical and atypical respectively

Streaks: Streaks is a term used interchangeably with radial streaming or pseudopods. Radial streaming is a linear extension of pigment at the periphery of a lesion radially arranged linear structures in the growth direction, and pseudopods represent finger-like projections of dark pigment (brown to black) at the periphery of the lesion. In order to ensure accurate recognition, streaks are numerated only when at least 3 near linear and parallel structures are clearly visible. Streaks are local dermoscopy features of skin lesions; however they can correlate with a global pattern of skin lesions called a starburst pattern if symmetrically arranged over the entire lesion. Mathematical definition of streaks: The above clinical definition is translated to mathematical concepts with justified parameters to be captured by image processing techniques:

1) Streaks are 3 or more linear structures co-radially oriented in the boundary which is a contour with the thickness equal to 1/3 of the minor axis of the lesion.

2) Streaks are darker than their neighbourhood.

3) Streaks are shorter than the 1/3 of the minor axis of the lesion and they should be longer than 1% of the major axis.

4) Streaks do not branch and their curvature is smaller than one. Following figure shows examples of lesions with no streaks (Absent), regular (Present), and irregular (Present) streaks. (a) shows a lesion without streaks, (b) illustrates a lesion with complete symmetric regular streaks pattern called starburst (present) and (c) shows a melanoma lesion with irregular streaks and partial pattern. Following Figure 2.17 exemplifies absent , regular and irregular streaks.

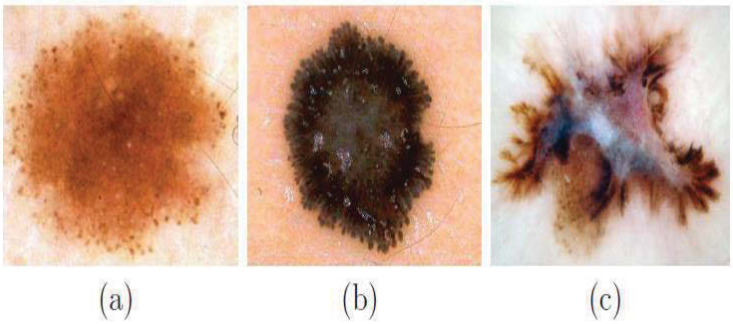


Figure 2.15 Examples of Streaks (a) absent, (b) regular and (c) irregular

- **2.6.3.3 The Seven-Point Checklist**

The seven-point checklist is a diagnostic algorithm that provides a quantitative scoring system as shown in Table 2.3. This method was developed for the dermoscopic diagnosis of melanoma by Argenziano [29]. As this method requires the identification of only seven dermoscopic criteria, thus it enables even less experienced clinicians to use the model following a relatively short learning curve.

Calculation of the Seven Point Score

The differences between melanomas and nevi can be evaluated by a univariate statistical test. The significant variables were used for stepwise logistic regression analysis to determine their different diagnostic weight in the diagnosis of melanoma and were expressed here by odds ratios. Using the odds ratios calculated with multivariate analysis, a score of 2 is given to the three criteria with odds ratios of >5, called ‘major’ criteria, and a score of 1 to the four criteria with odds ratios of <5 which is called ‘minor’ criteria. By simple addition of the individual scores a minimum total score of 3 is required for the diagnosis of melanoma and total score of less than 3 indicates non-melanoma.

Table 2.3 Seven-Point Checklist for the dermoscopic differentiation between benign melanocytic lesions and melanoma

Criterion	Definition	Odd ratios [a]	Seven point Score [b]
Major criterion			
Atypical pigment network	black, brown or gray network with irregular holes and thick lines	5.19	2
Blue-white veil	irregular, structureless area of confluent blue pigmentation with an overlying white 'ground-glass' film. The pigmentation cannot occupy the entire lesion and usually corresponds to a clinically elevated part of the lesion	11.1	2
Atypical vascular pattern	linear-irregular or dotted vessels not clearly seen within regression structures	7.42	2
Minor criteria			
Irregular streaks	brown to black, bulbous or finger-like projections irregularly distributed at the edge of a lesion. They may arise from network structures but more commonly do not	3.01	1
Irregular dots/globules	black, brown, round to oval, variously sized structures irregularly distributed within the lesion	4.90	1
Irregular blotches	black, brown and/or gray structureless areas asymmetrically distributed within the lesion	2.93	1
Regression structures	white scar-like depigmentation and/or blue pepper-like granules usually corresponding to a clinically flat part of the lesion	3.89	1

[a] Odds ratios measuring the capacity of each criterion to increase the probability of melanoma diagnosis.

[b] The score for a criterion is determined on the basis of the odds ratio: >5 (score 2), and

<5 (score 1). Simply add the scores of each criterion that is present within a pigmented lesion

With regard to texture features, skin lesions have been classified by physicians manually into three classes as shown in Figure 2.19, namely, reticular, globular, and homogeneous.

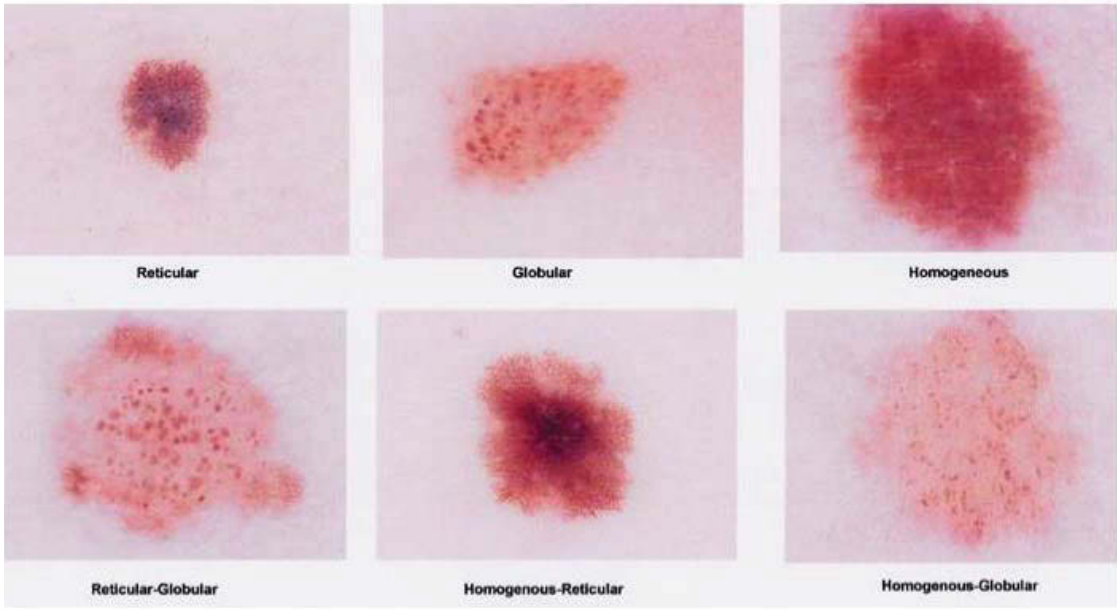


Figure 2.16 Texture characteristics of skin lesions [30]

Lesions having only one or two of the above features are found to be benign with a probability of 90% and 70%, respectively, while lesions possessing all three features have a 90% probability of being malignant and should definitely be excised. In other cases, whereby only partial features are detected, lesions have a 70% probability of being benign. Despite the high diagnostic value of texture information, it is very difficult to utilize it in automated skin cancer detection due to complexity of the texture and high noise and interference of the obtained lesion images [30]. Figure 2.20 shows application of 7 point checklist for classification of clark nevus and melanoma.

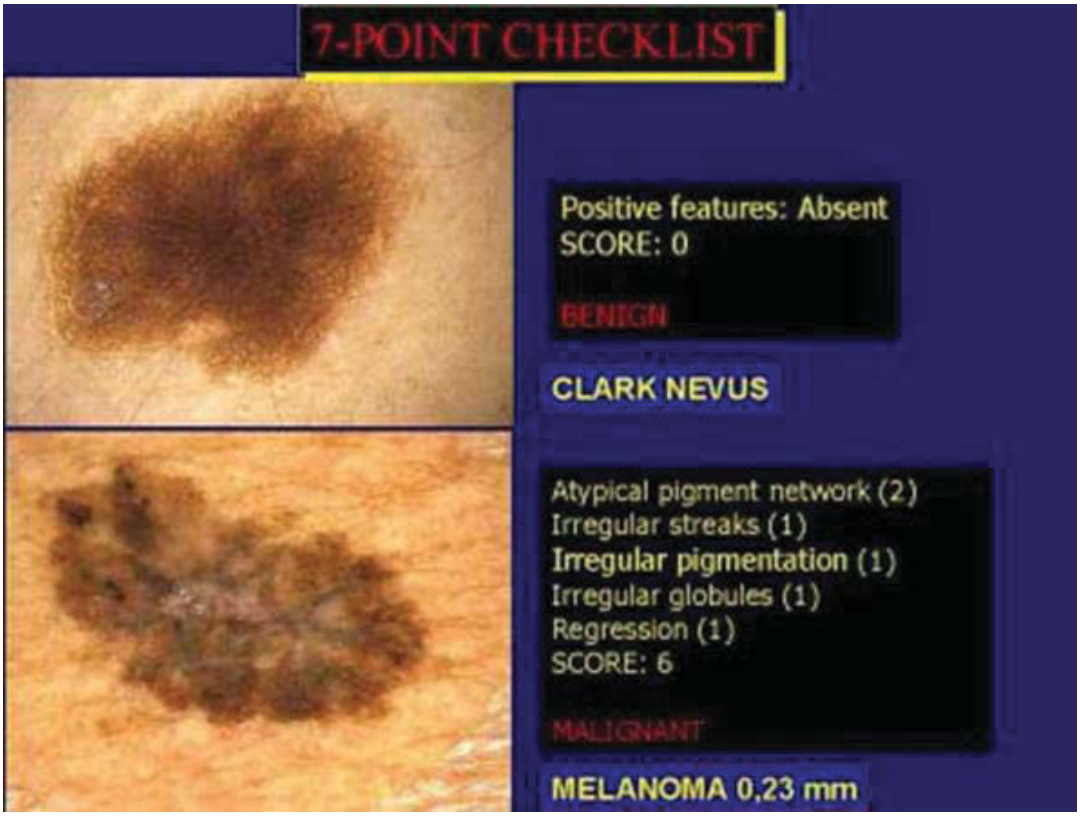


Figure 2.17 Classification using 7-point checklist [27]

- **2.6.3.4 The Menzies method**

Pattern analysis requires significant expertise for its use. A simplified dermoscopy method for the diagnosis of melanoma was developed by Menzies [31] for inexperienced clinicians. This method gives a sensitivity of 92% and specificity of 71% for the diagnosis of melanoma. An atlas has been produced that allows even inexperienced clinicians to learn the method [32]. This enabled primary care physicians to increase their sensitivity for the diagnosis of melanoma by 38% as compared to the standard clinical visualization[33].

The Menzies method is briefly described in the Table 2.4. For a melanoma to be diagnosed, it must have neither of the two morphological negative features and one or more of the nine positive features. Figure 2.21 shows application of Menzies Method for classification of clark nevus and melanoma.

Table 2.4 Menzies Method

Negative features (in melanoma, neither can be found)
Symmetry of pigmentation pattern (lesion axial symmetry, lesion colour symmetry)
Presence of only a single colour
Positive features (at least one feature found)
Blue-white veil
Multiple brown dots
Pseudopods
Radial streaming
Scar-like depigmentation
Peripheral black dots/globules
Multiple (5–6) colours
Multiple blue-grey dots—'peppering'
Broadened network

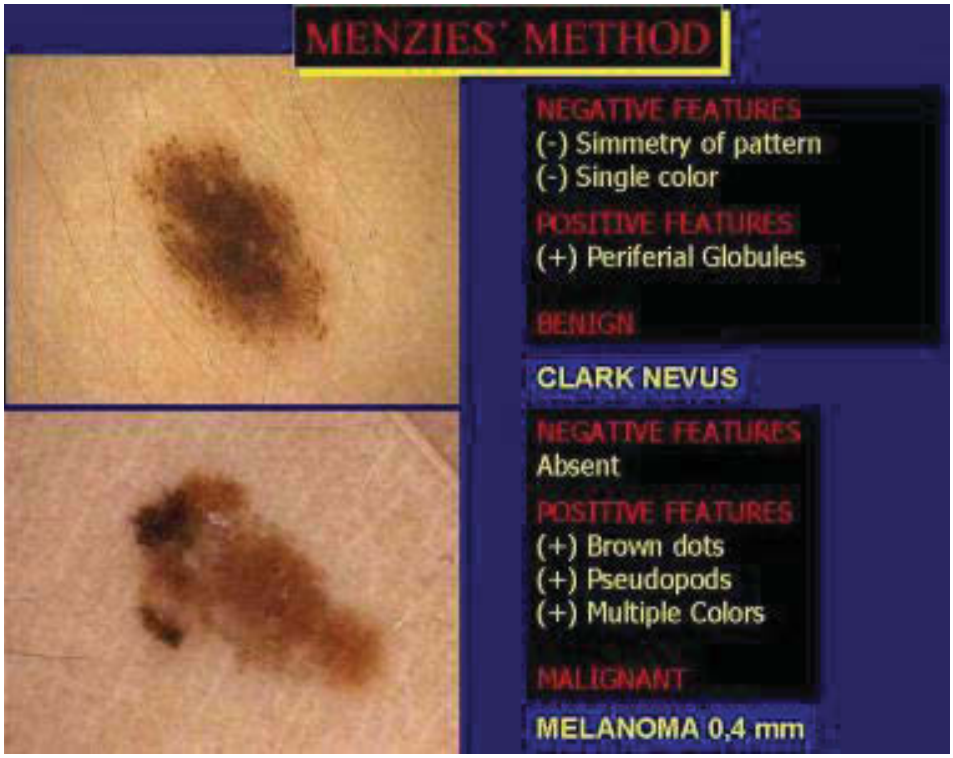


Figure 2.18 Application of Menzies method to melanocytic lesions [27]

- **2.6.3.5 The Three-Point Checklist**

The three-point checklist was developed for inexperienced dermoscopists. This is meant to encourage them to use dermoscopy by teaching them a simplified algorithm for the evaluation of pigmented skin lesions. General practice physicians can easily be taught dermoscopy and the three-point checklist, which can be utilized as a screening tool to determine whether a given pigmented lesion should be further evaluated by a dermatologist.

The three-point checklist [34] features three criteria that are found to be particularly important in the differentiation between malignant and benign pigmented skin lesions by participants in the Consensus Net Meeting on Dermoscopy[35].

Table 2.5 Definition of Three Point Checklist

Criterion	Description
Asymmetry	asymmetrical distribution of colours and dermoscopic structures in any one axis
Atypical network	pigmented network with irregular holes and thick lines
Blue-white structures	any type of blue and/or white colour including white scar-like depigmentation, blue-whitish veil and blue pepper-like granules (regression structures)

In the three-point checklist shown in Table 2.5 the presence of any two of the following three criteria: asymmetry; atypical pigment network; and blue-white structures indicates that the lesion under investigation may be a melanoma. Figure 2.22 shows application of 3 point checklist for diagnosis of melanoma.

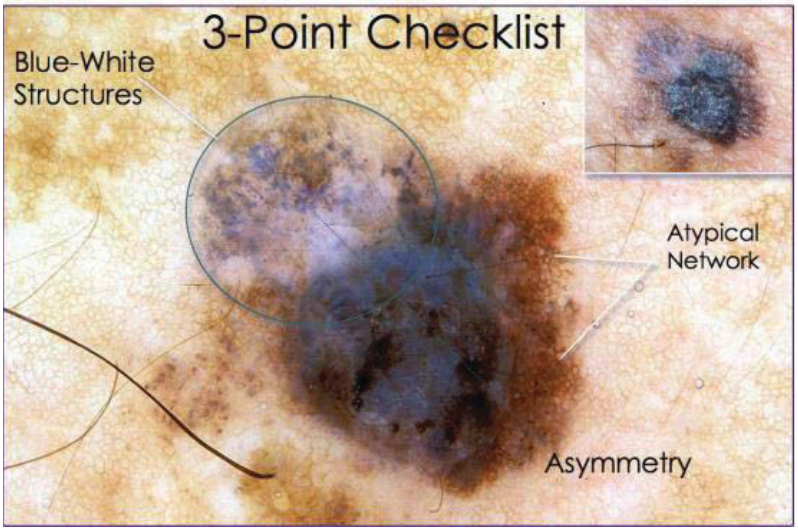


Figure 2.19 Diagnosis of Melanoma using Three-point Checklist
(Source: dermoscopy-ids.org)

Use of the three-point checklist demonstrates sensitivity (96%) for detecting melanoma that is comparable to the other algorithms and thus suggests that it may prove to be useful as a screening tool for non-experienced observers. The different dermoscopy classifications definitely have their own worthy internal coherence; however, the use of the different diagnostic scores can be affected by interobserver and intraobserver variability when only a single guideline is used for evaluation (i.e, limited qualitative and quantitative agreement). Furthermore, all of these classifications can prove to be very sensitive but not very specific; thus, they do not allow 100% accuracy. Although very useful to detect intraepidermal lesions, dermoscopy is limited in regard to nodular lesions or clearly dermal lesions, lesions without pigmentation, very dark lesions in which the amount of pigment does not allow the observation of ELM signs, and faintly pigmented seborrhoic warts.The efficiency of dermoscopy is closely related to an integrated diagnostic synopsis for trained clinicians. The user must think in global diagnostic terms when considering the accuracy of dermoscopy findings, independent of the methods used; a broader aim is to include case histories and clinical assessment. In fact, the information needed to be integrated by dermatoscopic evaluation are Anamnesis (personal and family background) ; Photo-type ; Previous sunburn history ; Complete analysis of the entire skin surface (with the intent of finding the so-called "ugly duckling sign") ; Careful evaluation of the timing of first appearance (and eventual widening) of the lesion being examined and/or modifications of previously existing lesions: A magnifying glass is to

be used to carefully observe and evaluate the shape, color, and dimensions of the lesion. Such a combination of the traditional clinical diagnostic procedures and dermoscopy allows better classification of suggestive melanocytic lesions and a significant reduction in the percentage of unneeded biopsies, but only in pigmented skin lesion centers.

2.6.4 Diagnosis based on Pathological Analysis

Since the 1960s, the clinical characteristics of melanoma, its histopathology and its biological basis have been the subject of intense study at pigmented lesion clinics in North America, Europe, and Australia[36]. The purpose of this was to bring to the attention of anatomic pathologists the essential characteristics of the pathology report for primary cutaneous melanoma in the modern era.

The Researchers on [37] [38] recognized the importance of anatomic compartment of attack to the melanoma. In 1970 the researcher on [39] described the use of an optical micrometre to specifically measure primary tumour thickness. These two parameters became the recognized prognostic variables available to histo-pathologists in predicting the biologic behaviour of melanoma [40]. Throughout the 1980s, advances in biology and pathology of melanoma made it apparent that tumours undergo specific growth phases. In particular, radial growth phase were proven unable of generating metastatic events unless getting worse was seen at the primary site. Further work carried out by Massachusetts General Hospital and the University of Pennsylvania established that the existence of microscopic satellites and brisk mitotic activity could modify the behaviour of a tumour.[36]

The skin is a very organized structure consisting of three main layers, called the Epidermis, the Dermis and the Hypodermis. According to pathologists, the easiest ways is by understanding the structure of epidermis layer. So our primary task is to accurately segment the skin tissue image into dermal and epidermal layers. Pathologists use microscope to perform histo-pathological examination and provide diagnostic information based on their observations. Figure 2.23 shows manual segmentation performed by pathologists using ‘Aperio Image Scope’ software. [41] but there are many problems associated with manual segmentation, like lengthy time-consuming and depends on the skill and training of the pathologist doing the analysis. Therefore automatic segmentation is needed to bring some kind of consistent accuracy within the

process. In the technical report [42] author said the automated cancer diagnosis consists of three main computational steps: pre-processing, feature extraction, and diagnosis. The aim of the diagnosis step is (i) to distinguish benignity and malignancy or (ii) to classify different malignancy level by making use of extracted features. This step uses statistical analysis of the features and machine learning algorithms to reach a decision.

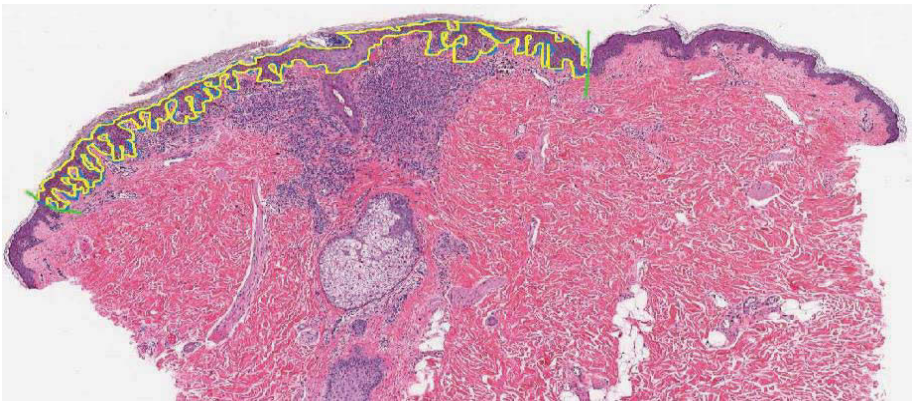


Figure 2.20 Manual Segmentation by Pathologist

- **2.6.4.1 Histo-pathological Images**

Malignant melanomas are the most serious form of skin cancer accounting for the majority of skin cancer related deaths. Histo-pathological images of skin tissues are analysed for detecting various types of melanomas. The automatic analysis of these images can greatly facilitate the diagnosis task for dermato-pathologists. The first and primary step in automatic histo-pathological image analysis is to accurately segment the images into dermal and epidermal layers along with segmenting other tissues structures such as nests and melanocytic cells which indicate the existence of cancer.

An efficient automatic segmentation procedure proposed for histo-pathological images of skin tissue. [41] the proposed technique is based on Orientation Sensitive Fuzzy C-Means (OS-FCM) algorithm along with other refinement techniques

- **2.6.4.2 Radial Growth Phase Melanoma**

The radial growth phase growing in a horizontal array as single cells and small nests predominantly in an intra-epidermal location.[36] The radial growth phase is incapable biologically of generating metastatic events, safe for specific and uncommon circumstances. In this example, there is pagetoid spread of fully transformed malignant

epithelioid melanocytes within the epidermis, see Figure 2.24 the basement membrane zone in concert with a level II component in the papillary dermis (i.e. Clark level II). Note that the dermal component comprises small nests of neoplastic melanocytes (red arrow Figure 2.24) with a similar cytology to those seen above the basement membrane zone which separates them from the overlying epidermal component (black arrow). The dermal nests of radial growth phase melanoma are smaller than those seen in the overlying epidermal component.

- **2.6.4.3 Vertical Growth Phase**

Clark in [43] described the early vertical growth phase and distinguished it from the radial growth phase by criteria that included, most importantly, the existence of a dominant nest within the papillary dermis. This expansible nest is larger than any nest within the epidermis or the surrounding dermis, see Figure 2.25. Clark [43] also suggested that the nodule should have 25–50 cells to aid in its identification, and that the cells comprising this nodule are cytologically separate from those of the intraepidermal component by virtue of their different characteristics, be they of shape, size, nuclear or cytoplasmic features, or the presence or absence of pigment.

The vertical growth phase means the point at which the melanoma becomes biologically capable of producing metastatic events. Note that there is a main able to expand dermal nest of neoplastic melanocytes (***arrow***) which is larger than any of the overlying the quiz at the dermal–epidermal junction or within the epidermis.

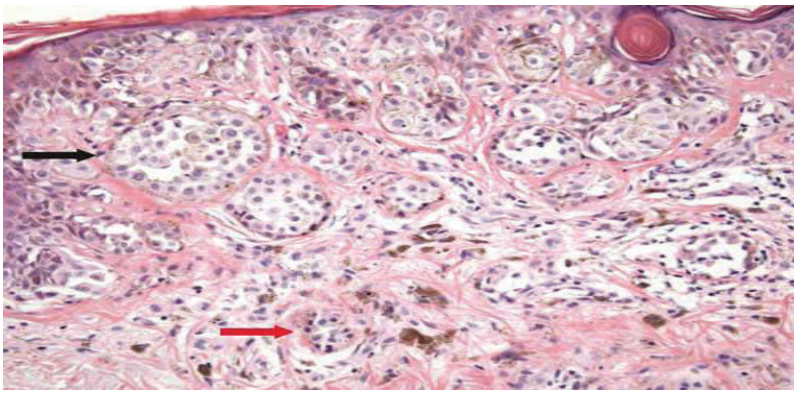


Figure 2.21 Radial Growth Phase of Melanoma

- **2.6.4.4 Histologic type (The microscopic structure) of melanoma**

Since the 1970s melanomas have been histologically sub classified as belonging to four major groups: superficial spreading, lentigo maligna, acral lentiginous, and

nodular melanoma.[38], [44], [45] Some observers suggest that the relatively greater survivorship seen in lentigo maligna melanoma vs the other histologic subtypes is reflective of smaller thickness at time of diagnosis.[43], [46], [47]

The lifetime risk for developing invasive disease in a lesion of lentigo maligna has been estimated to be only in the 5% range. [48] The early work of Fidler[49] suggests that those are the blood vessels capable of accommodating the large tumor bolus conducive to the establishment of the metastatic seed and so suggest that this is a critical factor in the mortality of any given acral melanoma.

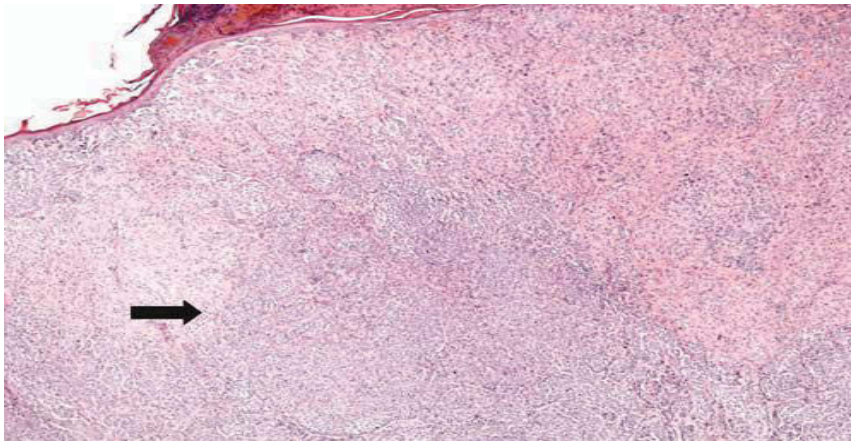


Figure 2.22 Vertical growth Phase of Melanoma

- **2.6.4.5 Thickness (Breslow)**

In the accumulated experience of several investigators over a long period of time, ***tumor thickness*** as measured by ***ocular micrometer*** has emerged as the most powerful predictor of outcome in primary cutaneous melanoma. [39], [50], [43], [51], [52] [53] [54] [55, 56] [57] [58] [59] [60] The Breslow measurement is taken from the epidermal surface or, in the event that the surface is ulcerated, from the base of the ulcer, and is made with a calibrated ocular micrometer. [39] The Breslow measurement is the most important means of prognosticating mucosal melanomas that lack the anatomic compartmentalization seen in the skin. [61] In most prognostic studies, the measured depth emerges as the most powerful independent factor for prediction of lymph node metastasis and survival.[62].

- **2.6.4.6 Microscopic Satellites**

Microscopic satellites, characterized by reticular dermal and/or sub-circular nodules of tumor greater than 0.05mm in width separated from main vertical growth

phase component are associated with lymph node metastasis, with decreased disease-free and overall survival. [43],[59] ,[63] [64] [65] [66] Clark et al [10] found a reduction of actuarial 8-year survival from[67] to 40% when satellites were identified. (Figure 2.26) It is suggested that this is a difficult criterion to assess, as tumor tongues radiating from main vertical growth phase nodule may mimic satellites due to artefacts of the sectioning [68].

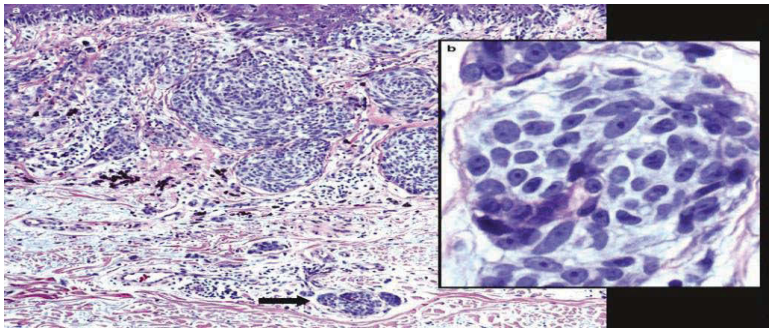


Figure 2.23 Microscopic Satellites (a) Shows neoplastic group is discontinuous (arrow) from the overlying vertical growth phase component. (b) shows the melanocytes must have a malignant cytomorphology

- **2.6.4.7 Blood Vessel and Lymphatic Invasion**

The presence or absence of blood vessel and lymphatic invasion should be reported Figure 2.27. Regression of over 75% of the tumor volume of a melanoma is considered a bad prognostic sign. Regression is characterized by haphazard, pattern less fibrosis, accompanied by vertically oriented, ectatic blood vessels (straight arrow), melanophages (curved arrow) and patchy lymphocytic infiltration of the stroma. In complete, as opposed to partial regression, there is a total absence of neoplastic melanocytes in the dermis and in the overlying epidermis.

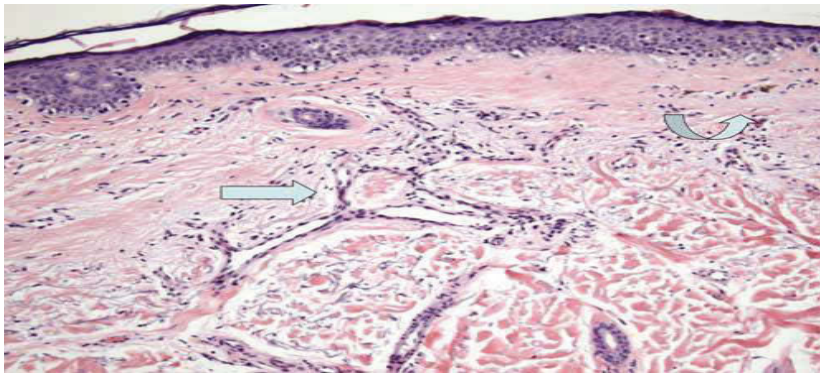


Figure 2.24 Blood Vessel and Lymphatic Invasion

- **2.6.4.8 Pathologic features of a melanoma**
- Cellular atypia: cancerous melanocytes have atypical cellular features compared to normal melanocytes, such as:
    - Increased nuclear: cytoplasm ratio
    - Found migrating in throughout the layers
- Nests: atypical melanocytes are usually found in nests suggesting uncontrolled growth

2.7 Summary

This chapter provided the essential elements that form the medical basis for melanoma diagnosis and it includes the statistically proven analytical variables. As mentioned, clinicians and oncologists must recognize specific risk factors for metastatic disease so as to guide adjuvant therapy and further surgical therapy including complete removal of the affected area when necessary. The skin level diagnosis along with the pathological analysis of the effected skin area has now been shown in large trials to offer significant survivorship benefits. It is therefore considered essential that patients receiving melanoma diagnoses be provided with the accurate and comprehensive diagnostic assessment upon which the novel therapeutic strategies will be based. Several scoring based diagnosing systems had been reviewed. In order to compliment the diagnosis made by the medical experts a lot of research is being done for the development of computer aided diagnosis systems that will be discussed in the next chapter.

Chapter 3

Methodological Review

3.1 Introduction

Computer based skin cancer diagnosis systems can provide significant contribution in early recognition of malignant melanoma. This chapter reviews the state of the art in such systems and examines current practices, problems, and prospects of image acquisition, pre-processing, segmentation, feature extraction & selection and classification of skin cancer images. It reports statistics and results of the most important implementations that exist in literature. The performance of few classifiers specifically for skin lesion diagnostic problem is observed and the corresponding findings are discussed. Whenever available, indication of various conditions that affect the techniques performance is reported. A framework is suggested based on useful considerations for critically assessing the quality of skin cancer diagnostic models and the results based on these models.

Parts of work presented in this chapter are also published as a review paper in International Journal of Biomedical Imaging [69] and as book chapter in Lecture notes in Computer Science [70].

3.2 Importance of Computer Aided Diagnosis

The incidence of skin cancer (malignant melanoma (MM])) has increased dramatically over the past few decades [4] . Australia is one of those countries in which skin cancer is widely spread in comparison to other types of cancer[71]. In results of the Health Benefits and Risks of Sun Exposure experiments, researchers reported that solar radiation is the main cause of skin cancers. However, it also is a main source of vitamin D for humans [72]. Authorities should pay attention not only to skin cancer research, but also to research on vitamin D–sun–health relationships occurring worldwide. Researchers found that primary prevention and early detection continue to be of paramount importance in addressing the public health threat of skin cancer[73, 74].

Doctor's diagnosis is reliable, but this procedure takes lots of time and efforts. These routines can be automated. It could save lots of doctor's time and could help to diagnose more accurately[75]. By using computerized means there is a good opportunity to store diagnostic information in order to use it for further investigations or creation of new methods of diagnosis.

The main problem is that the diagnosis is highly dependent on subjective judgement, and is scarcely reproducible [76, 77]. Several scoring systems and algorithms such as the ABCD-E rule [26, 78, 79], the seven-point checklist [80-82], three-point checklist [83] and the Menzies method [31, 84] have been proposed to improve the diagnostic performance of less experienced clinicians. Although this simplification has enabled the development of these diagnostic algorithms with good accuracy but still they showed problems that have not yet been solved. The most important is that the purpose for which they were designed was not achieved, because the within- and between-observer concordance is very low, even for expert observers [26, 85-87]. Despite extensive research in investigating the varied presentations and physical characteristics of melanoma, the clinical diagnostic accuracy remains sub optimal. Thus, a growing interest has developed in the last two decades in the automated analysis of digitized images obtained by Epiluminance microscopy (ELM) techniques to assist clinicians in differentiating early melanoma from benign skin lesions.

Application of computational intelligence methods helps physicians as well as dermatologists in faster data processing to give better and more reliable diagnoses. Studies related to the automated classification of pigmented skin lesion images have appeared in the literature as early as 1987 [88]. After some successful experiments on automatic diagnostic systems for melanoma diagnosis [88-94], utility of machine vision and computerized analysis is getting more important every year. The importance of the topic is patent if the enormous quantity of researches related with the melanoma diagnosis is analysed. Numerous computerized diagnostic systems have been reported in the literature where different border detection, feature extraction, selection and classification algorithms are used. Some researchers [89, 95-100] reviewed and tried to critically examine image analysis techniques for diagnosis of skin cancer and compared diagnostic accuracy of experts dermoscopists with artificial intelligence and computer aided diagnosis. More research, however, is needed to identify and reduce uncertainties in the automatic decision support systems to improve diagnosis accuracy.

3.3 Development of Computer-aided diagnosis system

Computer aided decision-support tools are important in medical imaging for diagnosis and evaluation. Predictive models are used in a variety of medical domains for diagnostic and prognostic tasks. These models are built from experience which constitutes data acquired from actual cases. The data can be pre-processed and expressed in a set of rules, such as it is often the case in knowledge-based expert systems and consequently serve as training data for statistical and machine learning models.

The general approach of developing a CAD system for diagnosis of skin cancer is to find the location of a lesion and also to determine an estimate of the probability of a disease. **The first step in this research was to establish a standard general scheme of a CAD system for skin lesions. This thesis proposed general skin cancer diagnositc scheme is shown in Figure 3.1.** The inputs to the Computer aided System are digital/dermoscpic images, with the possibility to add other system such as histopathalogical images obtained from slides of the biospsy samples from patients taken by the pathalogists. In the first phase analysis of the pre-processing of image is done which can help to reduce the ill effects and various artifacts like hair etc. present in the images. It is followed by the detection of the region of interest by image segmentation technique. Once the lesion/the effected area is localized, different features can be quantified and used for classification.

Differentiation of malignant melanoma images demands efficient image processing and segmentation approach, detailed feature extraction and reliable classification algorithms. A detailed research is necessary to make the best choice and setting the benchmarks for diagnostic system development and validation. Following sections focus on the description of the major steps that may be involved in development of an automated skin cancer diagnosis process.

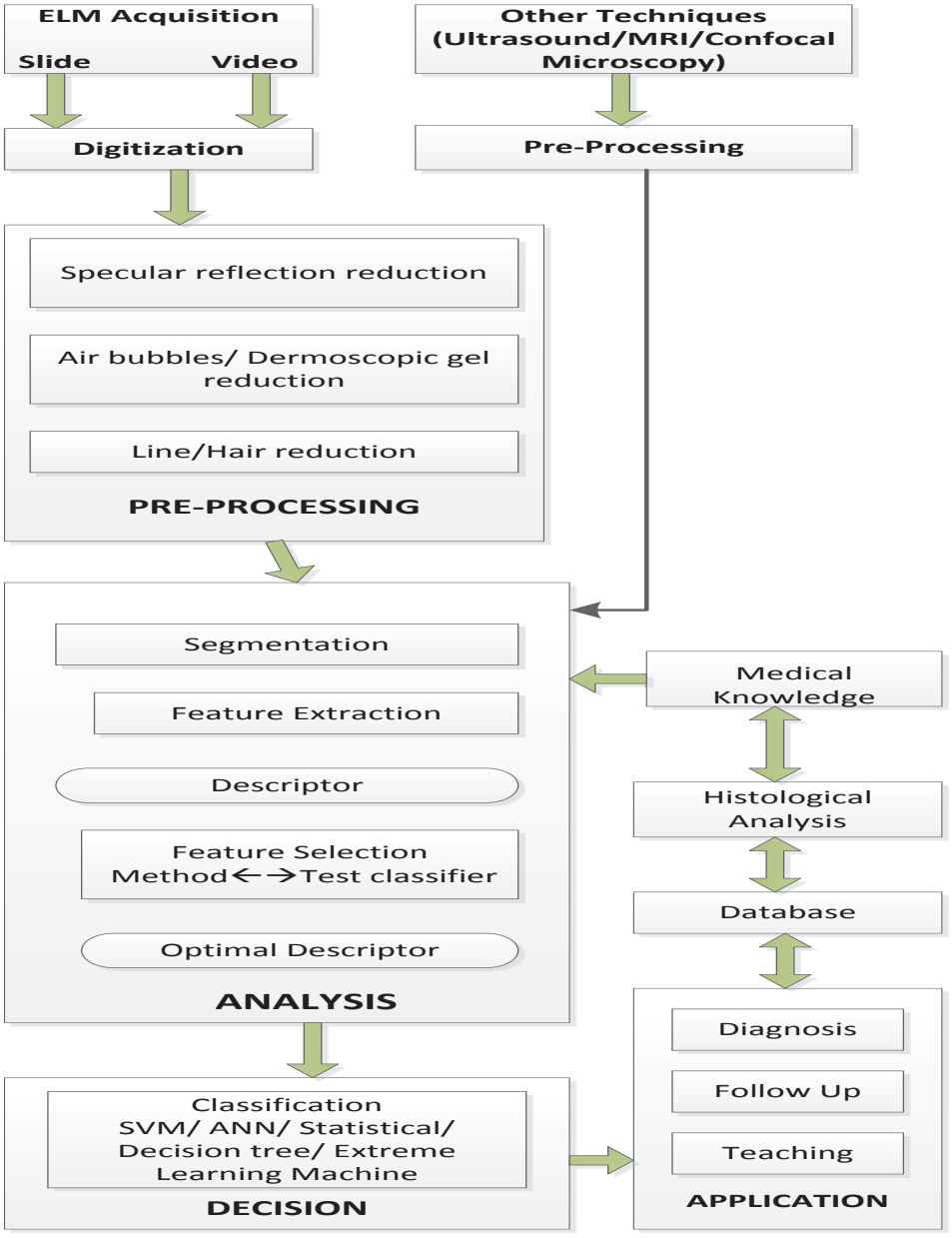


Figure 3.1 Computer Aided Diagnostic Support System for Skin Cancer Diagnosis

3.4 Image acquisition/ methods for screening skin lesions

Unaided visual inspection of the skin is often suboptimal for diagnosing melanoma. Numerous imaging modalities are under investigation to determine their usefulness in imaging and ascertaining a correct in vivo diagnosis of melanoma. These include total-cutaneous photography, dermoscopy, confocal scanning laser microscopy (CSLM), ultrasound, Magnetic Resonance Imaging (MRI), Optical Coherence Tomography (OCT), and multispectral imaging. Each technique has certain pros & cons.

These are now being harnessed to improve early detection. **A bird eye view of the currently available cutaneous imaging devices and new frontiers in non-invasive automated diagnosis of melanoma is provided in Table 3.1.** Readers may refer to [84, 101-104] for analysing performance comparison of some of the existing screening techniques.

Table 3.1 In Vivo Imaging Techniques for Diagnosis of Skin Cancer

Method	Advantages	Limitations
Photography [105-107] **Total body photo graph (2-D TBP** **3-D TBP)** **baseline photographs of individual lesions**	Affordable, easy data management. Monitor patients with many dysplastic nevi. Useful in the follow-up management & easy comparison for detecting change in size, shape or colour that may be suggestive of malignancy. 3D representation of the patient's entire cutaneous surface may reduce work time & clarify documentation.	Limited morphologic information.
Dermoscopy [85, 108-113] **ELM (oil/slide mode & polarizing mode)**	Facilitate a 10-fold magnification of the skin. Melanoma dermoscopic characteristics are well correlated to histo-pathologic features. Identify foci of melanoma for helping pathologist as to where to section the specimen so as to minimize false-negative results as a result of sampling error. Liquid immersion provides increased illumination and	Qualitative and potentially subjective. Low magnification in routinely used instruments and the limited scope of observable structures restrict the usefulness and diagnostic applicability of the method.

	resolution, sharper and less distorted colours. Polarizing mode avoid a potential source of nosocomial infections.	
Multispectral Imaging [114-116] **Melafind** **Solar Scan** **Spectrophotmetric intracutaneous analysis**	Spectral imaging is quantitative and more objective. Less inter-physician variability. Melafind can create multispectral sequence of images in less than 3 seconds. SIA scope can help in diagnosis of lesions as small as 2mm. Analyse the location, quantity and distribution of skin chromophores within epidermis and papillary dermis.	Difficult interpretation because of the complexity of the optical processes of scattering and absorption.
Laser-based enhanced diagnosis [117-119] **Confocal Scanning Laser Microscopy** **Reflectance Confocal Microscopy** **spectrally encoded confocal microscopy**	In vivo imaging of skin lesions at variable depths in horizontal planes & examination at a quasi-histological resolution without biopsy. High resolution allows imaging of nuclear, cellular and tissue architecture of the epidermis and underlying dermal structures without a biopsy and allow recognition of abnormal intra-epidermal melanocytic proliferation. No tissue damage because of low-power laser.	Processes in the reticular dermis & tumor invasion depth cannot be evaluated reliably. Technically sensitive & expensive to use in routine clinical application. Formal training and experience is required to become proficient in this new technique.

Optical Coherence Tomography [120-122]	Depth of invasion can be better measured with OCT than CSLM. Noninvasively assess and monitor inflammatory skin diseases.	Limited resolution does not allow a differential diagnosis between benign and malignant lesions. Limited to thin tumors because of strong scattering of epidermic tissue.
Ultrasound imaging [123, 124]	Can provide information about perfusion patterns of lymph nodes and other soft tissues that can be used to stage the tumor.	May overestimate or underestimate tumor thickness, accuracy of results depend heavily on skill of examiner & anatomic site of lesion.
Magnetic Resonance Imaging [125-127]	Obtain information on the depth and extent of the underlying tissue involvement & can be used to measure melanoma thickness or volume.	Need sufficient resolution & adequate number of images per sequence for discriminating skin lesions.

Relative to other specialties, dermatologists have been slow to adopt advanced technologic diagnostic aids. Thus, so far dermoscopy/ELM is one of the main methods used to image skin. Sometimes simple ELM does not sufficiently increase the diagnostic accuracy in distinguishing pigmented Spitz naevi (PSNs) from melanoma. For obviating the problems of qualitative interpretation, methods based on the mathematical analysis of pigmented skin lesions (PSLs), such as digital dermoscopy analysis (DDA) and D-ELM, have been developed [128, 129]. The visual evaluation of content of DDA is very complex. Efficient image processing techniques must therefore be developed to help physicians making a diagnosis. The introduction of digital ELM and sophisticated image processing software has opened up a new horizon in the evaluation of cutaneous benign and malignant pigmented skin lesions (PSLs) as it enables the observation, storage and objective evaluation of many parameters.

The main focus in this chapter is on automatic diagnostic system based on digital dermoscopy images normally collected from different dermoscopy atlases [28, 32] or from dermatologists. However, it is anticipated that multimodal systems that combine different imaging technologies can further improve the ability to detect melanoma at an earlier stage and reduce the trauma of dermatologic diagnosis.

3.5 Pre-Processing

The main processing step towards a complete analysis of pigmented skin lesion is to differentiate the lesion from the healthy skin. Detection of the lesion is a difficult problem in dermatoscopic images as the transition between the lesion and the surrounding skin is smooth and even for trained dermatologist, it is a challenge to distinguish accurately. It has been observed that dermoscopy images often contain artefacts such as such as uneven illumination, dermoscopic- gel, black frames, ink markings, rulers, air bubbles, as well as intrinsic cutaneous features that can affect border detection such as blood vessels, hairs, and skin lines and texture. These artefacts and extraneous elements complicate the border detection procedure, which results in loss of accuracy as well as an increase in computational time. Thus, it requires some pre-processing steps to facilitate the segmentation process by removal of unwanted objects or artefacts and colour space transformation.

Everything that might corrupt the image and consequently affect the results of image processing must be localized and then removed, masked, or replaced. It may include image resizing, masking, cropping, hair removal (or attenuation), and conversion from RGB colour to intensity grey image. It is done to reduce noise and the effect of reflection artefacts. It is meant to facilitate image segmentation by filtering the image and enhancing its important features.

The most straightforward way to remove these artefacts is to smooth the image using a general purpose filter such peer group filter (PGF)[130], mean filters, median filter [131-133], Gaussian filters [134, 135] or anisotropic diffusion filters (ADF). A major issue with these aforementioned filters is that these filters are originally formulated for scalar images. For vector images one can apply a scalar filter to each channel independently and then combine the results, a strategy referred to as marginal filtering. Although fast, this scheme introduces colour artefacts in the output. An alternative

solution is to use filters that treat the pixels as vectors [136]. Another noteworthy thing is setting of mask size proportional to the image size to manage a trade-off between smoothing of image and blurring of edges. Despite of taking care of all the fore mentioned things it is still not guaranteed to get an image free of all artefacts.

An alternative strategy for artefact removal is to use specialized methods for each artefact type. Many studies suggested their works, very few [137-139] discussed different aspects of artefacts together but none of them have discussed about all cases of artefacts. For this rationale, an overview is presented for effective pre-processing method namely colour space transformation, colour quantization, contrast enhancement, and artefact removal which are being used for reducing all the possible ill effects present in the dermoscopic images.

Dermoscopy images are commonly acquired using a digital camera with a dermoscope attachment. Due to the computational simplicity and convenience of scalar (single channel) processing, the resulting RGB (red-green-blue) colour image is often converted to a scalar image using different methods like retaining only the blue channel as lesions are often more prominent in this channel or applying the luminance transformation or Karhunen-Loéve (KL) transformation and retaining the channel with the highest variance. Skin lesions come in a variety of colours but absolute colours are not very useful in segmenting images. Normally the analysis is based on changes in colour within the lesion or with the surrounding skin particularly colour changes belonging to the lesion boundary. Therefore, it is quite common to transform the images that are in RGB colour coordinates into other colour spaces like CIEL*a*b*, CIEL*u*v*, KL, and HSI (Hue-Saturation- Intensity).

Typical 24-bit colour images have thousands of colours, which are difficult to handle directly. For this reason colour quantization is commonly used as a pre-processing step for colour image segmentation[140]. The process of colour quantization consists of two phases-palette design (i.e. selection of a small set of colours that represents the original image colours) and pixel mapping (i.e. assignment of one of the palette colours to each input pixel). Celebi et al. [130] showed that for skin lesion the colour quantization method should reduce the number of colours in image to 20 for getting precise quantization.

One of the factors that complicate the detection of borders in dermoscopy images is insufficient contrast. The contrast of image is enhanced to ensure that edges of the lesion are eminence. Recently, Delgado et al. [141] proposed a contrast enhancement method based on independent histogram pursuit (IHP). An easy, yet powerful way to enhance the image contrast is histogram stretching, a mapping of the pixel values onto [0, 255]. Another very popular technique is histogram equalization, which alters pixel values to achieve a uniform distribution. Homomorphic filtering [142], Fast Fourier Transform (FFT) and high pass filter can be used to compensate for uneven illumination or specular reflection variations in order to obtain the high contrast lesion images.

For the removal of black frames produced in the digitization process, Celebi et al. [132, 143] proposed an iterative algorithm based on the lightness component of the HSL (Hue-Saturation-Lightness) colour space. In order to remove air bubbles and dermoscopic-gels, adaptive and recursive weighted median filter developed by Sweet [144] can be utilized. This type of median filter has an edge persevering capability. A method that can remove bubbles with bright edges was introduced in [24] where the authors utilized a morphological top-hat operator followed by a radial search procedure. Line detection procedure based on the 2-D derivatives of Gaussian (DOG) [145] and exemplar-based object removal algorithm [146] can be used for removing dark lines like ruler marking. In most cases, image smoothing effectively removes the skin lines and blood vessels.

One of the most undesirable elements that are most commonly present in dermatoscopic images is hair. Lee et al. [147]and Schmid [148] used mathematical morphology. Fleming et al. [24] applied curvilinear structure detection with various constraints followed by gap filling. Erosion/Dilation with straight line segments can efficiently eliminate (or at least weaken) the hairs [149, 150]. Schmid et.al [151, 152] suggested a scheme based on a morphological closing operator [153] applied to the three components of the L*u*v* uniform colour space [154]. Zhou et al. [155] and Wighton et al.[156] proposed more sophisticated approaches based on inpainting. However, it is being observed that most of these techniques often leave behind undesirable blurring; disturb the texture of the tumor; and result in colour bleeding. Due to these problems, it is very difficult to use the colour diffuse image for further skin tumor differentiation. In contrast, an artifact removal algorithm that focuses on accurate detection of curvilinear artifacts and pays special attention to lesion structure during the removal stage has been

introduced by Zhou et al. [157]. This approach effectively removes artifacts such as ruler markings and hair, but it has high computational requirements.

To address all these issues Abbas et.al [139]developed a novel method that automatically detects these visible artifacts and removes them. Abbas et al. [158] presented a comparative study about hair removal methods which indicate that hair-repairing algorithm based on the fast marching method achieve an accurate result.

All the above mentioned strategies are meant to facilitate the segmentation and feature extraction stages which consequently lead to better diagnostic results.

3.6 Segmentation

Segmentation is the process of partitioning a digital image into multiple segments (sets of pixels). The goal of segmentation is to simplify and/or change the representation of an image into something that is more meaningful and easier to analyze. Thus, image segmentation is to cluster pixels into salient image regions, i.e., regions corresponding to individual surfaces, objects, or natural parts of objects. Image segmentation is typically used to locate objects and boundaries (lines, curves, etc.) in images. More precisely, image segmentation is the process of assigning a label to every pixel in an image such that pixels with the same label share certain visual characteristics. Thus the result of image segmentation is a set of partitions that have the whole information about the image. Each pixel in a particular region is the same kind in terms of its some characteristics or calculated property which includes colour, appearance, brightness, etc. while the nearest areas are varying with respect to the identical characteristics.

Image segmentation is one of the most important and difficult task in image analysis process. The success of image analysis process and the accuracy of the subsequent steps highly depend on the success of autonomous image segmentation technique. The initial step in the computerized analysis of skin lesions is the detection of lesion borders. The importance of border detection for the analysis is twofold[159]. Firstly, the border structure provides important information which helps in accurate diagnosis. Many clinical features such as asymmetry, border irregularity, and abrupt border cut-off in lesion are calculated from the border. Secondly, the extraction of other important clinical features such as atypical pigment networks, globules, and blue-white areas critically depends on the accuracy of border detection techniques [130].

Segmentation is difficult because of the great variety of lesion shapes, sizes, and colors along with different types and textures of human skin [130]. Other difficulties that make it a challenging task include, low contrast between the lesion and the surrounding skin; Smooth transition between the lesion and the skin; Irregular and fuzzy lesion borders; Reflections and shadows due to wrong illumination; Artifacts such as skin texture, air bubbles and hair; Variegated coloring inside the lesion.

Various image segmentation algorithms are presented in literature. However, on a broader perspective, image segmentation algorithms can be divided into two main categories (Figure 3.2), i.e. discontinuity based segmentation and similarity based segmentation.

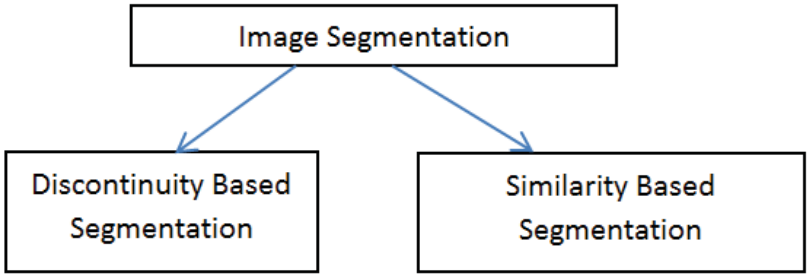


Figure 3.2 Categorization of Image Segmentation Algorithms

3.6.1 Discontinuity Based Segmentation

In discontinuity based segmentation algorithms the partition or subdivision of image is carried out based on some abrupt changes in the intensity levels in an image, or abrupt changes of gray levels of an image. So the main objective of these algorithms is to identify isolated points, lines or edges in an image. Some of the discontinuity based segmentation algorithms present in literature include Method based on zero crossing of Laplacian-of-Gaussian (LoG) [129, 160], Active contour based methods like Gradient Vector Flow (GVF), Adaptive Snake (AS), Geodesic active Contour model (GAC), Radial search technique. [135, 149, 161, 162]. Discussion on the application of GVF [163] [164] and Radial Search Technique [165] for skin image segmentation can be found in the references provided.

Limitations of Edge based approach

This approach performs poorly when the boundaries are not well defined, for instance, when the transition between skin and lesion is smooth. In these situations, the

edges have gaps and the contour may leak through them. Another difficulty is the presence of spurious edge points that do not belong to the lesion boundary. They are the result of artifacts such as hair, specular reflections or even irregularities in the skin texture and they may stop the contour preventing it to coverage to the lesion boundary.

3.6.2 Similarity Based Segmentation

In similarity based segmentation algorithms those pixels in the image are grouped together which are similar in some sense. The similarity based algorithms can be further divided in two categories, (Figure 3.3) i.e. thresholding and region based approach.

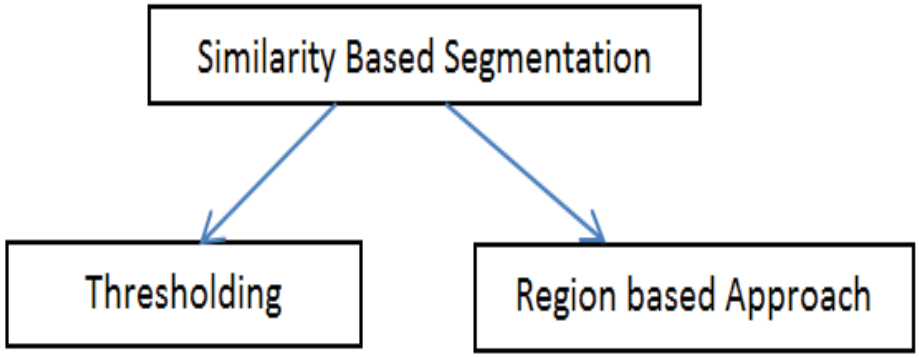


Figure 3.3 Categorization of Similarity based Segmentation Algorithms

- **3.6.2.1 Thresholding Approach**

Thresholding is the simplest approach of segmentation. A threshold value is decided under some criterion and then the pixels are divided into groups based on that criterion. Thresholding is based on the fact that the values of pixels that belong to a skin lesion differ from the values of the background. Thus, by choosing an upper and a lower value it is possible to isolate those pixels that have values within this range. The information for the upper and the lower limits can be extracted from the image histogram, where the different objects are represented as peaks. The bounds of these peaks are good estimates of the limits. Although, it should be noted that, simple thresholding that is described here cannot be used in all cases because image histograms of skin lesions are not always multi-modal.

Thresholding operation is testing a pixel against a function T, where

$$T= T[(x,y), p(x,y), f(x,y)] \quad (3.1)$$

Thus T is a function of any combination of three terms, that is;

(x,y) ➔ pixel location

f(x,y) ➔ pixel intensity at (x,y)

p(x,y) ➔ local property in a neighbourhood centered at (x,y). e.g. average intensity value within neighbourhood.

Depending upon this combination T can be local threshold, Global threshold or adaptive/dynamic threshold.

T[f(x,y)] ➔ Global threshold

T[p(x,y), f(x,y)] ➔ Local threshold

T[(x,y), p(x,y), f(x,y)] ➔ Adaptive/ dynamic threshold

Thresholding is one of the most prominent methods used for skin image segmentation in various studies in literature [166],[167], [168], [169], [170] [171]..

Limitations of Thresholding

Thresholding methods achieve good results when there is a good contrast between the lesion and the skin, thus the corresponding image histogram is bimodal. But it usually fails when the modes from the two regions overlap.

- **3.6.2.2 Region based segmentation**

In region based segmentation approach, the image is usually divided into regions that fulfil a homogeneity criterion. Region based segmentation can be further categorized into region splitting and merging approach and region growing approach, (Figure 3.4) that will be discussed briefly in the following sections.

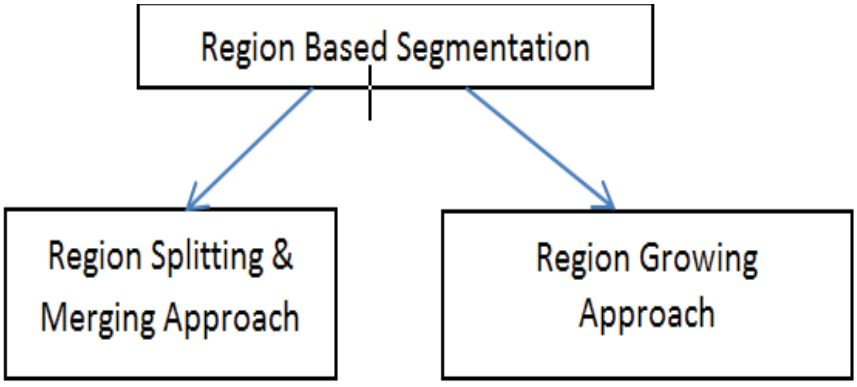


Figure 3.4 Categorization of Region based Segmentation Algorithms

A. Region Splitting & Merging Approach

In this method first the image is split into a number of different components following some criteria. Then after splitting the image into smaller components, merge some of those sub-images which are adjacent and which are similar in some sense.

Region Growing Approach

Region Growing is a procedure that groups pixels or subregions into larger regions. The procedure starts with a set of "seed" pixels and grows regions from these by appending neighbouring pixels that have similar properties such as gray level, color and texture. In other words, in this method, the start is made with any one particular pixel in the image and then group all other pixels which are connected to this particular pixel (connected pixels are those which are adjacent to this pixel and are similar in intensity value).

Some of the region based algorithms present in literature include Level Set segmentation, Statistical region merging , multiscale region growing and morphological flooding [132, 143], [130-133, 148, 151, 172, 173].

Limitations of Region Based Segmentation

Region based approaches have difficulties when the lesion or the skin region are textured or have different colours present, which leads to over-segmentation. According to some critics [131], when an image is segmented by region-based methods, the resulting image might look too square. This is not desirable because demoscopic images can contain very small objects and irregular structures.

3.6.3 Segmentation Using Soft-Computing Approaches

Soft Computing is an emerging field that consists of complementary elements of fuzzy logic, neural computing and evolutionary computation. One of the important applications of self-computing is image segmentation [174]. Three different soft computing approaches to image segmentation which are most frequently used are Fuzzy based Approach [175] [131] [176], Genetic Algorithm based approach [171, 177, 178] and Neural Network based Approach [179] [180].

Some researchers [181]argued that manual border detection is better than computer-detected borders in order to separate the problems of feature extraction from the problems of automated lesion border detection. However, for the development of automated diagnostic system for skin lesion detection it is very important to develop automatic segmentation algorithms. As segmentation is a crucial early step in the analysis of lesion images, so, it has become one of the important areas of research and many algorithms and segmentation techniques are available in literature. **A brief overview of various segmentation algorithms being used for dermoscopic image analysis in literature is provided in Table 3.2 [69] for quick reference.**

Table 3.2 Methods for Segmentation of Dermoscopic Images

Method	Description	Related References
Thresholding	Determining threshold and then the pixels are divided into groups based on that criterion. It include bi-level and multi-thresholding	Histogram thresholding ([182],[183], [166],[167], [184], [185]) Adaptive Thresholding ([134],[186], [187], [188], [189], [168], [169], [170])
Color-based segmentation algorithms	Segmentation based on colour discrimination. Include Principle component transform/ spherical coordinate transform	[190], [191],[192],[193], [194], [195]
Discontinuity Based Segmentation	Detection of lesion edges using Active contours/ radial search techniques/ zero crossing of Laplacian of Gaussian (LoG)	Active contours ([196],[135],[197], [186],[139],[198], [133]) Radial search ([199], [200],[165]) LoG ([201],[202], [203],[204])

Region-based segmentation	Splitting the image into smaller components then merging sub-images which are adjacent and similar in some sense. It include Statistical region merging, Multi-scale region growing, morphological flooding	Split and merge ([205], [206]) SRM ([132],[143], [207], [170]) Multi-scale ([208], [209]) Morphological flooding([151])
Soft computing	methods involve the classification of pixels using soft computing techniques including neural networks, fuzzy logic, and evolutionary computation	Fuzzy logic ([210], [211],[131], [148], [152], [187],[157]) Neural Network ([179],[180]) Optimization algorithms ([212], [179])

Several comparative studies [132, 134, 138, 162, 186] are also present in literature which provides performance analysis of several segmentation algorithms. There are several issues that should be kept in mind for selecting a suitable algorithm for example, scalar vs. vector processing, automatic vs. semi-automatic and number of parameters whose values need to be determined a priori [137]. Interested readers may check relevant references to identify a suitable approach for a specific study.

3.7 Feature extraction

Melanoma is visually difficult to differentiate from Clark nevus lesions which are benign. It is important to identify the most effective features to extract from melanoma, melanoma in situ and Clark nevus lesions, and to find the most effective pattern-classification criteria and algorithms for differentiating those lesions. Thus the next stage of the image analysis process is extracting the important features of the image.

The purpose of feature extraction is to reduce the original data set by measuring certain properties, or features, that differentiate one input pattern from another. The feature extraction is performed by measurements on the pixels that represent a segmented object allowing various features to be computed. Unfortunately, the feature extraction step is often subject to error. In most of the publications dealing with this topic, many features are extracted to feed a sophisticated classifier, but there is very little discussion about the real meaning of those features and about objective ways to measure them. Thus, this topic is investigated in detail to come up with a guideline for future researchers.

ifferent feature extraction methods found in literature include statistical, model-based, and filtering-based methods. Various researchers used principal component analysis (PCA) of a binary mask of the lesion, , wavelet packet transform (WPT) [166, 184, 207, 213, 214], grey level co-occurrence matrix (GLCM) [134, 215] , Fourier power spectrum [216], Gaussian derivative kernels [217]and decision boundary feature extraction[218-220] in order to reduce data redundancy. Some of the typically used filter banks are Laws masks, the dyadic Gabor filter bank, and wavelet transform [221]. A particular problem in the related literature is that a significant number of studies do not report the details of their feature extraction procedure.

The ABCD-E system[26, 78, 222], 7-point checklist [81, 223], 3 points checklist[34], pattern analysis [76] and Menzies method [224] offers alternative approaches in deciding the differentiating features that need to be extracted.

According to the conclusion made by Johr [80] the automatic extraction of characteristics that take into account the rule ABCD [26, 222, 225] is computationally less expensive than the ones that take into account 7 points checklist [81, 223] or the Menzies method[31, 226]. Furthermore, the reliability in the clinical diagnosis is very high for ABCD-E rule. So, most of the automated decision support systems also use ABCD rule as the base of their feature extraction step. Pigmented skin lesions are typically evaluated by dermatologists using the “ABCD” rule; which analyse the Asymmetry, Border irregularity, Colour variation and Diameter of a lesion. Most of the automated decision support systems also use ABCD rule as the base of their feature extraction step. Types of features utilized in this research include Border/shape Features which cover the A and B parts of the ABCD-rule of dermatology, Colour Features which correspond to the C rules and Textural Features.

Melanoma is often defined primarily by asymmetry and variations in colour, texture and border regions so most researchers focus on features that best represent this variability. In automated diagnosis of skin lesions, feature design is based on the ABCD-rule of dermatoscopy. ABCD represent the asymmetry, border structure, variegated colour, and dermatoscopical structures and define the basis for a diagnosis by a dermatologist.

However, ABCD is more prone to over classification of atypical melanocytic nevi as melanomas. Pehamberger et al. [76]and Menzies et al [224]define further eight standard ELM criteria— pigment network, brown globules, black dots, radial streaming, pseudopods, overall pigmentation, and depigmentation—for the diagnosis with the help of ELM. Dolianitis [227] showed in a comparative study that Menzies method showed the highest sensitivity, 84.6%, for the diagnosis of melanoma, followed by the 7-point checklist (81.4%), the ABCD rule (77.5%), pattern analysis (68.4%), and assessment of a macroscopic image (60.9%). Pattern analysis and assessment of the macroscopic image showed the highest specificity, 85.3% and 85.4%, respectively. So many researchers [167, 170, 185, 189, 228, 229] are trying to develop efficient automatic diagnostic systems based on 7-point criteria and pattern analysis.

Numerous methods for extracting features from clinical skin lesion images have been proposed in the literature. Several studies have also proven the efficiency of border shape descriptors for the detection of malignant melanoma on both clinical and computer based evaluation methods [200, 230]. Very simple parameters, such as area and perimeter, are extracted in [208, 231] [201]. Measurements of shape features are also used like fragmentation index [232-234], thinness ratio/ circularity factor [113, 134, 168, 235] asymmetry index [150, 230, 234], aspect ratio [169, 208], compactness [169, 208], symmetry axis [236] , bulkiness score [237] , irregularity index [238, 239], fractality of borders [201], convex hull ratio [168] and skin line pattern [240] . Some groups use the sharpness of the transition from the lesion interior to the skin [113, 134, 235] as descriptors of the structure and irregularity of the border. Hall et al. [89]calculate fractal dimensions to represent border irregularity. Lacunarity [241] is another measure that can be used to characterize a property of fractals and quantifies aspects of patterns that exhibit changes in structure.

Colour features are mainly statistical parameters calculated from different colour channels, like average value and standard deviation of the RGB [113, 168, 232-234] or HSV colour channels [235]. Other colour features used in different studies include colour asymmetry [208] , centroidal distance [208], LUV histogram distance [208] etc. Cotton and Claridge [242] found that all normal skin colours lie on a two–dimensional surface patch within a three- dimensional (3-D) colour space (CIE–LMS). Atypical skin structures result in colour coordinates that deviate from the normal surface patch. Some researchers [134, 195, 208, 243] [201] , used GLCM based texture features [244-246] like dissimilarity, contrast, energy, maximum probability, correlation, entropy etc.

Parameters for the description of dermatoscopic structures and ELM criteria are difficult to find in literature. Major issues are concerned with the difficulty in relating such information as lesion shape and colour to medical structures (tissues, vessels, etc.) which experts are more familiar with. Some of the dermoscopic feature extraction studies include atypical pigment networks [24, 185, 247], globules/dots/blotches [24, 210, 248-250], streaks[251], granularity [252] and blue-white veil [181, 253]. It is noteworthy that diagnostic systems based on extraction of critical high level features show an increase in the diagnostic accuracy of computerized dermoscopy image analysis systems. Thus, in addition to general features like area, border, shape, colour etc. these high level features can also be integrated in the automated diagnostic system to gain greater clinical acceptance.

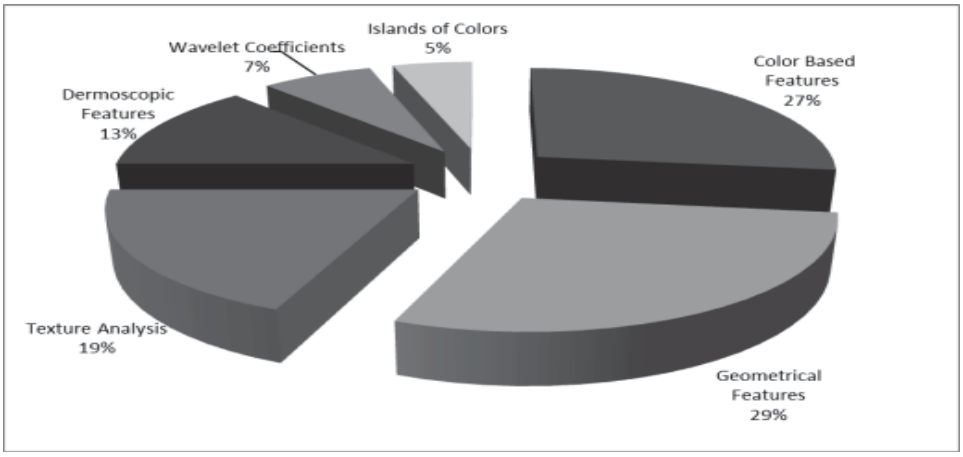


Figure 3.5 Illustration of feature distribution used in dermoscopic studies in literature

Figure 3.5 shows relative distribution of various types of features used in dermoscopic studies. It is based on the literature review conducted for this thesis.

Some researchers used some unique features for classification but it is found from skin cancer research that a unique feature is not sufficient to diagnose precisely skin cancer, and that the combination of different criteria is the key to the early detection of malignant melanoma and other types of skin cancer. The evolution of competing dermoscopic algorithms with variable definitions of specific attributes complicates dermoscopic diagnosis. It is necessary to identify features that are the most reproducible and diagnostically significant.

3.8 Feature selection

For clinical purposes it is arguable that use of most relevant features is a desirable characterisitcs of a good predictive model [254]. Similarly, features selection is a critical step for successfully distinguishing between malignant melanoma, benign and dyplastic nevi. Many potential features may be used but it is important to select a reasonable reduced number of useful features while eliminating redundant, irrelevant or noisy features. However, it is important to make sure that there may not be loss of significant information.

Feature selection is an important pre-processing step in many machine-learning tasks. The purpose is to reduce the dimensionality of the feature space by eliminating redundant, irrelevant or noisy features. From the classification perspective, there are numerous potential benefits associated with feature selection: (i) reduced feature extraction time and storage requirements, (ii) reduced classifier complexity for better generalization behaviour, (iii) increased prediction accuracy, (iv) reduced training and testing times, and (v) enhanced data understanding and visualization.

Feature selection can be defined as a process to select the best optimal subset of M features from the original set of *N* features. The original feature set can be formed by concatenating the features formed by different feature extraction methods. Feature selection consists of choosing, the subset of features among the input features that has maximum prediction power for the output. When the dimensionalities of a domain expand, then the number of feature set *N* also increases and finding the best feature subset is usually difficult.

All features could be important for some problems, but for some other problems a small subset of features is usually relevant. Feature selection is carried out for

eliminating redundant and irrelevant noisy features. The standard data sets can contain hundreds of variables of which many of them could be highly correlated with other variables. For example, when two features are perfectly correlated, only one feature is sufficient to describe the data. The dependant variables provide no extra information about the classes and it just adds to the computational cost. Hence by eliminating the dependent variables, the amount of data can be reduced which can lead to efficient performance of the classifier. On the other hand, variables which have no correlation to the classes may serve as noise and may result in introducing bias in the predictor and consequently, reduce the classification performance. This usually happens when there is a lack of information about the process being studied. By applying feature selection techniques some insight into the process can be gained and can improve the prediction accuracy and reduce the computation requirements. Removing redundant and irrelevant features can help improve the performance of classifiers by reducing the potential for over-fitting and selecting only those features that work best.

The importance of feature selection phase has led to the development of a variety of techniques for selecting an optimal subset of features from a larger set of possible features. In this section a brief overview is presented for the feature selection methods found in literature.

In the broader context, two approaches of feature dimension reduction are found in the literature. The first approach projects the feature along the principal components (i.e. the eigenvectors corresponding to the largest, second largest and so on) of the data matrix, that is constructed by column wise augmenting the data points. Several algorithms like Principal Component Analysis (PCA), Singular Value Decomposition (SVD) etc. has been developed with the primary idea of projecting the data points along the principal components [255] Since maximum information is carried by the eigenvectors corresponding to large eigenvalues, usually the projection is restricted to the few large eigenvectors. Naturally the data dimension is changed by the selected number of eigenvectors corresponding to the few large values.

It is noted that principal component based feature reduction does not preserve the original features of the data points. Further, the principal component based methods presume only linear dependence of the features. Although recently Kernelized PCA and its variants attempted to take care of the nonlinear dependence of features, but they may

not ensure the right selection of the features as it is detrimental to the choice of kernel functions. Thus, sometimes the algorithms that remove irrelevant variables cannot be compared with dimension reduction methods such as Principal Component Analysis (PCA) [255] since good features can be independent of the rest of the data [256].

On the other hand, an alternative to the above dimension reduction techniques is feature selection that reduces dimensionality by selecting important subsets of existing features. Feature selection does not create new features since it uses the input features itself to reduce their number. It was chosen to use feature selection approach based on the real features rather than the projected versions as that of PCA, because we intended to do task dependent feature selection. While approaches like PCA, component analysis, clustering of inputs mainly looks at inputs of a classification problem and tries to reduce their description without regard to output.

Once a feature selection criterion is selected, a procedure must be developed which will find the subset of useful features. Directly evaluating all the subsets of features (2N) for a given data becomes an NP-hard problem as the number of features grows. Hence a suboptimal procedure must be used which can remove redundant data with tractable computations. In any feature subset selection method, there are some factors that need to be considered, the most significant are: the evaluation measure and the search strategy to find that subset. In other words, the search for the optimal feature subset requires an evaluation measure to estimate the goodness of subsets and a search strategy to generate candidate feature subsets.

3.8.1 Evaluation Measures

Feature selection algorithms are divided into filters, wrappers and embedded approaches [257]. Filter methods act as pre-processing to rank the features wherein the highly ranked features are selected and applied to a predictor. Filters approaches evaluate quality of selected features using measures and estimations which are solely dependent on intrinsic properties of the data and independent of any learning algorithm. Wrapper approaches require application of a classifier to train the given feature set to evaluate this quality. In wrapper methods the feature selection criterion is the performance of the predictor i.e. the predictor is wrapped on a search algorithm which will find a subset which gives the highest predictor performance. Embedded approaches perform feature selection during learning of optimal parameters. Some classification algorithms have

inherited the ability to focus on relevant features and ignore irrelevant ones. Embedded methods used in literature [258],[259], [260] include variable selection as part of the training process without splitting the data into training and testing sets. Brief overview of all there type of evaluation measures is presented as follows.

- **3.8.1.1 Filter Methods**

Filter methods use feature ranking techniques as the principle criteria for feature selection and they are applied before classification to filter out the less relevant features. Filter feature selection methods apply a statistical measure to assign a score to each feature. The features are ranked by that score and a threshold is used to remove ones below the threshold score. A basic property of a unique feature is to contain useful information about the different classes in the data. This property can be defined as feature relevance [257] which provides a measurement of the feature's usefulness in discriminating the different classes under consideration. These methods are often univariate and consider the feature independently, or with regard to the dependent variable. They can be split into two categories, those that evaluate all the features in one pass and those that evaluate multiple proposed feature-subsets combined with a heuristic search.

Example of some filter methods include the Chi squared test, information gain, mutual information and correlation coefficient scores. In [261] a random variable called probe is used to rank the features using Gram-Schmidt orthogonalization. The RELIEF algorithm [262], [263] is another popular filter based approach wherein a feature relevance criterion is used to rank the features but the drawback of the RELIEF algorithm is in selecting a threshold.

Ranking methods are usually used due to their simplicity and speed. The other advantages of feature ranking are that it is computationally light and avoids overfitting and is proven to work well for certain datasets [258], [264], [265] . Filter methods do not rely on learning algorithms which are biased which is equivalent to changing data to fit the learning algorithm.

On the other hand, the drawback of ranking methods is that the selected subset might not be optimal and chances are high that a redundant subset might be obtained. Some ranking methods such as Pearson correlation criteria and Mutual Information do not discriminate the features in terms of the correlation to other features. The features in

the subset can be highly correlated and if a more advanced selection method is used a smaller subset would suffice [264] [266]. In feature ranking, important features that are less informative on their own but are informative when combined with others could be discarded [258], [267]. Finding a suitable learning algorithm can also become hard since the underlying learning algorithm is ignored [266]. Also, there is no ideal method for choosing the dimension of the feature space.

Next, overview is provided for the second type of evaluation method for feature selection called wrapper methods. Unlike filter methods which use feature relevance criteria, wrapper methods rely on the classification for obtaining a feature subset.

- **3.8.1.2 Wrapper Methods**

Wrapper methods wraps around the induction algorithm that is used as the final classifier. A predictive model is used to evaluate a combination of features and assign a score based on model accuracy. Wrapper methods use the predictor as a black box and the predictor performance is used as the objective function to evaluate the feature subset. Since evaluating subsets can become NP-hard problem, exhaustive search methods can become computationally intensive for larger datasets. Therefore, suboptimal subsets are found by employing search algorithms which find a subset heuristically. A number of search algorithms can be used to find a subset of variables which can maximize the classification performance.

The search process can be methodical such as a best-first search, it may be stochastic such as a random hill-climbing algorithm, or it may use heuristics, like forward and backward passes to add and remove features. For example, the Branch and Bound method [257], [268] used tree structure to evaluate different subsets for the given feature selection number. But the search would grow exponentially [257] for higher number of features. Therefore, simplified algorithms such as sequential search or evolutionary algorithms such as Genetic Algorithm (GA) [269] or Particle Swarm Optimization (PSO) [270] are employed which can produce good results and are computationally feasible.

Naturally, the wrapper method can be more expensive as compared to filter method. It usually consists of K-fold cross-validation on the training data. But it also has advantages. They are usually more accurate and also evaluate the set of variables selected and not just one at a time, so redundant features can be removed, and variables whose usefulness depend on other features can also be found. It will also find features that might

be particularly useful with the proposed induction algorithm, so e.g. if a support vector machine (SVM) and a decision tree algorithm get best performance with different features the wrapper method could take advantage of this.

- **3.8.1.3 Embedded Methods**

Embedded methods learn which features best contribute to the accuracy of the model while the model is being created. The most common type of embedded feature selection methods are regularization methods. Regularization methods are also called penalization methods that introduce additional constraints into the optimization of a predictive algorithm (such as a regression algorithm) that bias the model toward lower complexity (less coefficients). Examples of regularization algorithms are the LASSO, Elastic Net and Ridge Regression.

Filter approach is computationally faster. Wrapper and embedded methods are computationally more expensive than filters; however, they are usually more accurate. On the other hand, although, wrapper and embedded approach can provide smaller subset of features with higher accuracy, but it is argued that Filter approaches have more generalization tendency as it is independent of any learning algorithm.

3.8.2 Search Strategy

Search is a key topic in the study of feature selection methods. An efficient search strategy is needed to explore the feature space. Searching for the optimal subset, which can achieve the best performance according to the defined evaluation measure, is a quite challenging task. Various search methods that differ in their optimality and computational cost have been developed to search the solution space. These methods include: Exhaustive Search, Branch and bound method, sequential search algorithms and Heuristic/stochastic search algorithms. A brief description of all these methods are provided here for reference, but the main focus will be on evolutionary computation and population based search procedures like Differential evolution [271], Genetic methods (GA) [269], Ant Colony Optimization (ACO) [272], Particle Swarm Optimization (PSO) [270], as this is the focus of research in recent years. The proposed feature selection method in this research is also based on adaptive version of differential evolution algorithm which also comes under the category of stochastic/population based search algorithms.

- **3.8.2.1 Exhaustive Search**

All the possible $\binom{D}{d}$ subsets are evaluated and the best one among them is chosen. The exhaustive search, which considers all possible subsets, guarantees the optimal solution. However, it is impractical to run, even with moderate size feature sets and computational time is intractable when the problem size is not small. A number of other search strategies that differ in their computational cost and optimality have been proposed in the literature.

- **3.8.2.2 Branch and Bound**

One of the early search strategies is the branch and bound [268], which requires the evaluation function to be monotonic. This algorithm generates a search tree that identifies the features being removed from the original set. It achieves a substantial reduction in the number of subset evaluations by pruning those sub trees that will never be superior to the current best solution. However, the main problem with this algorithm is its exponential time complexity. This method can be computationally expensive for large datasets. Additionally, this algorithm requires the strict assumption of monotonicity, i.e., adding new features never degrades the performance.

- **3.8.2.3 Sequential search algorithms**

Sequential search methods, such as sequential forward selection (SFS) [273], [274], Sequential Floating Forward Selection (SFFS) [273], [274] and sequential backward elimination (SBE) [257], have been widely used as they are simpler and the computational cost is relatively smaller. These algorithms are called sequential as they follow iterative algorithms.

To avoid the nesting effect of the usual sequential search methods, an adaptive version of the SFFS was developed in [275], [276]. The Adaptive Sequential Forward Floating Selection (ASFFS) algorithm used a parameter r which specified the number of features to be added in the inclusion phase and that parameter was calculated adaptively. Another parameter o was used in the exclusion phase to remove maximum number of features if it increased the performance. The ASFFS attempted to obtain a less redundant subset than the SFFS algorithm. It can be noted that a statistical distance measure can also be used as the objective function for the search algorithms as done in [260], [259], [273], [275]. Theoretically, the ASFFS should produce a better subset than

SFFS but it was noted that it is dependent on the objective function and the distribution of the data.

A lightly more reliable sequential search method is the plus-l-minus-r , which considers removing features that were previously selected and selecting features that were previously eliminated [277]. The Plus-L-Minus-r search method [275] , [278], [279] also tries to avoid nesting. In the Plus-L-Minus-r search, in each iteration L number of features are added and r number of features are removed until the desired subset is achieved. The parameters L and r have to be chosen arbitrarily. In [278] the authors tried to improve the SFFS algorithm by adding an extra step after the backtracking step in the normal SFFS in which a weak feature is replaced with a new better feature to form the subset.

- **3.8.2.4 Evolutionary Computation algorithms**

Another trend of search strategies is the stochastic search, where it has been found that including some randomness in the search process makes it less sensitive to the dataset [280], and hence helps avoid local minima. Stochastic methods provide a more efficient solution even in case of noisy data, but at the expense of computational cost. But, stochastic methods are the best choice when optimal or minimal subset is required. This method does not perform a search on all feasible feature subsets. It randomly generates subsets of features based on some defined parameters, and then assigns suitable features to the subsets in an iterative search process to find the optimal subsets. Evolutionary Computation algorithms have been applied to feature selection problems. When it comes to comparative performance Kudo and Sklansky [281] presented a rather fair comparison and concluded that SFFS is the best among the sequential search algorithms and is suitable for small and medium-sized problems, while the stochastic methods like GA etc. are suitable even for large-sized problems.

Some of the famous stochastic methods used in feature selection are: simulated annealing [282] [283], genetic algorithm (GA) [284-286], ant colony optimization (ACO) [287, 288], particle swarm optimization (PSO) [270, 289] and recently differential evolution (DE) [290].

A detailed review of evolutionary optimization can be found in [291, 292]. Here description of four main evolutionary computation methods (GA, ANT, PSO and DE) is presented for reference as they form the basis of most of the stochastic feature selection

algorithm and form the basis of the experiments presented in this thesis. Genetic algorithm, Ant Colony Optimization and Particle Swarm Optimization is discussed briefly and details will be presented only for differential evolution algorithm as this forms the basis of the proposed algorithm suggested for feature selection for skin cancer image analysis.

A. Genetic algorithm

GAs is stochastic search algorithms mainly inspired by the genetic process of biological organisms. Genetic algorithms has successfully been used to solve various search, optimization, and machine learning problems [293]. It searches for the best solution in search space intelligently and works in an iterative order, to obtain a new generation created from the old one. The GA consists of several solutions called chromosomes. Genetic Algorithm (GA) [269] can be used to find the subset of features [276] wherein the chromosome bits represent if the feature is included or not. The global maximum for the objective function which is normally the predictor performance, gives the best suboptimal subset. In GA's, a fitness function computes the fitness of every string. The standard operators used in GA's are mutation, crossover and selection. The binary genetic algorithm (BGA) [294] employs binary chromosomes which includes several genes with binary values 0 and 1, which determines the attributes for each individual. Many variations of GA-based algorithms were introduced in the literature [286, 295, 296] [297] [298] that can be used for feature selection. One of the constrained version of GA denoted as hybrid genetic algorithm (HGA) was introduced in [299] that limits the number of selected features to a predefined subset size.

B. Ant Colony Optimization

Ant Colony Optimization (ACO) [287] based feature selection methods represent the features as nodes interconnected with links. The search for the optimal feature subset is implemented by using a number of ants that would traverse through the graph which in feature selection case is the feature space. This is usually performed by utilizing previous knowledge, that is, the pheromone trails, and local importance that are measured with respect to the features (nodes) that have already been visited. The main advantage of such representation is that the pheromone laid by the ants while traversing the graph represents a global information sharing medium, which can eventually lead the ants to the vicinity of the best solution.

One main problem associated with use of original ACO for feature selection problem is that the solutions represented by the ants are constructed sequentially, thus, an optimal performance is not guaranteed. Thus, modifications to ACO-based feature selection were needed to be introduced to overcome this problem. In addition to that, many ACO based feature selection algorithms employ some sort of prior estimation of features' importance. For large datasets with thousands of features, this property will increase the memory requirement of the algorithm. For example, the algorithm presented in [300] requires the computation of mutual information between feature pairs, as well as between feature pairs and the target classes. One can easily notice that computational cost and memory requirement increase exponentially with the dataset size. Hence, the algorithm may not be practical to run for very large datasets. More details about this algorithm can be found in [300].

C. Particle Swarm Optimization

Particle Swarm Optimization (PSO) first introduced by Kennedy and Eberhart [270, 301] is a biologically inspired population- based stochastic optimization technique derived from the collective behaviour of bird flocks. A particle in search space can be considered as an individual bird of a flock. PSO consists of a set of solutions (particles) called population. Each solution consists of a set of parameters and it represents a point in the multidimensional space. A group of particles (population) makes up a swarm.

The PSO was originally introduced for the optimization of problems in continuous multidimensional search space. To extend that concept to feature selection, it was required to be developed to deal with the binary data, in which 0 and 1 denote the absence and presence of a feature, respectively. For feature subset selection, binary particle swarm optimization (BPSO) [302] algorithm employs a set of particles that adjust their own positions according to two fitness values, a local fitness value and a global fitness value. An inertia weight that controls those two values is fine tuned to enhance the exploration capability of PSO.

However, it is noted that the performance of PSO degrades when the dimensionality of the problem is too large, and that PSO can easily get trapped in local minima. Thus, an improved version of the binary particle swarm (IBPSO) was presented in [303]. This modified algorithm retires global best after certain number of iterations to avoid being trapped in a local optimum. However, there is no guarantee that this modification would produce better results.

The main shortcoming of PSO is the premature convergence of a swarm. The main reason for this shortcoming is that the particles try to converge to a single point, located somewhere on a line between the global best and the personal best positions. However, this point is not always guaranteed to be the optimum point. Another reason could be the fast rate of information flow between particles, which can lead to the creation of similar particles. This results in a loss in diversity. Thus, the possibility of being trapped in local optima is increased.

D. Differential Evolution

Differential evolution (DE) algorithm, proposed by Storn and Price [304, 305] emerged as a very competitive form of evolutionary computing method. It is a simple but powerful stochastic, population-based, evolutionary search algorithm for solving global optimization problems.

DE based algorithms have been frequently used to tackle different kind of optimization problems and its effectiveness and efficiency has been successfully demonstrated in many application fields [306-308] including recent applications in pattern recognition [309-312] which is the concerned area of this research.

DE has proven to outperform many other optimization algorithms in terms of convergence speed and robustness over many common benchmark problems and real world applications [308, 313, 314]. The strength of the algorithm lies in the fact that it has easier implementation and comparatively lesser parameter tuning is required. It provides speed of finding the optimal or suboptimal points of the search space and it is quite robust, i.e. it produces nearly same results over repeated runs. The algorithm uses only primitive mathematical operators and is conceptually very simple. In addition to that the performance does not deteriorate severely with the growth of the search space dimensions. These issues perhaps have a great role in the popularity of the algorithms within the domain of machine intelligence and pattern recognition. However, it should be noted that the control parameters and learning strategies involved in DE are highly dependent on the problems under consideration. This method is not applied so far for skin cancer image analysis, but keeping an eye on the advantages of this method an adaptive version of DE algorithm is proposed in this thesis and it proved to provide superior performance.

3.9 Feature Normalization

In classification tasks the features that characterize the samples quite often have different ranges. Many classifiers such as k-nearest neighbours and neural networks require that the features be normalized so that their values fall within a specified range. In the case of SVMs, feature normalization is an important pre-processing step that is necessary to prevent features with large ranges from dominating the calculations and also to avoid numerical instabilities in the kernel computations. : A variety methods were used to normalize the data, including the unit-vector, statistically based (statistic-based), min-max and softmax scaling (softmax) methods [315]. One of the most common normalization methods is the z-score transformation [316] given by (3.2):

$$z_{ij} = \frac{((x_{ij}-\mu_j)/(3\sigma_j)+1)}{2} \tag{3.2}$$

where x_{ij} represents the value of the j_{th} feature of the i_{th} sample; μ_j and σ_j are the mean and standard deviation of the j_{th} feature, respectively.

Assuming that each feature is normally distributed, this transformation guarantees 99% of z_{ij} be in the [0,1] range. The out-of-range values are truncated to either 0 or 1. The normality of each feature distribution was verified using the moments of skewness and kurtosis (5% significance level). For more information about these tests, the reader is referred to [317].

Another form of z-score is given in [169] . The characteristics of the described values show uneven levels and unities, which can be a negative influence in the subsequent classification process. To solve it, an objective scale is applied between 1 and -1, obtaining the z-scores of the values of all the variables (3.3):

$$z_{ij} = \frac{x_{ij}-m_j}{\sigma_j} \tag{3.3}$$

where x_{ij} represents the ith value of the characteristic j. m_j and σj represent the average and typical deviation, respectively.

3.10 Classification

Classification phase of the diagnostic system is the one in charge of making the inferences about the extracted information in the previous phases in order to be able to produce a diagnostic about the input image. In the classification step, a classifier is trained

and tested, and then can be used to classify a new image. The supervised classification step includes two stages, namely, training and testing. At the training stage, a classifier is trained on a set of training feature vectors with an assigned label. At the testing stage, the classifier is applied to a set of test feature vectors that have been reserved for testing purposes (and are different from training feature vectors), and performance of the classifier is evaluated.

Predictive models are used in a variety of medical domains for diagnostic and prognostic tasks. These models are built from ‘‘experience’’, which constitutes data acquired from actual cases. The data can be pre-processed and expressed in a set of rules, such as it is often the case in knowledge-based expert systems, or serve as training data for statistical and machine learning models. Among the options in the latter category, the most popular models in medicine are logistic regression (LR), artificial neural networks (ANN), K nearest neighbour, decision trees, discriminant analysis and support vector machines. These models have their origins in two different communities (statistics and computer science), but share many similarities.

The task of classifying data is to decide class membership y’ of an unknown data item x’ based on a data set $D = (x_1, y_1) \ldots (x_n, y_n)$ of data items x_i with known class memberships y_i. For ease of discussion, only dichotomous classification problems are considered, where the class labels y are either 0 or 1. The x_i are usually m dimensional vectors, the components of which are called covariates and independent variables (in statistics parlance) or input variables (by the machine learning community).

In most problem domains, there is no functional relationship $y= f(x)$ between y and x. In this case, the relationship between x and y has to be described more generally by a probability distribution P(x,y); one then assumes that the data set D contains independent samples from P. From statistical decision theory, it is well known that the optimal class membership decision is to choose the class label y that maximizes the posterior distribution P(y|x) [318].

There are two different approaches to data classification: the first considers only a dichotomous distinction between the two classes, and assigns class labels 0 or 1 to an unknown data item. The second attempts to model P(y|x); this yields not only a class label for a data item, but also a probability of class membership. The most prominent representatives of the first class are support vector machines. Logistic regression, artificial neural networks, k-nearest neighbours, and decision trees are all members of the

second class, although they vary considerably in building an approximation to P(y|x) from data.

Classification typically uses standard pattern recognition techniques such as statistical methods like nearest neighbour and discriminate analysis as well as neural networks, decision tree based, SVM based classification. Numerous classification approaches are possible and are described in several text books. Examples of some other classification schemes reported in the literature are Bayes classifiers assuming multivariate Gaussian feature distributions, Fisher transformation nearest neighbour classifiers, classification trees, feed-forward neural networks, and the Learning Vector Quantization scheme. In some cases, even thresholding and simple extreme picking are applicable as classification schemes. Yet, other researchers report their results using clustering schemes, i.e., unsupervised schemes [319].

It is not intend in this chapter to delve deeply into the technical aspects of all the classification algorithms. However, to make the reader analyse the performance of algorithms that are mostly used for dermoscopic image analysis, it is believed that it is helpful to air them briefly. Readers who wish to have a detailed description of a specific classification approach should refer to the cited references.

3.10.1 K-nearest neighbor algorithm

The K-nearest-neighbour classifier [320, 321] is a nonparametric method of pattern recognition. For a lesion belonging to the test set (query vector), it finds the K vectors closest to the query vector in the training set. The unclassified sample is then assigned to the class represented by the majority of the K closest neighbours.

The most critical requirement of the K-nearest-neighbour classifier is to have a training set including enough examples of each class of pigmented lesions to adequately represent the full range of measurements that can be expected from each class. Optimizing the procedures of feature selection and weight definition could additionally improve the performance of the K-nearest-neighbour classifier [202].

The k-nearest-neighbours algorithm [321] is a popular density estimation algorithm for numerical data. In contrast to the other methods, this algorithm does not implement a decision boundary, but uses the elements of the training set to estimate the density distribution of the data. They implicitly combine this information with class prevalence in Bayes' rule to obtain the posterior (class membership) probability estimates

of a data point. The density estimation uses a distance measure (usually Euclidean distance). For a given distance measure, the only parameter of the algorithm is k, the number of neighbours. The parameter k determines the smoothness of the density estimation: larger values consider more neighbours, and therefore smooth over local characteristics. Smaller values consider only limited neighbourhoods. Generally, the choice of k can only be determined empirically. In some experiments [322] researchers used values of k 5 10, 20. . . 100. In medicine, most applications use nearest-neighbour algorithms as benchmarks for other machine-learning techniques [323, 324]. K-NN classifiers are especially successful while capturing important boundary details that are too complex for some other classifiers [187] . Classification based on the k-nearest neighbour algorithm differs from the other methods considered here, as this algorithm uses the data directly for classification, without building a model first [321, 325][326].

The advantage that k-nearest neighbour have over other algorithms is the fact that the neighbour can provide an explanation for the classification result. The major drawback of k-nearest neighbour lies in the calculation of the case neighbourhood: for this, one needs to define a metric that measures the distance between data items. In most application areas, it is not clear how to, other than by trial and error, define a metric in such a way that the relative importance of data components is reflected in the metric [322]. K-NN classification results are normally presented as specificity, sensitivity or classification rate in relation to the area size (k). For example Ruiz [169] showed k=7 showing good result, see Table 3.3.

Table 3.3 Algorithm output of KNN according to the area size [169]

Area (K)	Classification Rate (%)	Specificity (%)	Sensitivity (%)
5	35.71	31.91	39.22
7	73.47	70.21	76.47
9	70.41	68.09	72.55
11	66.33	63.83	68.63
13	34.69	34.04	35.29

3.10.2 Decision Trees

The decision tree approach is one of the most common approaches in automatic learning and decision making and it belongs to the supervised machine learning techniques. A Decision Tree Classifier is a predictive model, trained (or induced) by adopting a suitable dataset with respect to which classification results are already available. The decision tree paradigm constructs classifiers by dividing the data set into smaller and more uniform groups, based on a measure of disparity (usually entropy). This algorithm repeatedly splits the data set according to a criterion that maximizes the separation of the data, resulting in a tree-like structure [327-331].

The advantage of decision trees over many of the other methods used here is that small decision trees can be interpreted by humans as decision rules. They therefore offer a way to extract decision rules from a database. This makes them especially well-suited for medical applications, and advantages and disadvantages of decision trees in medicine have been widely investigated [332, 333].

Advantages and disadvantages of decision trees in medicine have been widely investigated [332, 333]. Compared with the other machine learning methods mentioned here, decision trees have the advantage that they are not black-box models, but can easily be expressed as rules. In many application domains, this advantage weighs more heavily than the drawbacks. This makes them especially well-suited for medical applications. In many classification tasks decision Tree Technique has been preferred to other solutions (also including Artificial Neural Networks and Support Vector Machines) because Decision Tree Classifiers are often fast to train, apply and generate easy to understand rules.

A major disadvantage of decision trees is given by the greedy construction process: at each step, the combination of single best variable and optimal split-point is selected; however, a multi-step look ahead that considers combinations of variables may obtain different (and better) results. Given a large training set, decision tree classifiers, in general, generate complex decision rules that perform well on the training data but do not generalize well to unseen data [334]. In such cases, the classifier model is said to have

overfit the training data. A further drawback lies in the fact that continuous variables are implicitly discretised by the splitting process, losing information along the way.

3.10.3 Logistic Regression

Logistic regression is an algorithm that constructs a separating hyper plane between two data sets, using the logistic function to express distance from the hyper plane as a probability of class membership.

Logistic regression is widely used in medical applications for the ease with which the parameters in the model can be interpreted as changes in log odds, for the variable selection methods that are often available in commercial implementations, and for allowing the interpretation of results as probabilities. Although the model is linear-in-parameters and can thus only calculate linear decision boundaries, it is nevertheless widely used predictive model in medical applications [335-337].

The main advantages of this method over other algorithms are its ease of use (it is implemented in numerous software packages), allowing the interpretation of results as probabilities and its variable-selection capability. The latter has only limited importance for the PSL classification tasks, since the input variables are obtained from an image segmentation algorithm and are not directly interpretable as humanly visible features of an image. Nevertheless, it is desirable to eliminate input variables that contribute only random correlations to the overall result. Dreiseitl et.al [322] showed in a comparative study that logistic regression performs on about the same level as artificial neural networks and support vector machines, which are both capable of implementing nonlinear separating surfaces.

3.10.4 ANN

Neural network is one of the great vital parts of soft computing. The artificial neural network (ANN) is basically a group of artificial neurons that are internally connected and are close together. It has conspicuous capacity to obtain idea from complex data and used to take out patterns and determine trends that are too difficult to be noticed by humans or any other computer skills. The major goal of ANN is to mimic the human ability to adapt to the changing circumstances and the current environment [180, 325, 338-340].

The general working mechanism for artificial neural network is presented in Figure. 3.6. Artificial neural networks (ANNs) represent a means to calculate posterior class membership probabilities by minimizing a cross-entropy error function. The ANN consists of several small processing units (the artificial neurons) that are highly interconnected. Information flow in an ANN is modelled after the human brain, in that information is propagated between neurons, with the information stored as connection strengths (called weights) between neurons. The minimization process is implemented as an update rule for the weights in the network. Since this iterative process requires many presentations of the training set, the system is said to learn from examples. It has conspicuous capacity to obtain idea from complex data and used to take out patterns and determine trends that are too difficult to be noticed by humans or any other computer skills. A lot of research is being carried out nowadays on dermoscopic image analysis using ANNs.

This machine learning method has received considerable interest over the past few decades for its promise to automatically "learn" structure from data. However, many of the early implementations required a significant amount of parameter tuning to achieve satisfactory results, a process that needed too much time and expertise for a non-expert. Over the past years, statistically motivated Bayesian methods [341] and implementations of faster learning algorithms [342] have allowed non-experts the use of sophisticated methods that require little to no parameter-tuning. Various Neural networks based clustering techniques and algorithms are being used in this regard [343] which includes Back Propagation Network (BPN), Radial Basis Function Network (RBF) and Extreme Learning Machine (ELM) etc.

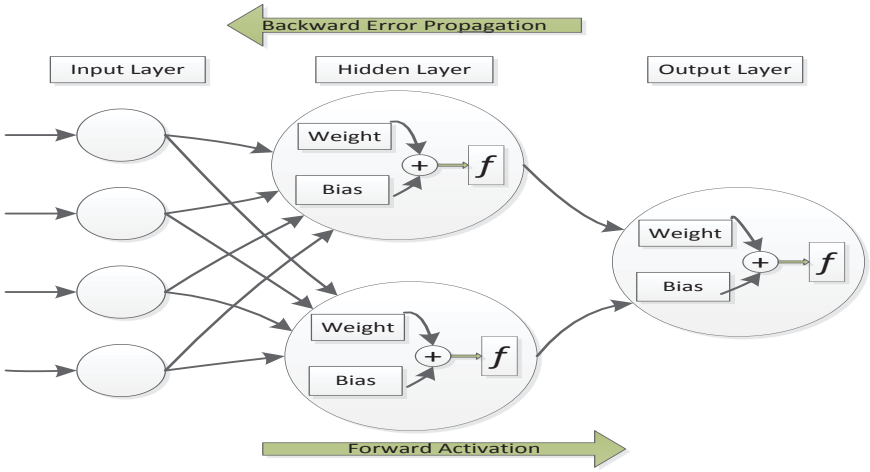


Figure 3.6 Working Mechanism of Artificial Neural Network

Back-propagation algorithm is one of the well-known algorithms in neural networks. Back-propagation algorithm has been popularised by Rumelhart, Hinton, and Williams in 1980's as a term for generalised delta rule [343].

BPN is a methodical system for multi-layer ANNs. It consists of three layers. The data is first given to the input layer. The data is then spread to the hidden layer and then to the output layer. It is also termed as the feed-forward network. BPN is a technique which is capable for altering the weights in a network, with distinguishable activation function, to learn data set [179, 180]. It works by propagating the input through the network and the error is calculated and then is propagated back through the network while the weights are adjusted in order to reduce the error [184].

The Back Propagation network has four stages during training phase. They include 1) weight initialization 2) Feed towards output, 3) errors of the network and 4) weight and biases updating.

3.10.4.1 Factors effecting neural networks training

In the research review carried out on existing automated skin lesion diagnostic systems[69], it was found that neural network performed well for skin lesion diagnostic systems. Because, neural network have better capability of handling complex relationships between the different parameters and making classification based on learning from the training data[322, 344]. Good training and validation of neural network can lead to better diagnostic results. It was found that success in training artificial neural networks can be based on the type of problem in the first step depends on model architecture (the number of neurons of output and input layers, the number of layers and the neurons of the hidden layer) and in the second step depends on used algorithm for training network. Some of the research results related to this are presented here.

Number of Input/output Neurons and hidden neurons:

When using ANN in different problems, the number of neurons in input and output layers is generally determined by the problem. High number of neurons in input and output layers makes the model size larger which may be harmful to learning quality and learning time. However, in applications like skin cancer diagnosis where several factors play role in diagnosis, it is usually difficult to restrict the inputs of the neural network and to ignore the impact of some factors for a better network learning. However, learning quality may improve by reducing the number of input variables, for which a

choice of minimal number of important input features is necessary.

There are no rigorous rules to guide the choice of number of hidden layers and number of neurons in the hidden layers. However, more layers are not better than few, and it is generally known that a network containing few hidden neurons generalizes better than one with many neurons. Also, the best network among a number of individually trained networks is then selected by trial and error.

Algorithm of training artificial neural network:

The learning method for popular artificial neural networks is often referred to as back-propagation algorithm. This was improved over time and various types of back-propagation algorithm were presented for training the network [345]. When using artificial neural networks, the proper learning algorithm is usually determined based on internal mental medications and via trial and error method [346]. The Perceptron neural networks learning algorithms chosen for this study are as follows:

A. LM (Levenberg–Marquardt) algorithm [347] is an approximation to Newton's method that finds a minimum of a function that is expressed as the sum of squares of linear functions. The Newton method approximates the error of the network with a second order expression, which contrasts to the former category which follows a first order expression. LM is popular in the ANN domain, which is a very simple, but robust, method for approximating a function.

B. SCG (scaled-conjugate gradient) algorithm: Scaled conjugate gradient algorithm developed by Moller [348], was designed to avoid the time-consuming line search. The basic idea is to combine the model-trust region approach (used in the LM algorithm), with the conjugate gradient approach. SCG differs from other conjugate gradient algorithms given by Hagan et al. [345] mostly in its use of nonlinear search technique. This makes the computations decreased in a single epoch.

C. Resilient Back Propagation (RP): When sigmoid functions are used as the functions for every activity on the set of hidden layer neurons, the large numbers face problem at the network entrance because they are ''squashing'' functions. This problem is solved in resilient back propagation algorithm only through employing the sign of derivation to determine the direction of the weight update; the magnitude of the derivative has no effect on the weight update.

This research study was mainly aimed at providing performance analysis of these learning algorithms for the case of pattern recognition for skin cancer diagnosis.

3.10.5 Support Vector Machines

Support vector machines (SVMs) are a machine learning paradigm based on statistical learning theory[349] [350]. Performances on par with or exceeding that of other machine learning algorithms have been reported in the medical literature. Algorithmically, support vector machines build optimal separating boundaries between data sets by solving a constrained quadratic optimization problem [351]. While the basic training algorithm can only construct linear separators, different kernel functions (i.e. Linear, Polynomial, Radial basis function, sigmoid) can be used to include varying degrees of nonlinearity and flexibility in the model. (Figure 3.14)

SVMs have several advantages over the more classical classifiers such as decision trees and neural networks. The support vector training mainly involves optimization of a convex cost function. Therefore, there is no risk of getting stuck at local minima as in the case of back propagation neural networks. SVMs are based on the structural risk minimization (SRM) principle which minimizes the upper bound on the generalization error. Therefore, SVMs are less prone to over-fitting when compared to the algorithms such as back propagation neural networks that implement the ERM empirical risk minimization principle. Another advantage of SVMs is that they provide a unified framework in which different learning machine architectures (e.g., RBF networks, feed forward neural networks) can be generated through an appropriate choice of kernel [208]. The disadvantage of support vector machines is that the classification result is purely dichotomous, and no probability of class membership is given.

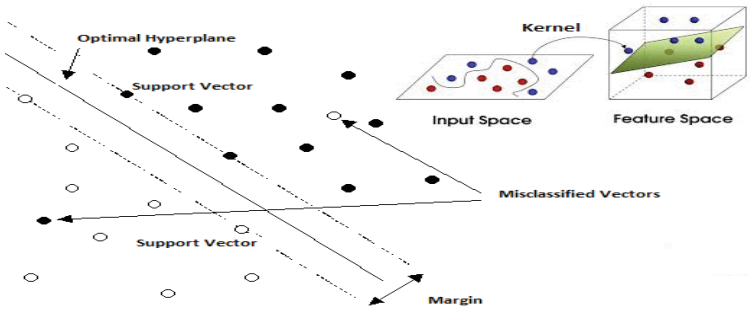


Figure 3.7 Principle of Support Vector Machine

Support vector machines (SVMs) are a machine learning paradigm based on statistical learning theory[349] [350]. Although the theory of support vector machines was developed more than 20 years ago, this paradigm has only recently been widely applied by the machine learning community. Many applications of this method in the medical domain have been reported in literature. The most attractive feature of this paradigm is that it is possible to give bounds on the generalization error of the model, and to select the best model from a class using the principle of structural risk minimization.

Support vector machines calculate separating hyper-planes that maximize the margin between two sets of data points. By using Lagrange multipliers, the problem can be formulated in such a way that the only operations on the data points are the calculation of scalar products. While the basic training algorithm can only construct linear separators, kernel functions can be used to calculate scalar products in higher dimensional spaces. If the kernel functions are nonlinear, the separating boundary in the original space will be nonlinear. Because there are many different kernel functions, there is a wide variety of possible SVM models.

Support vector machines (SVMs) have drawn considerable attention in the machine learning community due to their solid theoretical foundation and excellent practical performance. They are kernel-based learning algorithms derived from the statistical learning theory [350]. SVMs have several advantages over the more classical classifiers such as decision trees and neural networks. The support vector training mainly involves optimization of a convex cost function. Therefore, there is no risk of getting stuck at local minima as in the case of back propagation neural networks. Most learning algorithms implement the empirical risk minimization (ERM) principle which minimizes the error on the training data.

On the other hand, SVMs are based on the structural risk minimization (SRM) principle which minimizes the upper bound on the generalization error. Therefore, SVMs are less prone to over-fitting when compared to the algorithms that implement the ERM principle such as back propagation neural networks. Another advantage of SVMs is that they provide a unified framework in which different learning machine architectures (e.g., RBF networks, feed forward neural networks) can be generated through an appropriate choice of kernel. (obtained from [208])

Support vector machine performs classification by constructing an N-dimensional hyper-plane that optimally separates the data into two categories. The goal of SVM modelling is to find the optimal hyper-plane[30] that can separate clusters of vectors in such a way that cases with one category of the target variable are on one side of the plane and cases with the other category are on the other side of the plane. The vectors which are near the hyper-plane are the support vectors. SVM tries to minimize the empirical error while controlling the complexity of the mapping function. This enables it to achieve much better generalization performance on new data.

This subsection gives an overview of the SVM theory. SVM constructs a decision surface between samples of the two classes, maximizing the margin between them. Consider a set of n training data points [352] ∈Rd ×{−1, +1} i=1, . . ., n , where xi represents a point in d-dimensional space and yi is a two-class label. Suppose there is a hyper plane that separates the positive samples from the negative ones. Then the points x on the hyper plane satisfy w · x + b = 0, where w is the normal to the hyper plane, and |b|/||w|| is the perpendicular distance from the hyper plane to the origin. If two such hyper planes are considered between the positive and negative samples, the support vector algorithm's task is to maximize the distance (margin) between them. In order to maximize the margin, ||w||2 is minimized subject to the following constraints:

$$y_i(x_i \cdot w + b) - 1 \geq 0 \ \forall i \qquad (3.4)$$

The training samples for which (3.33) hold are the only ones relevant for the classification. These are called the support vectors. The Lagrangian function for the minimization of ||w||2 is given by:

$$L_1 = \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} y_i y_j \alpha_i \alpha_j x_i x_j \ \ subject\ to\ \ \alpha_i \geq 0 \forall \ \& \sum_{i=1}^{n} \alpha_i y_i = 0 \qquad (3.5)$$

Equation (3.34) applies only to linearly separable data. In order to handle non-linearly separable data, positive slack variables ξi, i=1, . . ., n are introduced into the constraints:

$$y_i(x_i . \omega + b) \geq 1 - \xi_i \ , \ \xi_i \geq 0 \forall i \ \text{ b} \qquad (3.6)$$

In order to control the trade-off between the model complexity and the empirical risk, a penalty parameter C (cost/penalty) is introduced into (3.36):

$$L_{nl} = \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} y_i y_j \alpha_i \alpha_j x_i x_j \ \ sub\_to\ 0 \leq \ \alpha_i \geq C \ \forall i \ \& \sum_i \alpha_i y_i = \ 0 \qquad (3.7)$$

To generalize these equations for non-linear decision functions, the concept of a kernel is introduced. The data seen in the equations so far appears in the form of dot products xi·xj. If the data was to be mapped to some other (possibly infinite dimensional) Euclidean space H, using a mapping _, the training algorithm would depend on the data through dot products in H, i.e. ∅ (xi)· ∅ (xj). Now, if there were a "kernel function" K such that K(xi,xj) =∅ (xi)· ∅ (xj), it would only need to use K in the training algorithm, and would never need to explicitly know what ∅ is. Substituting the kernel K into the dual SVM gives:

$$L_k = \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} y_i y_j \alpha_i \alpha_j K(x_i x_j) \; subject\ to\ 0 \leq \alpha_i \geq C \; and \sum_i \alpha_i y_i = 0 \tag{3.8}$$

This formulation allows us to deal with extremely high (theoretically infinite) dimensional mappings without having to do the associated computation. Some commonly used kernels are:

$$Linear: K(x_i, x_j) = {x_i}^T . x_j \tag{3.9}$$

$$Polynomial: K(x_i, x_j) = (\gamma {x_i}^T . x_j + r)^d , \gamma > 0 \tag{3.10}$$

$$Radial\ basis\ function\ (RBF): K(x_i, x_j) = e^{-\gamma \|x_i - x_j\|^2 / 2\sigma^2} , \ \gamma \tag{3.11}$$

$$Sigmoid: K(x_i, x_j) = \tan h\ (\gamma {x_i}^T . x_j + r) , \gamma > 0 \tag{3.12}$$

The most popular kernel functions are polynomials and Gaussian radial basis functions (RBFs). For polynomial kernels, the adjustable parameter is the degree of the polynomial; for Gaussian RBF kernels, it is the inverse variance. For any kernel function, it is also necessary to specify a cost factor C that determines the importance of misclassifications on the training set. Dreiseitl et. al [322]showed in a study that for the polynomial kernels, convergence times depended heavily on the degree of the kernel polynomial, with times for degree four kernels too slow. Gaussian RBF kernels were generally fast to converge, and did not depend as heavily on the choice of precision parameter γ.

In many studies, the radial basis function (RBF) kernel was adopted for various reasons. Firstly, the linear kernel cannot handle nonlinearly separable classification tasks, and in any case, is a special case of the RBF kernel. Secondly, the computation of the RBF kernel is more stable than that of the polynomial kernel, which introduces values of

zero or infinity in certain cases. Thirdly, the sigmoid kernel is only valid for certain parameters. Finally, the RBF kernel has fewer hyperparameters (γ) (γ-kernel width) which need to be determined when compared to the polynomial (γ,r,d) and sigmoid kernels (γ,r).

3.10.6 Evaluation of Classification Performance

Evaluation of classification results is an important process in the classification procedure. For skin lesion classification two different classification tasks are used as benchmarks: the dichotomous problem of distinguishing common nevi from dysplastic nevi and melanoma and the dichotomous problem of distinguishing melanoma from common nevi and dysplastic nevi.

The two criteria to assess the quality of a classification model are discrimination and calibration. Discrimination is a measure of how well the two classes in the data set are separated and calibration is a measure of how close the predictions of a given model are to the real underlying probability based on expert knowledge. Some of the common measures of analysing discriminatory power of different methods are reported here. See Table 3.5.

Sensitivity and specificity are the most commonly used performance evaluation parameter in literature. Accuracy can be used as a single parameter but if there is imbalance between the classes (melanoma, benign) then accuracy is not a suitable approach of evaluation. A better performance measure in unbalanced domains is the receiver operating characteristic (ROC) curve. The log diagnostic odds ratio is also sometimes used in meta-analyses of diagnostic test accuracy studies due to its simplicity (being approximately normally distributed). d_{class} is a measure to compare different classifiers presented by Sboner et. al [326]that enable us to give a simple estimation of how useful one classifier is with respect to another. By using this parameter instead of accuracy, the comparison between classifiers can be carried out in an accurate but intuitive way, avoiding the unbalanced class problem.

Table 3.4 Measures for Evaluating Performance of a Classifier

Evaluation Parameters
$\textbf{Accuracy} = \frac{\textbf{TP} + \textbf{TN}}{\textbf{TP} + \textbf{TN} + \textbf{FP} + \textbf{FN}}\ \textbf{x}\ 100\%$
$\textbf{Diagnostic Accuracy} = \frac{\textbf{TP}}{\textbf{TP} + \textbf{FP} + \textbf{FN}}$
$\textbf{Sensitivity} = \frac{\textbf{TP}}{\textbf{TP} + \textbf{FN}}\ \textbf{x}\ 100\%$
$\textbf{specificity} = \frac{\textbf{TN}}{\textbf{TN} + \textbf{FP}}\ \textbf{x}\ 100\%$
Classification rate = No of images correctly classified/ total no of images
$\textbf{Positive Predictive Value} = \frac{\textbf{TP}}{\textbf{TP} + \textbf{FP}}\ \textbf{x}\ 100\%$
$\textbf{Negative Predictive Value} = \frac{\textbf{TN}}{\textbf{TN} + \textbf{FN}}\ \textbf{x}\ 100\%$
$\textbf{Error Probability} = \frac{\textbf{FP} + \textbf{FN}}{\textbf{TP} + \textbf{TN} + \textbf{FP} + \textbf{FN}}\ \textbf{x}\ 100\%$
$\textbf{LR}_{+} = \frac{\textbf{Sensitivity}}{\textbf{1} - \textbf{Specificity}}$
$\textbf{LR}_{-} = \frac{\textbf{1} - \textbf{Sensitivity}}{\textbf{Specificity}}$
$\textbf{Diagnostic odds ratio DOR} = \frac{\textbf{TP/FN}}{\textbf{FP/TN}}$
distance of a real classifier from the idea one $\mathbf{d_{class}} = \sqrt{(\mathbf{1} - \mathbf{Se})^2 + (\mathbf{1} - \mathbf{Sp})^2}$

To provide an unbiased estimate of a model's discrimination and calibration there are some important considerations like effect of class imbalance, train/ test ratio and cross validation. Several studies have demonstrated that the accuracy degradation on

unbalanced data sets is more severe when the classes overlap significantly [353-355] this is the case in skin lesion classification. Most classifiers focus on learning the large classes which leads to poor classification accuracy for the small classes such as classifying the minority (melanoma) samples as majority (benign) implies serious consequences.

Train to test ratio is another important factor effecting the classification result. . It has been observed [195] that as the training-set size increases, however, it is observed that over training may also lead to less test accuracy.

There are two approaches for selecting training and test data; either to separate test and training feature vectors or pick training feature vectors as a subset of the test vectors. A classification result may be overly optimistic if performance cannot be measured on a data set not used for model building. In the ideal case, testing on a separate data set will provide an unbiased estimate of generalization error. If the original data set is too small for this approach, the recommended strategy is to use cross-validation [356] or bootstrapping [357] to make the best possible use of the limited amount of data. One way is to divide the whole data into n pieces, n-1 pieces used for training, and the last piece as the test set. This process of n-fold cross-validation builds n models; the numbers reported are the averages over all n test set. The extreme case of using only one data item for testing is known as leave-one-out cross-validation. Bootstrapping is rarely used in literature for skin lesion case but it has shown to be superior to cross-validation on many other data sets [358].

3.10.7 Selection of suitable classification method

The increasing number of electronic data bases containing dermoscopic images has led to an increasing interest in their utilization for building classification models that can “learn” from examples. The need to use data and learning techniques in order to make correct diagnosis requires proper choice of the learning algorithms and of their statistical validation. The problem is difficult given the relative paucity of lesion data and consequently the low quality of training data available and the imbalance between the classes. A variety of statistical and machine learning approaches are used for classification of dermoscopic images. See Table 3.6.

Table 3.5 Classification Methods used in Literature for Skin Cancer Diagnosis

Classification method	Related references
K – nearest neighbour	[202], [322], [187], [168], [169]
Decision trees	[322],[181],[253], [185] ADWAT [213], [359], Logistic Model tree (LMT) [188], [189], [170], CART [360], [98]
Statistical (discriminant analysis / logistic regression/ multifactorial analysis)	[195], [216],[232],[361],[196], [152], [250] DA [362], [233], [201], [363]mathematical classifier[364], logistic regression [90], [203], [204], [206], [15], [322]linear classifier [202]
Rule based classification	[182, 365, 366],[235],[365],[182],[367], [368]
Artificial Neural Network	[369],[370],[371], [372], [93],[361], [322],[166],[250], [211],[363],[114], [184], [214], [368], [210], [209], [194], [207], [169], [243],[180]
Support vector machine (SVM)	[150], [322], [92],[134],[30], [208], [184], [214], [187],[373] , [209], [194], [98]
Extreme learning machine	[180]
Others (Gaussian maximum likelihood, Bayseian classifier)	[187], [169], [98]

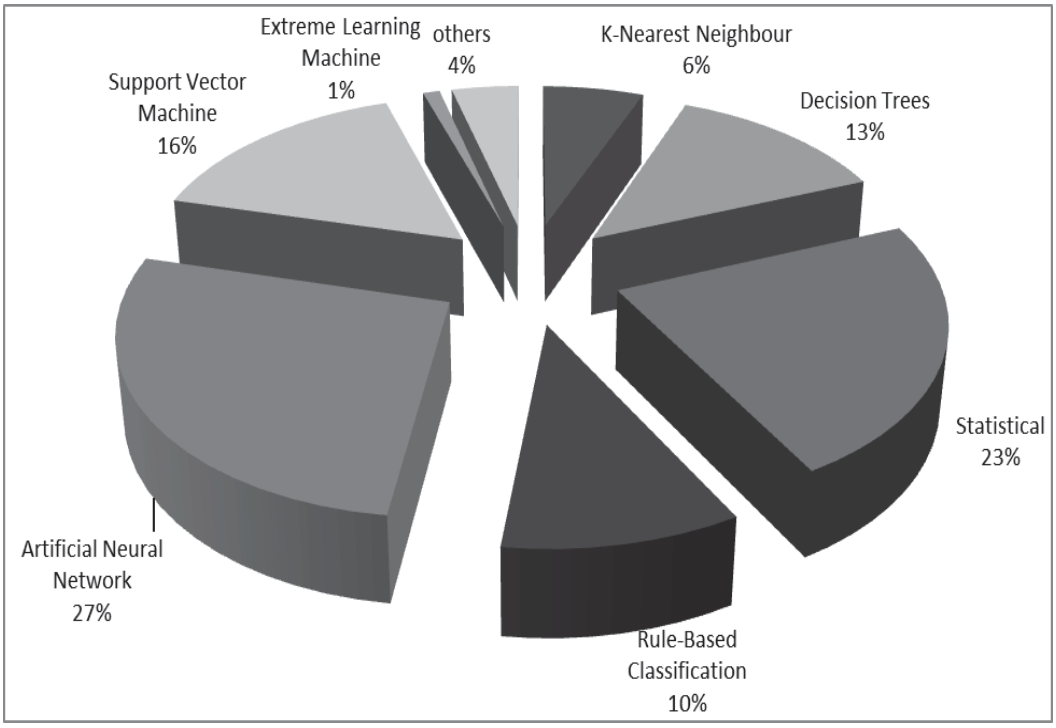


Figure 3.8 Illustration of Classification methods as used by existing diagnostic systems in literature

Different classification methods have their own merits**. Figure 3.15 shows percentage of different classification methods as used by existing diagnostic systems in literature, based on our review of around 300 relative studies.** The question of which classification approach is suitable for a specific study is not easy to answer. Different classification results may be obtained depending on the classifier(s) chosen, differences in sample sizes, proportion of melanomas in the sample and number of features used for discrimination. See Table.3.7. Many factors, such as different sources of obtaining dermoscopic images, availability of classification software, time consumption, computational resources and number of melanoma and benign images available for training must be taken into account when selecting a classification method for use.

Table 3.6 Summary of Classification Performance of some skin Cancer Detection Methods Present in Literature

Source	year	No. of features selected	classifier	Total images	Melanoma %	Dysplastic nevi %	Benign	Sens %	Spec %	accuracy
[364]	1993		CART	353	62		38	94	88	
[104]	1994		CART	404	59		41	90	88	80
[233]	1994	22	D.A.	164	11		89	88	89	
[369]	1994		ANN	200	40	30	30	95	88	
[239]	1994	14	ANN	240	50	16.7	33.3	79.5	86.3	82.9
				216	50		50	86	85.5	85.7
[374]	1997		Logistic regression	170	44		56	93	67	
[375]	1998	22	Discriminant analysis	917	7		93	93	95	
[371]	1998	16	ANN	120	32.5	48.4	27.5	90	74	

[376]	1999	26	Discriminant analysis	383	4.7		95.3	100	92	
[372]	1999	26	ANN	44	43.2		56.8	97.7	100	
[201]	1999	13	Discriminant analysis	147	38.8		61.2	88	81	85
[97]	2000	38	ANN	315	13.3		86.7	92.9	97.8	
[377]	2001	21	kNN	5363	1.8	18.8	79.4	87	92	
[116]	2001	13	Linear classification	246	25.6	45.1	29.2	100	85	
[93]	2002	13	ANN	588	36.9		63.1			94
[361]	2002	10	ANN	147	38.8		61.2	93	92.8	
[152]	2003	1	Linear classifier	100	50		50	78	90	
[326]	2003	38	LDA + k-NN + Decision tree	152	27.6		72.4	81	74	
[202]	2004	10	Linear classifier	840	46.5		53.5	95	78	
			KNN					98	79	

[150]	2004		SVM (third degree polynomial)	977	5.12		94.88	96.4	87.16	
[92]	2005	NR	SVM	477	8.8		91.2	84	72	
[363]	2005	20	ANN	34	41	59		86	100	
			Discriminant analysis					79	90	
[30]	2006	200	SVM	22	45		65			70
[181]	2006	28	Decision tree	224	51.8		48.2	51	97	
[204]	2006	3	LR+ multivariate model + ROC	132	17.4		82.6	60.9	95.4	89.4
[208]	2007	18	SVM	564	15.6		84.4	93.3	92.3	
[206]	2007	2	Logistic regression (LR)	260	17.7	18.1	64.2	91.3	81-91	
[187]	2008	10	Multiple classifiers (SVM, GML, kNN)	358	37.4	32.96	29.6			75.69

[168]	2011	33	kNN	83	55.4		44.6	60.7	80.5	66.7
[169]	2011	6	K-NN	98	52		48	76.4	70.21	73.47
			Bayesian					76.4	85.11	80.61
			Multi-layered perceptron					78.4	95.74	86.73
			Combination of three					78.4	97.87	87.76
[243]	2012	12	Automatic multi-layerperceptron	102	50		50	70.5	87.5	76

Very few researchers provided comparisons of different classification algorithms using same set of images [98, 169, 209, 214, 322]. The review of all these comparative studies reveal that Multiple Layer Perceptron (MLP) gives better performance than bayesian and kNN classifier, while SVM with RBF kernel normally outperform MLP, decision trees and other statistical methods. The results of an experimental assessment of the different designs can be the basis for choosing one of the classifiers as a final solution to the problem.

It had been observed in such design studies, that although one of the designs would yield the best performance, the sets of patterns misclassified by the different classifiers would not necessarily overlap. These observations motivated the relatively recent interest in combining classifiers. The idea is not to rely on a single decision making scheme. Instead, all the designs, or their subset, are used for decision making by combining their individual opinions to derive a consensus decision. Some classifier combination schemes

have been devised [169, 187, 326] for dermoscopic images and it has been experimentally demonstrated that some of them consistently outperform a single best classifier. However, there is presently inadequate understanding why some combination schemes are better than others and in what circumstances.

3.11 Model Validation

A vast number of diagnostic algorithms/models are published each year. Such models do not always work well in practice, so it is widely recommended that they need to be validated [378, 379]. To be useful, a prognostic index should be clinically credible, accurate, have generality (that is, be validated elsewhere), and the study should be described in adequate detail. To gauge the current state of reporting results in the literature, many papers were sampled on dermoscopic images data sets analysis.

31 publications were reviewed in detail which claimed fully automatic diagnostic models. It was found frequent shortcomings both reporting and methodology used, refer to Table 3.8. It includes lack of calibration in image acquisition, unspecified method for extracting and selecting variables in the model, and risk of over-fitting through too few events per variable. Many researchers did not specify the test/train or used uneven number of melanoma and benign images for training which may lead to biased classification. Some articles do not report comparisons and cross validation, instead they just reported the performance of a single method. It is imperative that these details be presented in papers as otherwise the validity of the claims in the papers cannot be assessed by the reader.

When assessing the quality of the results obtained using any diagnostic models one should consider the quality of the data set employed in model-building, the care with which adjustable model parameters were chosen, and the evaluation criteria used to report the results of the modelling process. This is important in distinguishing between overly optimistic claims (such as when performance is reported on the training set) and needlessly pessimistic ones (when model parameters are chosen in a suboptimal manner). The latter is especially common in studies that promote ‘‘new’’ algorithms.

Apart from all this, in order to judge the performance of an automatic diagnostic model it is important to mention that who is going to use that model. If automated diagnostic systems will be used by general practitioners or in pharmacies and shopping

centres, these systems should be worked with very high sensitivity and reasonably good specificity. That is, it should recognize the greatest number of melanomas in early stage, without misclassifying too many nevi so that unnecessary excision of benign lesions could be avoided. If the target is the expert user, studies should be designed with the aim to help clinicians in distinguishing between benign lesions, dysplastic nevi and malignant tumors of the skin. An increase in specificity might be the goal for an automated system directed to expert users together with sensitivity at least equal to that achieved by the expert.

Table 3.7 Percentage of Diagnostic Models Satisfying Quality Assessment Criteria

Criteria	Details Provided (% of models)	Details not Provided (% of models)
Image calibration	51	49
Pre-processing	45	55
Segmentation	78	22
Feature Extraction	71	29
Feature Selection	54	46
Test/train ratio	42	58
Taking care of balance in lesion classes for training	32	68
Comparative results	55	45
Cross validation	29	71

Overall, the objective is to get a classifier with the sensitivity and the specificity balanced. It should be noted that the ability to diagnose correctly melanoma is by far the most important property that an automated system must have. The consequence of failure to diagnose correctly a malignant tumor may lead to the eventual death of the patient. On the other hand, if a classifier gives a high sensitivity but a low specificity, it is not going

to be useful as a screening method to avoid biopsies (an invasive technique). And, of course, a classifier with a high sensitivity is required to avoid false negatives.

3.12 Summary

The basic aim of this chapter is to provide the background of the research contributions presented in the later chapters of the thesis. A review of the current practices, problems, and prospects of image acquisition, pre-processing, segmentation, feature extraction & selection and classification stages is provided for the development of automated diagnostic systems for skin cancer images. The statistics and results of some of the important implementations found in literature are discussed. Indications of various conditions that affect the techniques performance are reported. This research analysis forms the basis of the methods proposed in this thesis for getting improved performance during different stages of the overall process. The shortcoming in segmentation, feature selection and classification phases found during the literature lead to the development of more effective algorithms for these stages which will be present in the next three chapters.

Chapter 4

Thesis Contributions to Segmentation Methods

4.1 Introduction

This chapter presents the segmentation stage of the automated diagnostic system. The aim of this stage is to find the region of interest or in other words identify the affected area present in the overall image under consideration. Background of the proposed segmentation algorithm is presented. Afterwards, the details of the proposed algorithm are provided along with the segmentation results obtained. Finally, the proposed method is compared with other most popular segmentation methods used in literature for the detection of skin lesion area.

The proposed segmentation method is mainly designed for digital or dermoscopic images of skin lesions. However, as in the later diagnosis stages for skin cancer, the study of histopathological images taken from the biopsy samples of patients is also of great importance. Therefore, a modified version of the proposed segmentation method for the histopathological images is also provided in this chapter, which can deal with more complex details that need to be considered in case of histopathological image analysis.

Parts of the work presented in this chapter are published in peer reviewed articles that include 1 article in Journal of Signal and Information Processing [380], 1 book chapter in Lecture notes in Computer Science [381], and 2 other International Conferences [382, 383].

4.2 Background of Proposed Method

Segmentation of skin lesions is difficult because of the great variety of lesion shapes, sizes, and colors' along with different types and textures of human skin. Other difficulties that make it a challenging task include, low contrast between the lesion and the surrounding skin, smooth transition between the lesion and the skin, irregular and fuzzy lesion borders, reflections and shadows due to wrong illumination, artifacts such as skin texture, air bubbles and hair and variegated coloring inside the lesion.

The underlying objective of image segmentation is to separate the regions of interest from their background and from other components present in the image. As segmentation is a crucial early step in the analysis of lesion images, so, it has become one of the important areas of research and many algorithms and segmentation techniques are available in literature. As discussed in the methodology review in chapter 3, Image segmentation algorithms available in literature can be broadly classified into two categories: 1) discontinuity based segmentation 2) similarity based segmentation, which can further be classified into thresholding approach, clustering approach and region based approach. The former often depends on intensity gradients, while the latter takes advantage of pixel intensities directly. In addition to this, various soft computing methods have been applied to image segmentation too.

Cutaneous lesions come in a great variety of shapes, colors, and textures due to which the segmentation problem itself is very hard. It is unlikely that any one algorithm will be able to delineate all lesions satisfactorily. There appears to be a fairly general agreement that there is neither an ideal segmentation algorithm that can detect all kinds of skin lesion nor all methods are equally good for a particular type of image. A score of articles enumerating advantages and limitations of certain segmentation techniques are available in literature [25, 160, 174, 178, 384-388]. Each algorithm has advantages along with certain limitations that provide ground for future research and investigation.

Image segmentation using histogram-based thresholding [389] is probably one of the most common approaches because it requires less CPU resources and is easy to implement. Thresholding methods are generally based on maximization or minimization of certain criterion function and optimal threshold is the intensity level where the criterion function attains its maximum or minimum respectively. Thresholding can be global if a single threshold is calculated or it can be local thresholding if the image is divided into sub-blocks and a threshold is calculated for each block. Although thresholding is a simple approach for segmentation but the major disadvantage of this approach is that it usually disregards the spatial context within which the intensity value occurs. Methods based on 2-D histogram consider the contextual information to some extent but still more sophisticated approaches are required in this regard.

However, thresholding methods can form a good base for providing initial segmentation prior to the application of a more sophisticated algorithm and this can help in reducing the convergence time of complicated algorithms.

While considering more sophisticated methods, active contours or deformable models are among the most popular ones for image segmentation. Active contour is a popular approach used to estimate boundaries in medical images as well. Two types of algorithms lie under this category: 1) parametric active contours [390] which adapt a deformable curve until it fits the object boundary. 2) Geometric active contours based on level set theory [391, 392]. Some of the active contour models need user intervention for initialization. Thus automatic approaches like gradient vector flow algorithm based on anisotropic diffusion [393] and robust algorithms like adaptive snakes and shape probability association model [394] are taking important place in literature. The idea behind them is quite straightforward. An initial guess is specified, and then the contour or the model evolves by itself. If the initial model is parametrically expressed, it operates as snakes. In contrast, the level set methods do not follow the parametric model, and hereby own better adaptability.

Level set (LS) methods, as one of the automatic process approaches has shown effective results for medical image segmentation. But the level set methods, without the parametric form; often suffer from a few refractory problems, for example, boundary leakage and excessive computation. Thus, intensive computational requirements and regulation of controlling parameters make it a complex and time consuming method. To overcome such shortcomings, a good initialization is very important in level set methods used for image segmentation. Li.et al used fuzzy clustering to facilitate the LS segmentation for automatic segmentation of ultrasound, computed tomography and magnetic resonance imaging [395, 396]. Abbas et.al [139] proposed to use adaptive thresholding to initialize level set automatically. Application of these methods for skin lesion segmentation provides reasonable results. However, these techniques have several drawbacks including intensive computational burden and need of many pre-processing steps in the presence of intensity variety and several artifacts that are commonly observed in skin lesion images.

4.3 Proposed Approach

A new histogram analysis based fuzzy C mean thresholding method is developed to automatically initialize level set. A 3-class FCM based thresholding followed by some morphological operations is used for initializing the LS evolution and regulating the controlling parameters. As compared to other segmentation methods based on Fuzzy c-mean thresholding and level set algorithms, **this new algorithm is significantly improved in following aspects. 1) It unifies the advantages of fuzzy c mean clustering and adaptive thresholding along with reducing the computational complexity. 2) Effective threshold value can be calculated even in the presence of intensity variety. 3) Interfacing of few morphological operations within the segmentation method reduce the need of intensive pre-processing requirements for reducing effect of various commonly observed skin lesion artifacts like hair, skin lines etc., 4) computational complexity of level set evolution is reduced by providing good initialization and automatically calculating effective controlling parameters.** Finally, the algorithm was also verified on a large set of skin lesion images having almost all types of expected artifacts that may badly affect the segmentation results. Performance evaluation of proposed method is done by comparing the diagnosis results of this method with other state of the art methods, including FCM clustering [157], k-mean clustering[397], region based active contours [139] , adaptive thresholding [162], iterative thresholding based level set [139] and spatial fuzzy clustering based level set[395]. Results show that the proposed approach performs reasonably well as compared to the aforementioned methods for the segmentation of skin lesion images. The details of main parts of the proposed algorithm are provided as follows.

4.3.1 Image Pre-processing

In order to reduce the effect of extraneous artifacts such as skin texture, air bubbles, dermoscopic gel, presence of ruler markings and hair on segmentation results, the images were pre-processed with a smoothing filter. Among the different filters present in literature, it was found that a 7x7 mask median filter served the required purpose in the best way. Thus, the first step of the overall process is to get a filtered image. Various pre-processing steps are present in literature for specular reflection reduction (e.g. histogram equalization/stretching, independent histogram pursuit (IHP) [141] , homomorphic filtering [142] etc. for contrast enhancement and compensating uneven illumination

effects) and special hair-removing algorithms (e.g. mathematical morphology [147, 148, 151, 398], inpainting techniques[155, 156] and fast marching algorithms [158] etc.). The proposed segmentation algorithm does not need these special pre-processing steps separately for digital and dermoscopic skin lesion images as it manages to handle these problems during the segmentation process.

4.3.2 Histogram analysis based Fuzzy C Mean thresholding (H-FCM)

The classical segmentation algorithms perform excellently when classes are clearly distinguishable. However, in a number of real life situations, such as segmentation of skin cancer images, such a hard demarcation is quite difficult. Instead of the conventional deterministic assignment of a sample to a class, fuzzy partitioning strategies provide soft description of the classes, where each of the sample points is assigned a membership in each of the classes. Such a partition reflects the identities of the samples and thereby preserves the structural details in the original data set.

Clustering in general is also known as unsupervised classification that looks for some correlation or natural groupings in a multidimensional data set by employing a similarity or dissimilarity measure. K-mean clustering and Fuzzy C mean clustering are two well-developed methods under this category.

K-mean clustering is computationally very efficient and gives satisfactory results if the clusters are well separated and compact in the feature space. However, in the k-mean algorithm, every image pixel is limited to one and only one of k clusters, which is not justifiable in most of the cases in medical image segmentation. Unlike this, the Fuzzy C mean clustering (FCM) algorithm proposes a fuzzy membership that assigns a degree of membership for each class. The concept of degree of membership in fuzzy clustering is quite similar to the posterior probability in a mixture modelling setting. This clustering is achieved by iteratively minimizing a cost function that is dependent on the distance of the pixels to the cluster centres in the feature domain. By monitoring data points that have close membership values to existing classes, forming new clusters is possible. This is the major advantage of FCM clustering over K-mean clustering and the reason why it has been widely utilized for medical image segmentation where clear demarcation between organs is not easily visible.

The concept of fuzzy partitioning can be extended for digital image thresholding by visualizing the object and background regions as fuzzy sets, O and B, with each of the pixels showing a partial membership in each of the regions according to its gray value, i.e μ_O (I_{mn}) ϵ [0, 1], μ_B (I_{mn}) ϵ [0, 1]. With this sort of a partition, regions are no more guaranteed to be mutually exclusive; in other words, there may not exist a crisp boundary between regions.

A natural extension of fuzzy clustering for segmentation by considering the gray value alone as feature, leads to the thresholding formulation. A fuzzy thresholded description of an image can be characterized by two membership functions μ_O and μ_B in such a way that they reflect the nature of object and background gray distribution even after thresholding. Pal [399] extended the definition of classical thresholding from (4.1)

$$O = \{I_{mn} | I_{mn} \geq T\}, B = \{I_{mn} | I_{mn} < T\} \quad (4.1)$$

to fuzzy setting by defining a membership function as follows:

$$\mu_O(I_{mn}) < 0.5 \; if I_{mn} < T, \mu_O(I_{mn}) > 0.5 \; if I_{mn} > T \quad (4.2)$$

With the crossover membership distribution coinciding with the hard threshold T.

There are mainly two main approaches present in literature to implement FCM based thresholding. First one is by searching the extreme for all possible thresholds, while the second one is based on identifying the optimal combination by iteratively evaluating the criteria function and updating the threshold value accordingly. A look on the prospects of both iterative and the non-iterative implementation present in literature [400] supports the efficient iterative implementation, provided it converges.

On the other hand, preserving structural details is important for diagnosis of skin lesions because the boundary of lesion provides details about asymmetry and degree of abruptness in the structure which is an important differentiating feature between melanoma and benign lesions. The classical thresholding schemes assign the pixel unequivocally without distinguishing among pixels in a region even if their gray values are different in the original image. This leads to the loss of structural details on thresholding. Thus, use of simple hard thresholding schemes can result in loss of structural details, so such thresholding methods do not seem to be a good choice for segmentation of skin lesions. Structural details can be preserved in fuzzy partition space,

because the memberships assigned to the pixels depend on the difference between its gray value and the mean gray value of the region to which it belongs.

It can be seen that FCM clustering and histogram based thresholding both have some distinct benefits. Histogram based thresholding is based on finding hard thresholds which can be a good approach for images with well-separated modes in the histogram. On the other hand FCM clustering for gray level images allows categorizing the pixels in more than two classes based on intensity values. The method presented in this thesis unifies hard and fuzzy thresholding schemes. A 3-class histogram analysis based fuzzy c mean thresholding is proposed which combines the benefits the advantages of fuzzy c mean clustering and histogram analysis methods. First the pixels are divided into three classes based on their intensity value. Secondly histogram analysis of image is done to see average intensity distribution in the images and then the hard threshold is selected between classes with dominant intensity values. Thus, this method solves the difficulty that may arise in finding effective threshold automatically for different images regardless of intensity variety. The proposed thresholding algorithm is presented in Figure.4.1. The steps of the overall algorithm are explained in detail as follows.

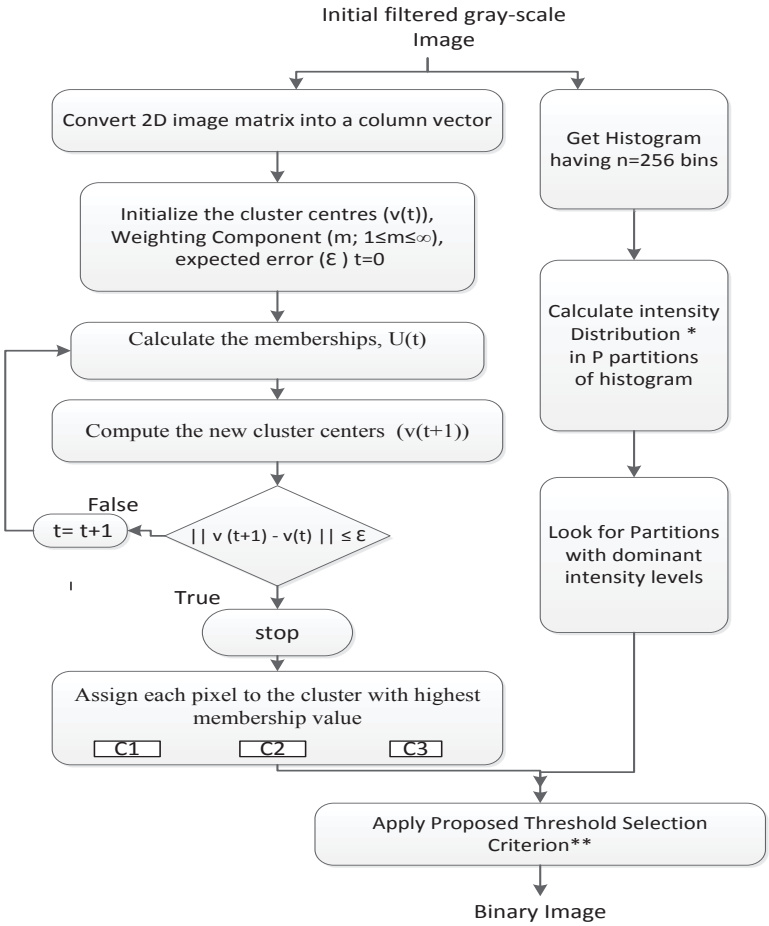


Figure 4.1 Histogram Analysis based Fuzzy C mean Thresholding

A. Fuzzy C mean Clustering

FCM clustering is used to partition N objects into C classes. The performance of FCM for medical image segmentation relies on the prefixed cluster number and the initial clusters substantially. It was found during experimental trials that random initialization was robust enough if 3-class clustering is used and it provides satisfactory results for appropriate segmentation of different skin lesion images including the one with blurred or highly abrupt and fuzzy boundaries. So, in this case N is equal to the number of pixels in the image i.e. $N=N_x$ x N_y and C=3 for 3-class FCM clustering. The FCM algorithm uses iterative optimization of an objective function based on a weighted similarity measure between the pixels in the image and each of the C-cluster centers. A local extremum of the objective function indicates an optimal clustering of the input data. The objective function that is minimized is given by (4.3)

$$Q = \sum_{i=1}^{C} \sum_{j=1}^{N} (u_{ij})^m \ \left\| z_j - v_i \right\|^2 \qquad (4.3)$$

where $z_j \in Z$ and $Z = \{z_1, z_2, z, \ldots \ldots z_N\}$ & $v_i \in V$ where $V = \{v_1, v_2, \ldots v_C\}$

$\|*\|$ is a norm expressing the similarity between any measured data value and the cluster centre; m∈ [1, ∞] is a weighting exponent and can be any real number greater than 1.

Calculations suggest that best choice of m is in the interval [1.5, 2.5], so m=2 is used here as it is widely accepted as a good choice of fuzzification parameter [401]. The fuzzy C-partition of given data set is the fuzzy partition matrix U= $[u_{ij}]$ with i=1, 2….C and j=1, 2, 3…N, where u_{ij} indicate the degree of membership of jth pixel to ith cluster. The membership functions are subject to satisfy the following conditions.

$$\sum_{i=1}^{C} u_{ij} = 1 \text{ for j=1,2,3,….N ; } 0 < \sum_{j=1}^{N} u_{ij} < N \text{ for i= 1,2….C; } 0 \leq u_{ij} \leq 1 \quad (4.4)$$

The aim of FCM algorithm is to find an optimal fuzzy C-partition by evolving the fuzzy partition matrix U= $[u_{ij}]$ iteratively and computing the cluster centres. In order to achieve this, the algorithm tries to minimize the objective function Q (4.3) by iteratively updating the cluster centres and the membership functions using (4.5,4.6).

$$v_i = \left(\sum_{j=1}^{N}(u_{ij})^m\right) / \left(z_j \sum_{j=1}^{N}(u_{ij})^m\right) \quad (4.5)$$

$$u_{ij} = 1/\left(\sum_{k=1}^{C}\left(\frac{\|v_i - u_j\|}{\|v_i - u_k\|}\right)^{\frac{2}{m-1}}\right) \quad (4.6)$$

After performing FCM clustering, finally each pixel is assigned to the cluster for which its membership value is maximum.

B. Selection of Threshold

Often the interest of image analysis methodologies restricts to the extraction of the object from the background so as to characterize the object with a set of features. Although a fuzzy threshold description is sufficient to some extent for this purpose, conventional feature extraction and object recognition methods may not be applicable as such with this description. Thus, in spite of the presence of elegant image analysis techniques developed based on fuzzy clustering description, hardening schemes are required to make the description useful for the conventional object recognition schemes. Details of few hardening approaches are presented by Jawahar et.al. [400]. Most of the techniques are limited to only 2-class fuzzy partitioning, e.g. (1) Finding a crisp threshold between two cluster centres, i.e. $U_1 < T_f < U_2$ and $\mu_o(T_f) = \mu_B(T_f)$. This provides mutually

exclusive and exhaustive object and background regions. (2) Hardening based on α – cuts of fuzzy sets, such as O= Oα = {(m,n); μ_o (I_{mn}) $\geq \alpha$ } and B= G-O. The parameter $\alpha \in (0,1]$ directly controls the size of the object region. As α increases, O approaches the core/skeleton of the object region. Not much attention is given for finding hardening schemes for general fuzzy segmentation, where the number of classes is more than two.

It was found that the practical significance of fuzzy clustering for multidimensional feature space is of great importance when it comes to skin cancer diagnosis. In this case deterministic misclassification can be very costly and a sophisticated thresholding procedure is required for increasing the accuracy of segmentation results. Apart from skin cancer diagnosis, formulation can provide very useful information for the high level vision. The proposed algorithm for threshold selection is as follows:

Algorithm:

- Transform the 2-D data containing intensity value for each pixel into column vector. Divide the pixels into C clusters based on the intensity value using above mentioned Fuzzy c-mean Clustering algorithm. (C=3 used for 3-class FCM clustering). Assign cluster label (1, 2 or 3) to each pixel data value.

- Get the histogram having n bins (n=256 for grayscale image), calculate for the intensity distribution in P partitions of histogram with P=C, using following formulation.

$\text{Distribution}_{p1,2} = \sum_{i=n/(C+1)}^{n/(\frac{C}{2})} p\_num(i,1)$, $\text{Distribution}_{p2,3} = \sum_{i=n/(\frac{C}{2})}^{n/(\frac{C}{3})} p\_num(i,1)$ …

$$\text{Distribution}_{p1,P} = \sum_{i=n/(C+1)}^{nC/(C+0.5)} p\_num(i,1) \quad (4.7)$$

- Where i is the value of intensity bin and p_num is the pixel numbers for that intensity value. For skin images, the case of three clusters is used and consequently three histogram partitions are obtained, as it gives best results for skin lesion segmentation. However, the mathematical formulations can be easily extended for n number of clusters for general segmentation problems.

- Look for the clusters with dominant intensity value, to identify location of appropriate threshold, using the following relations.

If ($Distribution_{p1,2}$ > $Distribution_{p2,3}$ && $Distribution_{p1,2}$ > $Distribution_{p1,3}$ ..)

Threshold value (T) = (max (data (label=1)) +min (data (label=2))/2

If ($Distribution_{p2,3}$ > $Distribution_{p1,2}$ && $Distribution_{p2,3}$ > $Distribution_{p1,3}$ ...)

Threshold value (T) = (max (data (label=2)) +min (data (label=3))/2

If ($Distribution_{p1,3}$ > $Distribution_{p2,3}$ && $Distribution_{p1,3}$ > $Distribution_{p2,3}$ ...)

Threshold value (T) = (max (data (label=1)) +min (data (label=3))/2

- Get the binary image on the basis of selected threshold value.

The major difference between the proposed thresholding algorithm and the popular ones is that it combines the advantages of iterative formulations of fuzzy C mean clustering with histogram analysis. Thus, it reduces the computational burden on the iterative procedure required for threshold selection in conventional FCM thresholding[400]. In addition to this it overcomes the need of extensive search for identification of appropriate threshold value as required in conventional histogram thresholding[397, 402]. It is evident (See Figures 4.2-4.4) that the proposed algorithm is efficient and provides superior segmentation results as compared to other thresholding algorithms reported in literature[397, 400, 403-410].

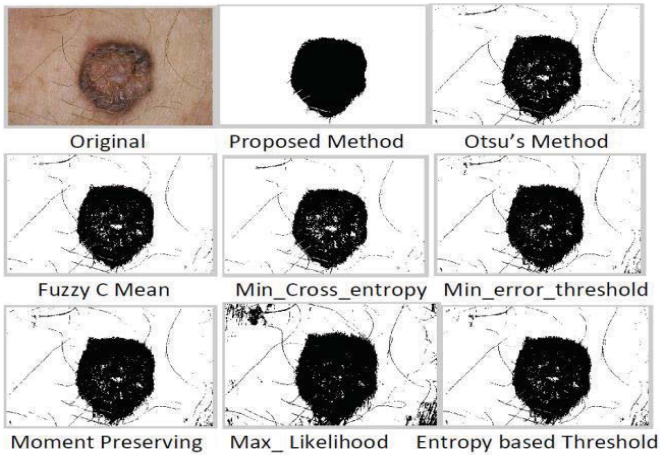


Figure 4.2 Thresholding of image with melanoma lesion with hair around

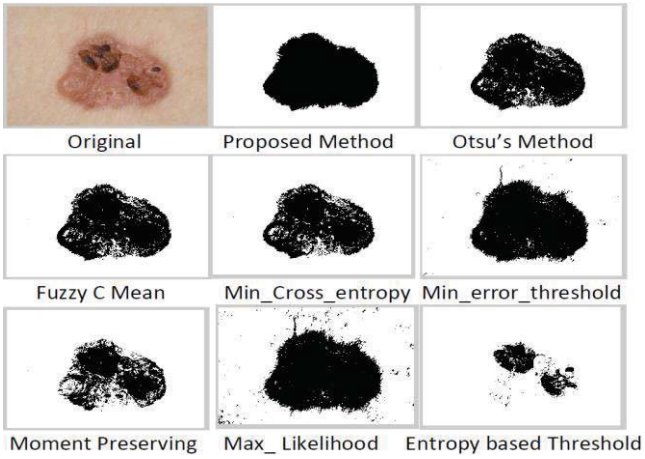


Figure 4.3 Thresholding of image with melanoma lesion with multiple colours

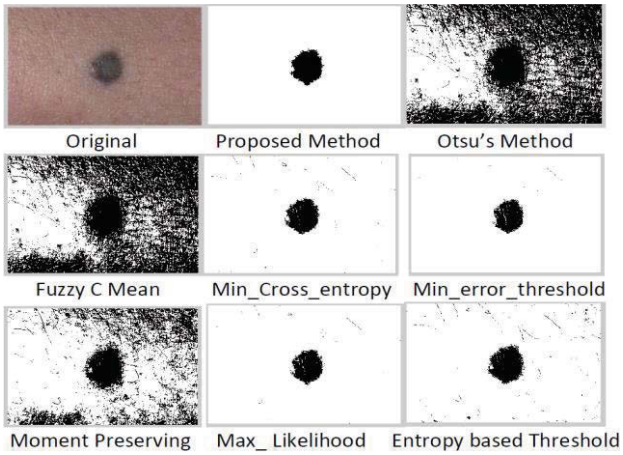


Figure 4.4 Thresholding of image with benign lesion

A good initial segmentation is essential for iterative sophisticated methods to achieve global convergence faster. These results can be used as a starting image for more sophisticated segmentation methods that utilize spatial information as well.

4.3.3. Morphological operations

The pixels on an image are highly correlated, i.e. the pixels in the immediate neighbourhood possess nearly the same feature data. Therefore, the spatial relationship of the neighbouring pixels is an important characteristic that can be of great aid in imaging segmentation. General boundary detection techniques have taken advantage of this spatial information for image segmentation. However, the FCM algorithm does not fully utilize this spatial information.

There are two major limitations associated with FCM based algorithms while using for image segmentation. Firstly, it lacks the spatial information and secondly, it pays more attention on intensity information only. Both of the issues can lead to increased effect of noise, artifacts and background inhomogeneity on the final image segmentation results. Li.et.al [395, 396] suggested two different ways of overcoming this limitation in case of liver tumor segmentation. One is the use of morphological operations and another one is by incorporating spatial information in fuzzy c-mean clustering.

It has been observed in experimental trials that for skin cancer detection using the proposed algorithm, interfacing thresholding results with morphological operations like erosion and dilation enhanced the overall accuracy of segmentation process. One reason for using morphological operations in place of incorporating spatial information in FCM clustering is the simplicity and computation ease of morphological operations [153]. Secondly, morphological operations like erosion and dilation have proved to enhance the skin lesion segmentation results especially in case of presence of hair which is a major artifact in case of skin lesions[398]. The major objective of morphological operations was to filter out the noise effects remained during FCM based thresholding and for preserving the genuine image components.

The output of this step is a binary image which will further be used for initialization and calculation of controlling parameter for level set evolution which is discussed further as follows.

4.3.4 H-FCM based fast Level Set Segmentation

Segmentation of images by means of active contours is a well-established approach. Active contours [411] are used to detect objects in a given image using techniques of curve evolution. The basic idea is to evolve a curve, subject to constraints from a given image, for detecting objects in that image. For instance, starting with a curve around the lesion to be segmented, the curve moves towards its interior normal and has to stop on the boundary of the lesion. Let ω be a bounded open subset of R^2, with $\partial\omega$ its boundary. Let $I_0: \bar{\omega} \rightarrow R$ be a given image, and C(s): [0, 1] $\rightarrow$ R^2 be a parameterized curve. Now in classical snake and active contour models, an edge detector (based on image gradient) is normally used to stop the evolving curve on the boundary of desired object. Instead of this parametric characterization of active contours, level set (LS)

methods can also be used in problems of curve evolution which embed active contours into a time dependent partial differential equation.

LS methods are established on dynamic implicit interfaces and partial differential equations (PDEs). In traditional LS formulation[391], the curve denoted by C is represented implicitly via a Lipschitz function $\emptyset$, by C= {(x,y)| $\emptyset$ (x,y)=0) , while the evolution of curve is given by the zero level curve at time t of the function $\emptyset$ (t,x,y). The evolving equation of the level set function $\emptyset$ can be written in the following general form (4.8)

$$\frac{\partial \emptyset}{\partial t} + F|\nabla \emptyset| = 0,\ \emptyset(0, x, y) = \emptyset_0(x, y) \tag{4.8}$$

which is called levels set equation, where $F = \text{div}\,(\nabla \emptyset(x, y)/|\nabla \emptyset(x, y)|$ is the speed function that represents the comprehensive forces which includes the internal force from the interface geometry and the external force from image gradient or/and artificial momentums [163].

In order to stop the level set evolution near the optimal solution, the advancing force need to be regularized by an edge indication function g. The edge indication function used here is given by (4.9) where I^* stands for the filtered image.

$$g = 1/(1 + |\nabla I^*|^2) \tag{4.9}$$

A geometric active contour model based on level set method and in particular the motion of mean curvature is given by the following equation [411].

$$\frac{\partial \emptyset}{\partial t} = g|\nabla \emptyset| \left(div \left(\frac{\nabla \emptyset}{|\nabla \emptyset|} \right) + \upsilon \right) \tag{4.10}$$

$$\emptyset(0, x, y) = \emptyset_0(x, y)\ in\ R^2 \tag{4.11}$$

Where $div \left(\frac{\nabla \emptyset}{|\nabla \emptyset|} \right)$ approximates the mean curvature and υ is customable balloon force for pushing curve towards the object.

The traditional LS method is computationally intensive and has certain limitations like need of re-initialization of level set function to signed distance function for stable curve evolution [391]. Therefore, in this algorithm fast level set formulation proposed by Li. et al. [392] is used. It is observed that this method is computationally more efficient and can be implemented by using simple finite difference scheme. In order to segment

the skin lesion image the overall iterative process for levels set evolution is given by (4.12).

$$\emptyset^{j+1} = \emptyset^{j} + \tau[\xi(g, \emptyset^{j}) + \mu\xi(\emptyset^{j})] \tag{4.12}$$

where $\xi(g, \emptyset) = \lambda\delta_{\varepsilon}(\emptyset)div\left(g\frac{\nabla\emptyset}{|\nabla\emptyset|}\right) + g\upsilon\delta_{\varepsilon}(\emptyset)$ is the term for attracting $\emptyset$ towards the variational boundary and $\xi(\emptyset) = \left(\nabla^2\emptyset - \frac{\nabla\emptyset}{|\nabla\emptyset|}\right)$ is the penalty term which forces $\emptyset$ to automatically approach the genuine signed distance function.

This fast level set formulation proposed in [392] provides a benefit of flexible initialization where region from thresholding can be used to construct the initial level set function. Taking advantage of this facility, binary image (Bi) obtained from the proposed thresholding algorithm (H-FCM), discussed in the previous section is used here for automatic initialization of the level set function $\emptyset$. The initial level set function (4.13) is used as in [395].

$$\emptyset_0(x, y) = -4\varepsilon(0.5 - Bi) \tag{4.13}$$

where ε is the regulator for dirac function [391] defined as follows:

$$\delta_{\varepsilon}(x) = \begin{cases} 0, & |x| > \varepsilon \\ \frac{1}{2\varepsilon}\left[1 + \cos\left(\frac{\pi x}{\varepsilon}\right)\right], & |x| \leq \varepsilon \end{cases} \tag{4.14}$$

Controlling parameters are also adaptively determined from the thresholded image (Bi) for regularizing the LS evolution process automatically. The weighting coefficient μ of penalty term $\xi(\emptyset)$ is taken as the ratio of area of on pixels in the binary image (Bi) to its perimeter pixels. The time step τ is taken as $0.2/\mu$ so that ($\tau x \mu$) remains always smaller than 0.25 which is necessary to ensure stable evolution as analysed in [392]. λ is the coefficient of contour length for smoothness regulation and its value is taken here as $0.1/\mu$, which is the best one found through experimental analysis. The value of λ can be increased for accelerating the evolution process but it may lead to over-smoothened contours. Thus, care must be taken especially for skin lesion images, where over smoothened images may lose significant details about boundary of lesion, which is an important feature for accurate cancer diagnosis. The balloon force which determines the advancing direction and speed of the evolving curve is given as (4.15)

$$\upsilon = -2(2 * \beta * Bi - (1 - \beta)) \tag{4.15}$$

where β is the modulating argument, taken here as 0.5, (found experimentally as best suited for skin cancer segmentation).

Figure 4.5 shows a systematic diagram that concludes the overall proposed segmentation method. Each block represents different steps of the algorithm.

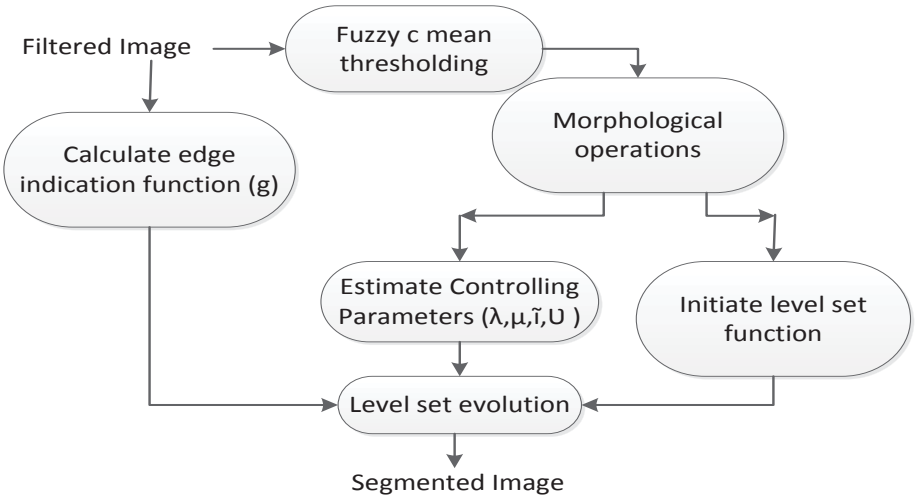


Figure 4.5 Flowchart of proposed algorithm

4.4 Segmentation of histopathological images for skin cancer

Traditionally, in the skin cancer diagnosis process, pathologists use histopathological images of biopsy samples removed from patients and examine them under a microscope. A pathologist typically examines the image to observe the deviations in the cell structures and/or the change in the distribution of the cells across the tissue under examination. However, these judgments depend on their personal experience and expertise and often lead to considerable variability [10, 11]. To overcome this problem and improve the reliability of diagnosis process, it is important to develop computational tools for automated diagnosis that operate on quantitative measures. Such tools can facilitate objective mathematical judgment complementary to that of pathologists, and help them in identifying the affected areas efficiently.

Detecting areas of interest and quantifying them plays an increasingly important role in modern medical diagnosis and treatment planning. Thus, image segmentation is an important measure toward the analysis phase in many medical image-processing tasks such as quantitative analysis and identification of the lesion tissues, image guidance operation, tumors radiotherapy and evaluation of therapies. Automatic grading of

pathological images has been investigated in various fields during the past few years, including brain tumor, breast cancer and oral sub-mucous fibrosis detection [412-414]. However, histopathological image analysis of skin has not been exhaustively reported in the literature, except few [415] [416-418]. This is partially due to the fact that automated segmentation of cancerous area of histopathological images is not a straightforward task because of the complex nature of image scenes and partly because the option of dermoscopic analysis for skin images is relatively easier than histopathological image analysis of underlying tissues.

There are a number of challenges faced during the segmentation process of histopathological skin images. The first challenge is the elimination of noise introduced by the staining process which is reflected in the lack of dark separation lines between a nucleus and its surrounding and can also cause non-nuclei stain artifacts. Secondly, typical tissues comprise touching and overlapping cells that demand extra effort in the detection of cancerous area. Thirdly, inhomogeneous interior of a nucleus adds to the difficulty of the overall process.

Numerous techniques are present in literature for tackling the problem of efficient and accurate image segmentation for separating an image into meaningful distinctive segments considering color, texture or spatial features of an image [69, 143, 162, 419, 420]. These include various thresholding, clustering, region growing and edge detecting methods. Although these methods have not been specifically designed for histopathological skin images, they can, however, provide the basis for segmenting histopathological images. The method proposed here is an extension of the method presented in the previous section, based on the concepts of orientation sensitive fuzzy C-means clustering and level set evolution to segment the cancerous area in the histopathological images of skin. The method is tested on a dataset of 150 images and segmentation results are compared with some state of the art methods used in literature for the segmentation of histopathological images. The details of the proposed method are as follows.

4.4.1 Pre-Processing

The main purpose of the pre-processing stage is to reduce the amount of noise that may badly affect the segmentation process. Considerable amount of noise arises from the staining process, which makes the focal area identification difficult, especially, at the

cellular level. The noise can cause the dark separation lines between a nucleus and it's surrounding to reduce (fade out). The staining process may also result in non-nuclei stain artifacts in the tissue. Thus, introducing more errors to the image. In order to ensure a successful segmentation process, the image should be pre-processed to account for the noise and the other bad consequences of the staining process. The success of segmentation is highly dependent on sufficient pre-processing.

The pre-processing process used is described in Figure 4.6. Initially, the RGB image is converted to Lab colour space (L is luminance component and a, b are chrominance channels). The Luminance component (L channel) is subjected to low pass filtering using 5 x 5 median filters. This helps in the correction of gray level shading and reducing the noise. Median filter replaces the current pixel under consideration by the median of its pixel neighborhood, including the current pixel itself. It is noted that selection of suitable shape and size of neighborhood plays an important in the performance of median filter depending upon the complexity of images. If the neighborhood is too large the median filter may smoothen the edges too much; on the other hand if it is too small it will be unable to sufficiently remove the noise. For present case, it is found that a median filter with square shape of 5 x 5 performed very well for most of the images.

After median filtering the intensity values are mapped from 0 to 255 for contrast enhancement. The L-channel is then combined with the chrominance channels (a-Channel and b-channel) thus converting the image back to RGB colour space. Next, the RGB image is converted to the gray scale image by forming a weighted sum of R, G and B components. Afterwards, anisotropic diffusion [421] is used for smoothing the homogenous parts of the images while sharpening the edges and enhancing important details. The approach is based on embedding the original image I in a family of derived images that are obtained by convolution of original image with a Gaussian kernel having variance t.

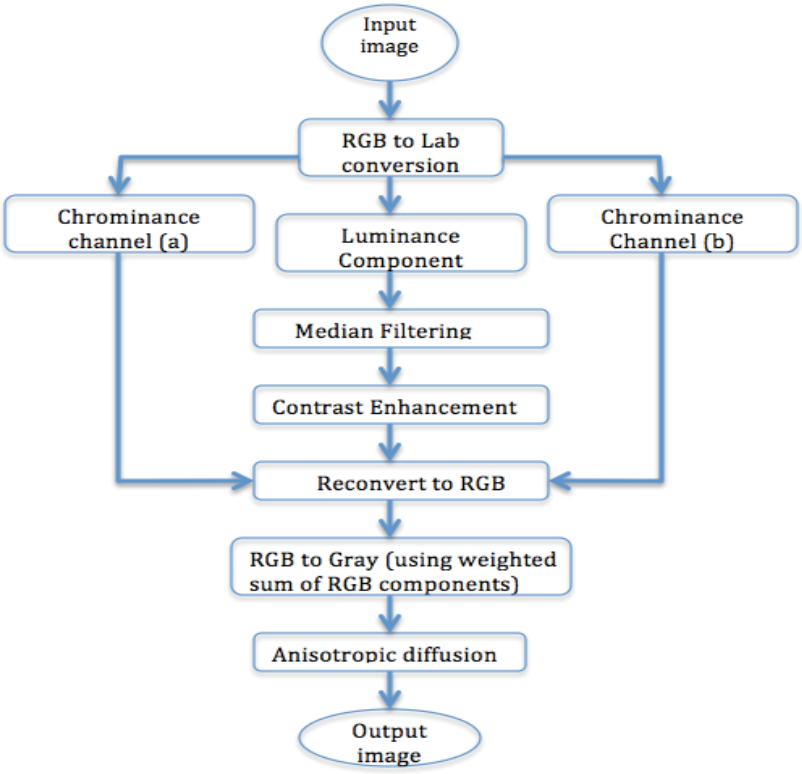


Figure 4.6 Pre-processing stage for Histopathalogical images

Coarser resolution results for larger values of t while lower values of t correspond to fine resolution. This can be represented by the solution of the anisotropic diffusion equation given in (4.16).

$$\frac{\partial \mathbf{I}}{\partial \mathbf{t}} = \mathbf{div}(\mathbf{c}(\mathbf{x},\mathbf{y},\mathbf{t})\nabla\mathbf{I}) = \nabla\mathbf{c}.\nabla\mathbf{I} + \mathbf{c}(\mathbf{x},\mathbf{y},\mathbf{t})\Delta\mathbf{I} \tag{4.16}$$

Where Δ denotes the Laplacian operator, ∇ represents the gradient and c(x,y,t) is the diffusion coefficient which controls the rate of diffusion. In this work, 8 nearest neighbour discretization of the Laplacian operator is used. Perona and Malik [393] proposed two functions for diffusion coefficients as the solution to the diffusion equation given in (4.17) and (4.18) below.

$$\mathbf{c}(\|\nabla\mathbf{I}\|) = \mathbf{e}^{(\frac{\|\nabla\mathbf{I}\|}{\mathbf{K}})^2} \tag{4.17}$$

$$\mathbf{c}(\|\nabla\mathbf{I}\|) = \frac{\mathbf{1}}{\mathbf{1}+(\frac{\|\nabla\mathbf{I}\|}{\mathbf{K}})^2} \tag{4.18}$$

Where the constant K controls the sensitivity to the edges. The first function gives more privileges to high contrast edges over low contrast edges, while the second gives more privileges to wider regions over the smaller ones.

4.4.2 Segmentation

The basic segmentation algorithm developed for digital/dermoscopic images was described in the flowchart in Figure 4.5. The same concept is extended for developing the algorithm for segmentation of the histopathological images. The modification is made to get efficient segmentation for histopathological images that have different characteristics as compared to dermoscopic images. Initially, Orientation Sensitive Fuzzy C Means clustering is applied on the pixels of the pre-processed image to classify them into clusters. It results in a rough image for the area under consideration. The rough image is then used as an initial level set function for the Level Set Evolution method and to estimate its controlling parameters. The level set evolution stage is identical to the one discussed in section 4.3.4 however the difference lies in the process of obtaining the course image used for level set initialization. The overall segmentation algorithm proposed for histopathalogical images is shown in Figure 4.7 and the details of the orientation Sensitive FCM clustering phase are presented below the output of which is used for the initialization of the level set evolution process.

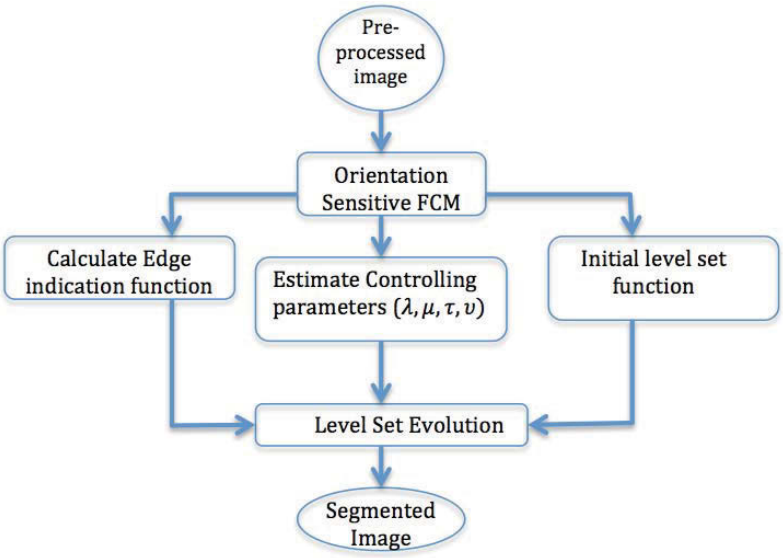


Figure 4.7 Proposed segmentation method for histopathological images

- **4.4.2.1 Orientation Sensitive Fuzzy C mean Clustering**

The aim of Fuzzy C mean Clustering (FCM) algorithm is to find an optimal fuzzy c-partition of the image by iteratively evolving the fuzzy partition matrix U= $[\mathbf{u_{ij}}]$ and computing the cluster centers. The fuzzy c-partition of given data set is the fuzzy partition

matrix U= [$\mathbf{u_{ij}}$] with i=1, 2….C and j=1, 2, 3…N, where $\mathbf{u_{ij}}$ denotes the membership value of ith pixel to jth cluster. The membership functions are subject to satisfy the following conditions (4.19)

$$\sum_{\mathbf{j=1}}^{\mathbf{C}} \mathbf{u_{ij}} = \mathbf{1} \forall \mathbf{i} \; ; \; \mathbf{0} < \sum_{\mathbf{i=1}}^{\mathbf{N}} \mathbf{u_{ij}} < \mathbf{N} \; \forall \mathbf{i}; \; \mathbf{0} \leq \mathbf{u_{ij}} \leq \mathbf{1} \; \forall \; \mathbf{j}, \mathbf{i} \tag{4.19}$$

Given the number of clusters, the following two recurrent equations (4.20) (4.21) are used.

$$\mathbf{U_{ij}} \; = \; (\mathbf{1} + \sum_{\substack{\mathbf{k=1} \\ \mathbf{k \neq j}}}^{\mathbf{C}} (\frac{\left\| \mathbf{x_i} - \boldsymbol{\mu}_\mathbf{j} \right\|^2}{\left\| \mathbf{x_j} - \boldsymbol{\mu}_\mathbf{j} \right\|^2})^{\mathbf{1}/\mathbf{m-1}})^{-\mathbf{1}} \tag{4.20}$$

$$\boldsymbol{\mu}_\mathbf{j} = \frac{\sum_{\mathbf{i=1}}^{\mathbf{N}} \boldsymbol{\mu}_{\mathbf{ij}}{}^{\mathbf{m}} . \mathbf{x_i}}{\sum_{\mathbf{i=1}}^{\mathbf{N}} \boldsymbol{\mu}_{\mathbf{ij}}{}^{\mathbf{m}}} \tag{4.21}$$

where $\mathbf{U_{ij}}$ is the fuzzy membership of xi to class j, $\boldsymbol{\mu}_\mathbf{j}$ is the jth class centre, C is number of classes taken here as 3, m is a weighting exponent that defines the fuzziness of the membership values and can be any real number greater than 1. Calculations suggest that the best choice of m lies in the interval [1.5, 2.5], so m=2 is used here as it is normally accepted as a good choice of fuzzification parameter.

FCM clustering has the ability to compute the cluster centres precisely. However, for the case of segmentation of the histopathological images the distance between clusters depends not only on the cluster centers, but also on their orientation, thus the fuzzy membership matrix is updated using (4.22) and the FCM algorithm becomes sensitive for the orientation. This modified FCM algorithm is referred here as Orientation Sensitive Fuzzy C Means (OS-FCM).

$$\mathbf{U_{ij}} \; = \; (\mathbf{1} + \sum_{\substack{\mathbf{k=1} \\ \mathbf{k \neq j}}}^{\mathbf{C}} (\frac{(\mathbf{x_i} - \boldsymbol{\mu}_\mathbf{j})^{\mathbf{T}} \mathbf{A_j} (\mathbf{x_i} - \boldsymbol{\mu}_\mathbf{j})}{(\mathbf{x_i} - \boldsymbol{\mu}_\mathbf{k})^{\mathbf{T}} \mathbf{A_j} (\mathbf{x_i} - \boldsymbol{\mu}_\mathbf{k})})^{\mathbf{1}/\mathbf{m-1}})^{-\mathbf{1}} \tag{4.22}$$

where the matrix $\mathbf{A_j}$ which is used for computing the distance taking into account the local orientation. $\mathbf{A_j} = \mathbf{V_j}^{\mathbf{T}} \mathbf{L_j} \mathbf{V_j}$ where $\mathbf{L_j}$ is a diagonal matrix containing the inverse of the eigenvalues of the co-variance matrix $\mathbf{C_j}$ given in (4.23) and the matrix $\mathbf{V_j}$ is composed of the corresponding eigenvectors [422]. The co-variance matrix $\mathbf{C_j}$ for each cluster j in the above equation is defined as

$$\mathbf{C_j} = \frac{\sum_{\mathbf{i=1}}^{\mathbf{N}} \mathbf{u_{ij}}^{\mathbf{m}} (\mathbf{x_i} \mathbf{x_j}^{\mathbf{T}} - \boldsymbol{\mu}_\mathbf{i} \boldsymbol{\mu}_\mathbf{j}^{\mathbf{T}})}{\sum_{\mathbf{i=1}}^{\mathbf{N}} \mathbf{u_{ij}}^{\mathbf{m}}} \tag{4.23}$$

The overall OS-FCM algorithm presented above works on iterative optimization of an objective function based on a weighted similarity measure between the pixels in the image and each of the c-cluster centers. A local extreme of the objective function indicates the optimal clustering of the input data. The objective function that is minimized is given by (4.24)

$$\mathbf{Q} = \sum_{\mathbf{i=1}}^{\mathbf{C}} \sum_{\mathbf{j=1}}^{\mathbf{N}} (\mathbf{u_{ij}})^{\mathbf{m}} \left\| \mathbf{x_j} - \boldsymbol{\mu}_\mathbf{i} \right\|^2 \tag{4.24}$$

In order to achieve this, the algorithm tries to minimize the objective function Q by iteratively updating the cluster centers and the membership functions. Once the optimization is obtained, the pixels in the pre-processed image are labeled in order to generate the initial coarse segmented image. The results of fuzzy clustering are used to regularize the level set evolution algorithm as presented in section 4.3.4. It will result in a refined segmentation of the cancerous regions present in the image.

In the following section the experimental analysis of the developed segmentation algorithms is provided for dermoscopic and histopathological images respectively.

4.5 Experimental Results for Dermoscopic Images

4.5.1 Image Dataset

The database used for analysis comprised of colour dermoscopic and clinical view lesion images, including images of melanocytic lesions, dysplastic nevi (atypical moles) and malignant melanoma lesions, which were collected from various sources but most images were obtained from Sydney Melanoma Diagnostic Centre, Royal Prince Alfred Hospital. Most of these skin cancer images have been captured from digital camera. The images have been stored in the 24-bit RGB format with dimensions varying from 720x472 to 720x484. There are different artifacts such as dermoscopy gel, specular reflection and outline markers, skin lines, hair etc. that can decrease the accuracy of border detection. Some of the difficulties are illustrated in Figure 4.8. Thus, in order to show significance of the proposed method for tumor delineation, it was ensured to select images with diverse skin types and textures with these different artifacts. It was assured that the proposed method could provide reliable segmentation results even in the presence of these unavoidable artifacts commonly present in skin lesion images.

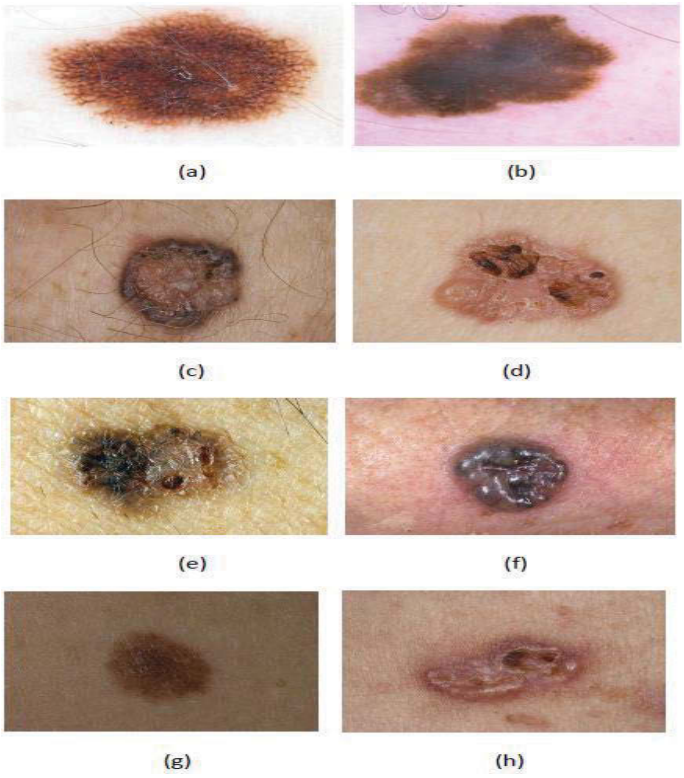


Figure 4.8 Difficulties in skin lesion images; (a) uneven border (b) presence of dermoscopic gel (c) presence of hair (d) multiple colours (e) skin lines and multiple colours (f) specular reflections (g) ill-illumination (h) smooth transaction between lesion and skin

In order to find out the efficiency of our method, the overall algorithm is implemented in MATLAB. Segmentation results obtained using the proposed algorithm is presented in Figure 4.9 - 4.17 for some of the skin lesion images selected from the dataset. It is clearly evident from the visual analysis of these results that the proposed segmentation method performs well in the presence of almost all possible difficulties that are expected to be present in a skin lesion image.

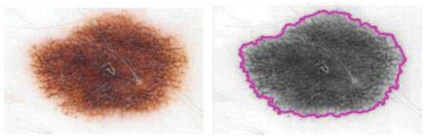


Figure 4.9 Segmentation of images with uneven and abrupt lesion border

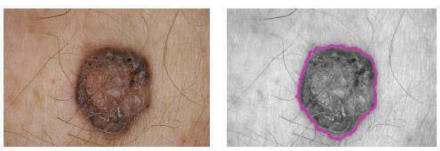


Figure 4.10 Segmentation of skin lesion surrounded by hair

Figure 4.11 Segmentation of lesion with multiple colours

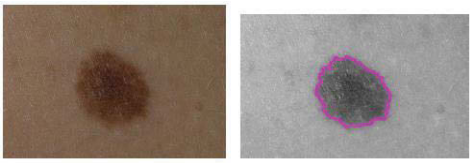


Figure 4.12 Segmentation of lesions with poor illumination

Figure 4.13 Segmentation of image having specular reflection

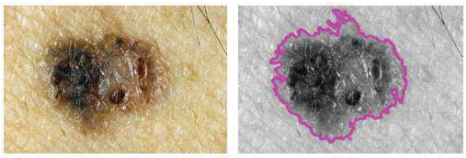


Figure 4.14 Segmentation of lesion with prominent skin lines

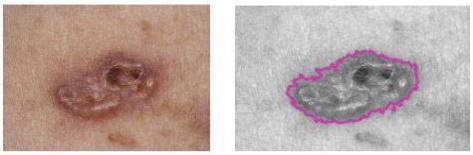


Figure 4.15 Segmentation of image with smooth transaction between lesion and skin

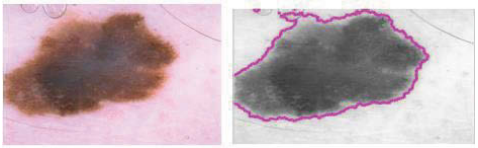


Figure 4.16 Segmentation of image with bubbles of dermoscopic gel

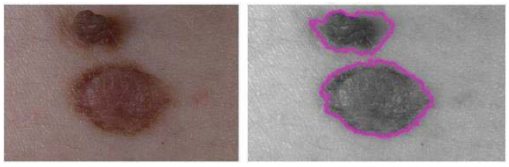


Figure 4.17 Segmentation of image with multiple lesions

4.5.2 Statistical Evaluation Parameters

In order to establish validity and clinical applicability of an algorithm, the objective evaluation of segmentation algorithms on a large set of clinical data is one of the important steps. Objective validation of medical image segmentation algorithms is somehow an open question, especially if one considers the variance of image characteristics and the related clinical requirements for accuracy. For evaluating the efficiency of the proposed method a performance comparison is provided with some of the well-known segmentation methods used here for segmentation of same collection of skin lesion images. The metric used here for measurements is based on pixel-by-pixel comparison of pixels enclosed in the segmented result (SR) and the ground truth image (GT). First, binary images are constructed for each boundary, where a pixel is considered non-zero if it lies inside the boundary and zero otherwise. The evaluation metrics are calculated as follows:

Hammoude Distance (HM): This metric makes a pixel-by-pixel comparison of the pixels enclosed by the two boundaries:

$$\mathrm{HM}(\mathrm{SR}, \mathrm{GT}) = \frac{\#(\mathrm{SR} \cup \mathrm{GT}) - \#(\mathrm{SR} \cap \mathrm{GT})}{\#(\mathrm{SR} \cup \mathrm{GT})} \tag{4.25}$$

It takes into account two types of error; pixels classified as lesion by automatic segmentation that were not classified as such by medical expert and pixels classified as lesion by medical expert that were not classified as such by automatic segmentation. The Hammoude distance gives equal importance to both types of errors.

However, from a clinical point of view, the 2nd type of error is more important since the lesion pixels should never be missed by the automatic diagnostic system. On the other hand, the experts demand that computer aided diagnostic system should help in reducing the rate of unwanted excision. So, the automated diagnostic system should not overestimate benign or dysplastic nevi as melanoma. Therefore, separate metrics should be used to take into account the two types of error. True positive rate, false positive error and false negative error are the parameters that can help in determining the efficiency of the segmentation method from the prospective of both types of errors.

True Positive Rate (TPR): This metric measure the rate of pixels classified as lesion by both the automatic and the medical expert segmentation. Higher TDR shows better performance of segmentation method.

$$TDR(SR, GT) = \frac{\#(\text{SR} \cap GT)}{\#(GT)} \tag{4.26}$$

False Positive Error (FPE): This metric determines the rate of pixels assigned as lesions by the segmentation method that were not assigned as lesion by the medical expert. Lower the value of FPE better is the performance of respective segmentation method.

$$\text{FPE}(\text{SR}, \text{GT}) = \frac{\#(\text{SR} \cap \overline{\text{GT}})}{\#(\text{GT})} \tag{4.27}$$

False Negative Error (FNE): This metric determines the rate of pixels categorized as lesions by the medical expert that were not assigned as lesion by the automatic segmentation. Again lower value of FNE shows that segmentation method is performing well.

$$\text{FNR}(\text{SR}, \text{GT}) = 1 - \frac{\#(\text{SR} \cap \text{GT})}{\#(\text{GT})} \tag{4.28}$$

Dice Similarity Coefficient (DSC): Dice similarity coefficient measures agreement between the ground truth and result of automated segmentation method. It is given by the formula

$$\text{Dice Similarity Coefficient} = \frac{2\ \text{TP}}{((\text{FP}+\text{TP})+(\text{TP}+\text{FN}))} \tag{4.29}$$

A value of 0 indicates no overlap; a value of 1 indicates perfect agreement. Higher number close to 1 indicates better agreement, and in the case of segmentation it indicates that the results match the gold standard (ground truth) better than results that produce lower Dice coefficients.

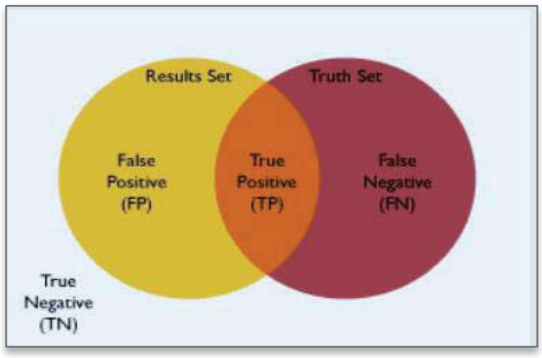


Figure. 4.18 Diagrammatical representation of evaluation parameters

Border Error: The automatic border can be compared with the manual ground truth border (marked by expert dermatologist) quantitatively by calculating border error value as follows

$$\text{Border Error} = \frac{\text{Area (SR)} \cup \text{Area (GT)} - \text{Area (SR)} \cap \text{Area (GT)}}{\text{Area (GT)}} \quad (4.30)$$

Where Area (SR) represents the area inside the automatic border and Area (GT) represents the area inside the manual border.

4.5.3 Comparative Analysis of segmentation results

The objective evaluation of segmentation algorithms on a large set of clinical data is one of the most important steps towards establishing validity and clinical applicability of the algorithm. Therefore, in order to evaluate the efficiency of the proposed method a performance comparison is provided with some of the well-known segmentation methods present in literature, that include, FCM clustering [157], k-mean clustering[397], region based active contours [139] and adaptive thresholding [162]. To ensure proper comparison, all these segmentation methods are applied on the same collection of skin lesion images. The algorithms were implemented in MATLAB R2012a (7.14.0.739) ®. All computations were performed on a 3.10GHz core (TM) i5 64-bit Intel processor with 4GB DDR2 RAM, running Windows 7 edition.

For the clarity of comparison and for supporting the claim about the proposed method, an experimental example of two of the results is presented here. The image considered in Figure 4.19 contain difficulties like smooth transaction between lesion and skin, presence of multiple colours, skin with prominent lines and patches and little bit of specular reflection as well. Although, the image in Figure 4.20 is heavily surrounded by hair, one can examine that the proposed algorithm extract the lesion border quite accurately, while other methods didn't succeed well in providing that accurate estimate of boundary when applied on the image with same degree of complexity.

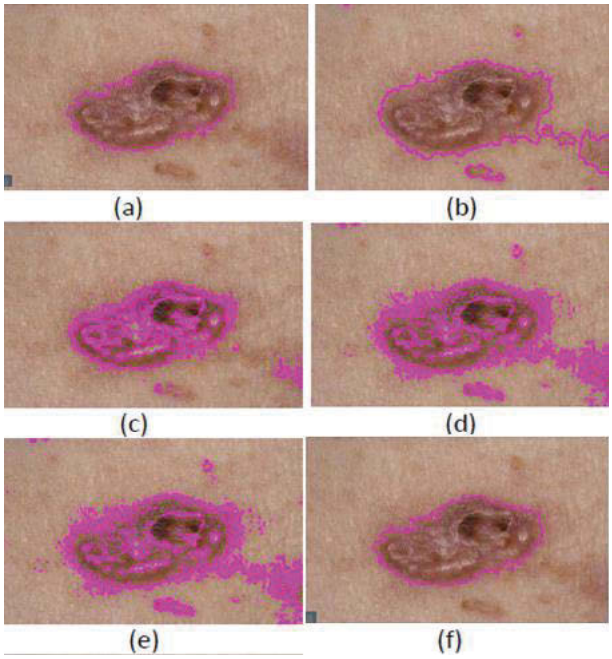


Figure 4.19 Border detection results obtained with (a) Ground truth by expert (b) Region based active contour (c) Adaptive thresholding (d) Fuzzy c-mean clustering (e) K-mean clustering (f) Proposed algorithm

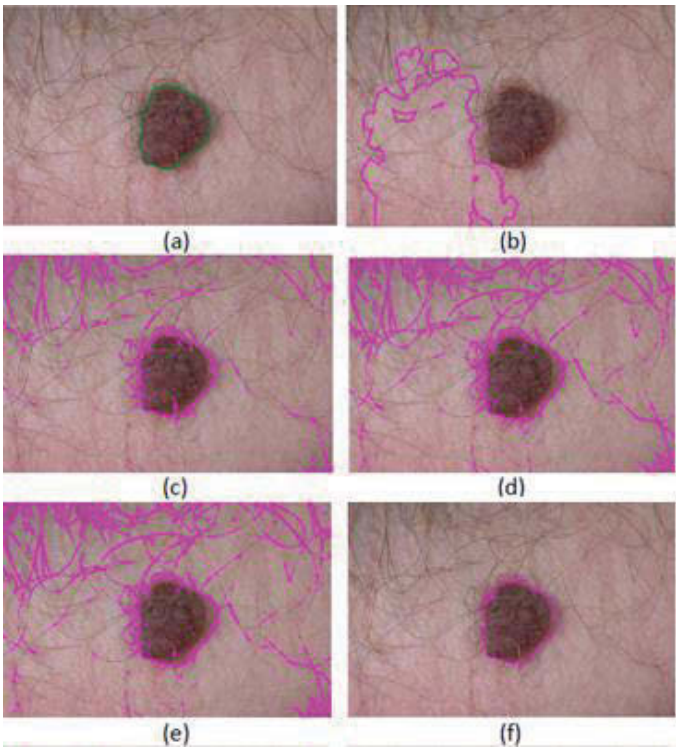


Figure 4.20 Border detection results obtained with (a) Ground truth by expert (b) Region based active contour (c) Adaptive thresholding (d) Fuzzy c-mean clustering (e) K-mean clustering (f) Proposed algorithm

Table 4.1 shows the comparative results of the proposed method with four of the well-known methods frequently being used in literature for segmentation of dermoscopic images. The evaluation was based on the measures described above i.e. HM, TDR, FPE and FNE. In order to evaluate these measurement parameters, the 4 segmentation methods and the proposed algorithm were tested on a dataset of 238 skin lesion images that include (1) 28 benign melanocytic lesions, (2) 67 dysplastic nevi (atypical moles), and (3) 143 malignant melanoma lesions. The segmentation results were compared with the reference images (ground truth) and average of segmentation scores are presented here for each method.

Table 4.1 Comparison results of segmentation methods. Value in bold correspond to best performance

Method	HM (%)	TDR (%)	FPE (%)	FNE (%)
FCM Clustering	27.84	85.25	8.52	14.75
K mean Clustering	38.15	74.83	13.14	25.16
Region based Active Contours	27.04	88.28	9.08	11.72
Adaptive Thresholding	23.23	86.86	6.95	13.14
Proposed FCM thresholding based fast LS Segmentation	**11.09**	**93.16**	**4.16**	**6.84**

It is evident from the results that the proposed method has shown reasonably better performance as compared to other methods. Fuzzy C mean Clustering and K mean clustering have not shown very good performance. Adaptive thresholding has shown relatively good performance in terms of FPE and border error but stills its average false negative error is much higher than the proposed method which makes it susceptible of declaring melanoma as benign, which can result in severe consequences as a potential cancer case can be diagnosed as benign. Region based active contour method showed comparatively lower false negative error but its false positive error is high, which is also not widely acceptable as it can result in excision when it is not needed. Its performance

can be improved by increasing the number of iterations for tracking the exact border but then the method becomes poor in respect of time efficiency. On the other hand, increase in number of iterations does not guarantee the convergence of active contour to the exact border if the lesion is surrounded by hairs (Figure 4.20 (b) result shown even after 25000 iterations). On the basis of both visual inspection as well as numerical comparison, it is evident that the proposed method provided significantly improved performance for the segmentation of skin lesion images, which consequently lead to better diagnostic results.

As one of the arguments was that the proposed method is a two-phase process, with coarse segmentation followed by fine border detection. So in order to further validate the comparative significance of the proposed method, various different techniques for level set initialization are also used and followed the same two phase approach for other methods as well. The algorithms used for comparative analysis are standard LS proposed by Chan (C-LS) [423], spatial fuzzy clustering based LS (SF-LS) [395], Adaptive Thresholding based LS (AT-LS), and K mean clustering based LS (K-LS) iterative thresholding based level set [139] and some of the results are presented in Figure 4.21 – 4.26 for some of the skin lesion images.

While presenting the results, it is tried to present images having different common problems of dermoscopic images, which can badly affect the segmentation process. Figure 4.21 represents segmentation of melanoma lesion with uneven border and hair on it. Figure 4.22 is image of dysplastic nevi with smooth transaction between lesion and skin. Figure 4.23 is a melanoma lesion present on a skin with spots and redness effect. Figure 4.24 is a benign lesion having a lot of hair in the surrounding. Figure 4.25 shows an image of melanoma with multiple colours present on skin with prominent skin lines. Figure 4.26 contains a dysplastic nevi image with specular reflections.

It can be observed that level set method proposed initially by Chan [411] provided good segmentation in some cases but this method cannot track the border in the presence of spotty skin, many hairs or image having specular reflections as illustrated in Figure 4.23 (e), 4.24(e), 4.26(e) respectively. Similarly, K mean based LS method may result in inappropriate segmentation when the lesion colour is close to skin as shown in Figure 4.22 (b), 4.23(b), Spatial fuzzy clustering and adaptive thresholding based initialization of LS improved the segmentation results in most cases but still its accuracy is less than

the proposed algorithm. Analysis of the segmentation results shows that proposed method gives quite accurate results even in the presence of all these difficulties.

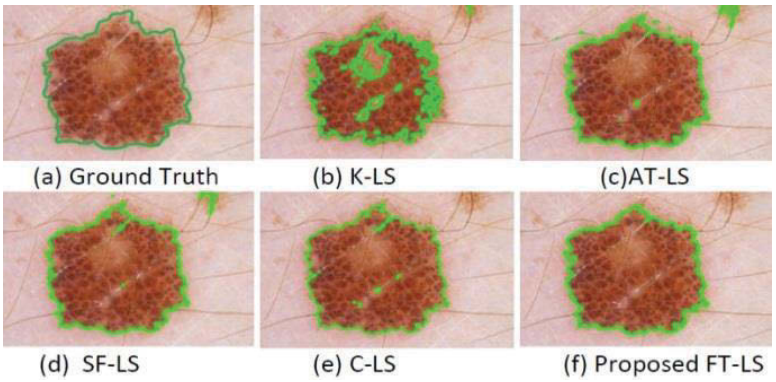


Figure 4.21 Segmentation results of a melanoma with hair and uneven border

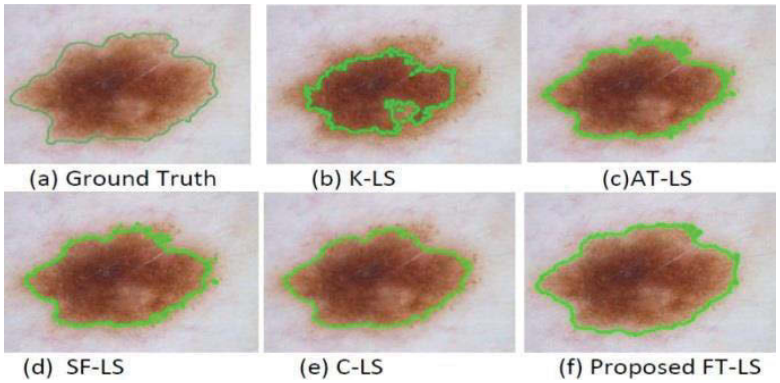


Figure 4.22 Segmentation results of a dysplastic nevi with smooth transaction between skin and lesion

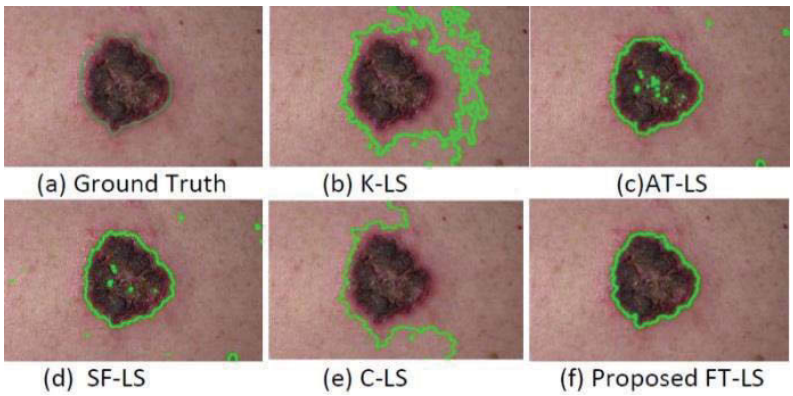


Figure 4.23 Segmentation results of a melanoma present on spotty skin with redness effect around lesion

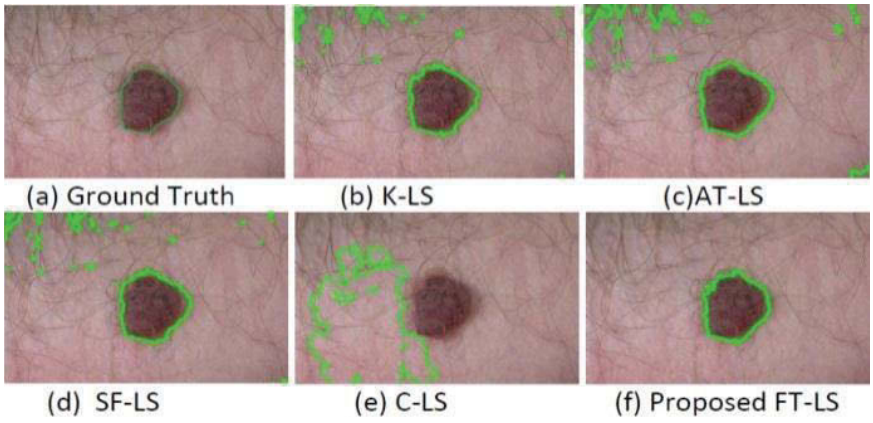


Figure 4.24 Segmentation results of a benign lesion surrounded by hair

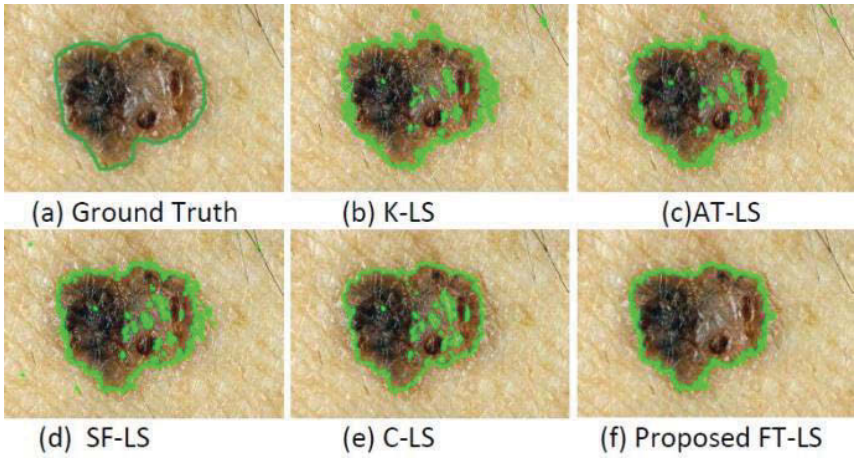


Figure 4.25 Segmentation results of a melanoma with multiple colours and prominent skin lines

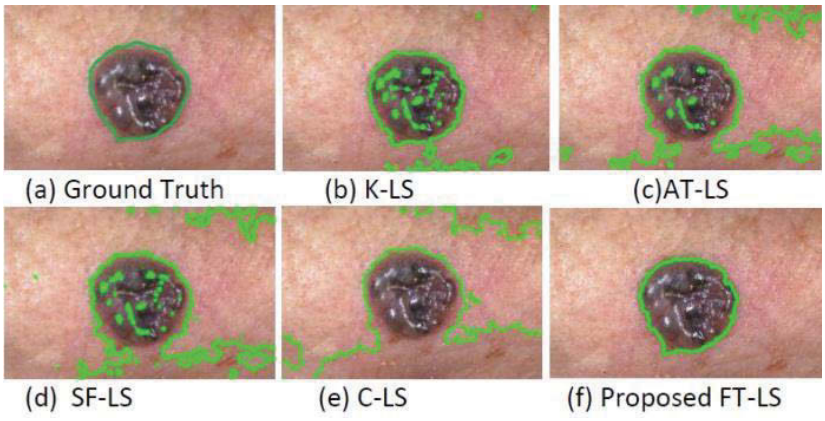


Figure 4.26 Segmentation results of a dysplastic nevi with specular reflection and uneven illumination

Table 4.2 shows the comparative segmentation results of the proposed method with those of four other used for LS initialization. In these experiments the database used

for analysis comprised of 270 dermoscopic and clinical view lesion images, which were collected from various sources, but most images were obtained from Sydney Melanoma Diagnostic Centre, Royal Prince Alfred Hospital. Most of these skin cancer images have been captured from digital camera. The segmentation results were compared with the reference images (ground truth) and average of segmentation scores are presented here for each method.

It is evident from the results that the proposed method has shown reasonably better performance as compared to other methods. Region based active contour method initially proposed by Chan. based on initialization using a rectangular region has shown good TDR but it has relatively high false positive error, which makes it susceptible of declaring benign lesions as melanoma. Secondly, it takes longer time to converge to the exact border while tracking the boundary as compared to the case when the initialization process is improved. Our experimental analysis showed that K-LS gave the worst results. One reason can be because it does not take into account the fuzziness present in the skin lesion images, while differentiating between lesion and skin. Thus, initialization using K-mean results can mislead the level set method parametric contours in tracking the boundary. This shows that misleading information during initialization can lead to even worse results.

The SP-LS segmentation method using spatial fuzzy clustering for initialization and AT-LS method using adaptive thresholding for initialization reduced the false positive error and the overall border error but true detection rate of these methods is not comparable to the proposed method. On the other hand, the proposed method that is a fusion of fuzzy clustering, thresholding and LS segmentation come up with better TDR and reduced false positive, false negative and overall border error. The Dice similarity coefficient of proposed method is quite high as compared to other methods. Thus, on the basis of the experimental analysis, it can be seen that the proposed method can show promising results for lesion segmentation in a computer aided diagnosis system.

Table 4.2 Results of Segmentation Methods. The values in bold correspond to the best performance

Method	Evaluated Parameters					
	HM (%)	TDR (%)	FPE (%)	FNE (%)	Border error (%)	DSC
C-LS	23.14	89.18	11.02	10.82	13.12	0.89
K-LS	31.33	80.12	22.6	19.88	18.79	0.79
SF-LS	22.33	88.57	7.97	11.43	9.54	0.90
AT-LS	16.71	88.92	8.02	11.08	6.22	0.90
Proposed FT-LS	**11.21**	**92.86**	**4.29**	**7.14**	**4.27**	**0.94**

The above performance evaluation has been carried out with different type of skin lesion images and the results were confirmed promising. However, sometimes skin the external skin image analysis is not sufficient and tissue and cell level analysis of the affected skin regions is also required for efficient diagnosis, which need more detailed segmentation process. Thus keeping an eye on the importance of tissue and cell level diagnosis of skin cancer, an extended version of this method was developed for segmenting histo-pathological images as well (in section 4.4) and the experimental results for segmentation of histopathalogical image are presented in the next section.

4.6 Experimental Results for Histopathalogical Images

The database used for analysis of proposed segmentation method includes 150 histopathological images taken from biopsies of skin cancer patients. The images were obtained from Sydney Melanoma Diagnostic Centre, Royal Prince Alfred Hospital. All the images were rescaled to a resolution of 720x 472 with bit depth 24. The whole process is implemented using MATLAB software R2013 and simulated by a system with corei5 3.10 GHz processor and 4 GB memory under Windows7 operating system. The segmentation results for some of the images obtained using the proposed method is

presented in Figure 4.27. It can be seen that the proposed method is able to identify the affected tissue areas in the image efficiently. The automated detection of the affected region of interest can be very helpful during the radiotherapy or other treatment stages.

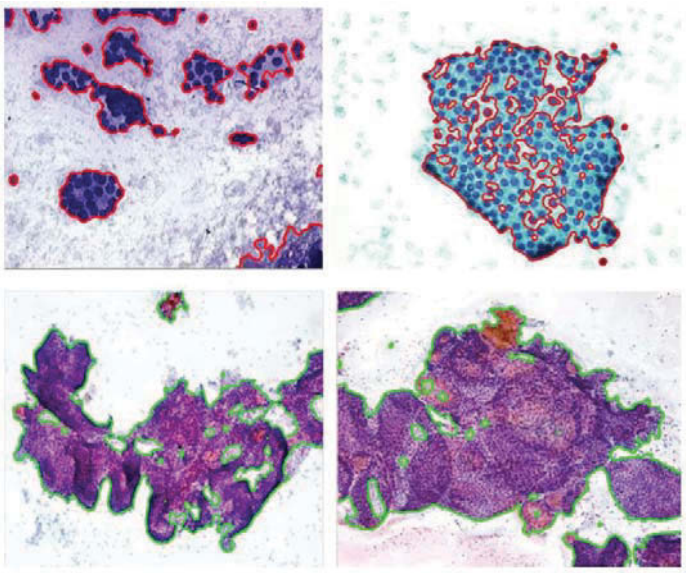


Figure 4.27 Segmentation results of the proposed method

Table 4.3 shows the comparative results of the proposed method with three methods used for the segmentation of histopathological images. First method used is watershed algorithm where the local minima of the gradient of the image are chosen as markers. It has been noticed that it is useful in detecting boundary lines between touching cells but sometimes resulted in over segmentation as found in some other medical imaging studies [424]. To overcome the over segmentation issue a second step involving region merging is used which help in reducing the false positive error. However, it is still much higher compared to the proposed method. Second method used in comparative analysis is the gradient flow vector (GVF) [198]. This method was unable to provide good segmentation for images with very small contrast between cancerous and non-cancerous regions but for images with better contrast it showed good conversion. The third method is expectation maximization based level set, which is one of the regions, based contour detection methods [425]. This method showed good true detection rate but the false positive rate was high. Similar to GVF, expectation maximization based level set (EM-LS) also showed relatively poor performance for images with low contrast. It is evident from the results that the proposed method has shown reasonably better performance compared to other methods. Based on the experimental analysis, it was found that the proposed method

could show reasonable results for segmentation of histopathological images of skin cancer.

Table 4.3 Comparative Analysis of Segmentation for Histopathalogical Images

Method	TDR (%)	FPE (%)	FNE (%)	DSC
Watershed algorithm	82.01	19.08	17.99	0.82
EM-LS	83.42	22.2	16.58	0.81
GVF	75.23	16.21	24.77	0.79
Proposed method	**87.66**	**11.66**	**12.34**	**0.88**

4.7 Summary

This chapter included the thesis contributions regarding the segmentation phase of the overall computer aided diagnosis process for skin cancer. The proposed segmentation method for digital/ dermoscopic skin lesion images and the extended version for the segmentation of histopathological images are discussed. The algorithms uses concepts of fuzzy C mean clustering; histogram intensity analysis based thresholding and level set evolution. Performance comparison has also been carried out with some of the popular methods used for segmentation of skin images and it can be observed that the results of the proposed algorithm are quite promising for both dermoscopic and histopathological images. Although all the experimental analysis is done for skin cancer images, however, future researchers can apply/extend this algorithm for the segmentation of other kind of images as well, which require precise segmentation.

Chapter 5

Thesis Contributions to Feature Extraction and Selection Methods

5.1 Introduction

This chapter presents the thesis contributions made for the feature extraction and selection phase of the automated diagnostic system. The chapter has two main sections. In the first section, the suggested features set extracted from the skin lesion dataset are presented. These features are further used in the analysis phase (i.e feature selection and classification stages) of the automated diagnostic system for skin cancer. While in the second section, the background and details of the proposed feature selection method are provided. Finally, experimental model and results are presented for skin cancer dataset analysis and the proposed feature selection method is also compared with other most popular feature selection methods when applied for analysis of various other types of cancer dataset as well.

Parts of work presented in this chapter are also published in peer reviews articles that include 1 article in Journal of Computer Science and Communication [426], 2 book chapters in Lecture notes in Computer Science [427, 428]and 2 international conferences [429, 430].

5.2 Feature Extraction for Skin Cancer Identification

Feature extraction is a crucial step in developing knowledge based automated diagnostic systems. This is because the particular features made available directly influence the efficacy of the classification stage. In order to classify melanomas from dysplastic nevi or regular benign nevi, it is required to extract and quantify those features that can separate the classes from one another. Unfortunately, the use of various different rules of diagnosis leave a wide room for interpretation even among dermatologists and it cannot be standardized of which features are more important and which are not. There is still a need to come up with feature sets that correspond to an established medical diagnostic rule or that can help in developing standard diagnostic rules for developing automated diagnostic systems. Therefore, in case of automated skin cancer diagnostic

systems, feature extraction stage can be used to calculate different parameters, which can reflect the physical parameters or patterns that are used in diagnosis by the dermatologists. In addition to this, different statistical parameter based on intensity, or texture analysis along with various image transforms based approaches can be used to characterize the images. These calculated parameters could help in differentiating cancerous and non-cancerous images. One of the main advantages of pattern classification is that it can also help in recognizing and quantifying notions that are very hard to describe even for a dermatologist.

The use of diverse range of feature extraction techniques can help in extracting various differentiating features and then efficient feature selection methods can help in selecting more relevant features that can help in better differentiation of different types of lesions. One of the research aims is to conduct experimental analysis using different feature extraction methods to propose the most relevant set of features that can help in identification of skin cancer.

The extracted feature set used in this research is mainly comprised of three main categories. 1) The physical features of lesions based on structure and colour distribution that are used in visual diagnosis. 2) Statistical parameters based on intensity analysis and mathematical modeling. 3) Features derived from transform based approach or spectral analysis of images.

5.2.1 Features based on Physical Shape and Colours

Melanoma and different types of benign tumours differ slightly in their physical characteristics and colours, (For medical details refer to Chapter 2). If any automated approach is to succeed in diagnosing melanoma, a collection of features covering different varying properties of lesions need to be used in order to obtain a satisfactory classification of the tumour images. The enormous variation in structure and colours of skin lesions makes the visual diagnosis very challenging. Thus, application of image processing techniques to these lesions can prove to be useful as an educational tool for teaching physicians to differentiate lesions, as well as for contributing information about the essential optical characteristics for identifying them. In the proposed automated diagnosis system different physical parameter are calculated which can reflect the parameters used in medical diagnosis [166]. These features consist of global features calculated for the

representation of the entire lesion and local features calculated for recognizing differences in several sectors of the lesion.

Melanoma is often defined primary by asymmetry and variations in colour, texture and border regions so most researchers focus on features that best represent this variability. In automated diagnosis of skin lesions, feature design is based on the ABCD-rule of dermoscopy. ABCD represent the asymmetry, border structure, variegated colour, and dermoscopical structures and define the basis for a diagnosis by a dermatologist. Pehamberger et al. [76]and Menzies et al [224]define further eight standard Epiluminescence microscopy criteria— pigment network, brown globules, black dots, radial streaming, pseudopods, overall pigmentation, and depigmentation—for the diagnosis with the help of Epiluminescence microscopy.

The ABCD system, 7-point checklist, 3 points checklist and Menzies method offers an alternative approach in extracting features of interest. According to the conclusion made by Johr [80] the automatic extraction of characteristics that take into account the rule ABCD [26, 222, 225] is computationally less expensive than the ones that take into account 7 points checklist [81, 223] or the Menzies method[31, 226]. Furthermore, the reliability in the clinical diagnosis is very high for ABCD rule. So, pigmented skin lesions are typically evaluated by dermatologists using the "ABCD" rule; which analyse the Asymmetry, Border irregularity, Colour variation and Dermoscopic structures of a lesion. Most of the automated decision support systems also use ABCD rule as the base of their feature extraction step. Three types of features are utilized in this study: Border/shape Features which cover the A and B parts of the ABCD-rule of dermatology, Colour Features which correspond to the C rules and Textural Features, which are based on D rules. In the following section, the physical features used in this research based on shape and colour variation are explained, while the features related to textural analysis are explained in the next section.

- **5.2.1.1 Shape features**

Shape is an important clinical feature in the diagnosis of skin lesions. In the following discussion, "object" refers to the binary lesion object obtained as a result of the border detection/segmentation. The geometry category features includes area (mm^2), border length (mm), irregularity, asymmetry (%), and circularity of borders (%), contour ratio and significant defects count of the lesion under analysis.

Area (A): The lesion area can be calculated by counting the number of pixels inside the border. However, it is shown in [431] that this method is not very accurate for objects with rough borders. In that case, the lesion area can be calculated using the method of bit quads [432] which has been shown to be one of the most accurate area estimators in the literature [431].

Border length: Border length is calculated simply by counting all boundary points on binary lesion mask. It is possible to estimate how irregular the border is by comparing it to the lesion size, e.g. circular lesion would have shorter border length comparing to irregular lesion of the same size.

Irregularity: The measure of the irregularity of the tumour border is estimated by the ratio of the square of the perimeter to the area of the tumour.

Asymmetry: A measure of the border asymmetry of the tumour is estimated as the percentage of non-overlapping area after a hypothetical(imaginary) folding of the border around the greatest diameter or along the best axis of symmetry (the maximum symmetry diameters).

Thinness Ratio/ Circularity index: It measures the isoperimetric Quotient of the lesion contour as a "Circularity Index (CI). The circularity of the skin lesion is defined as the ratio of the lesion contour area to the area of a circle with the same perimeter as that of the lesion contour:

$$\mathrm{CI} = \frac{4\pi A}{P^2} \tag{5.1}$$

Where A is the area within the lesion contour and P is the perimeter of the lesion contour. The maximum value of the CI is 1 and occurs when the lesion contour is a circle. The value of CI decreases as the lesion contour becomes more irregular.

Hull/Contour Ratio: This is defined as the ratio of the length of the convex hull of the lesion contour to the perimeter of the lesion contour. This is used as another measure of the 'raggedness' or 'spikiness' of the lesion border. Lesions whose shape can be closely approximated by their convex hull would have a higher HCR than those that have non-convex vertices in their polygon approximation.

Significant Defects: Irregular, notched or ragged edges are strong indicators that a lesion is not benign, while benign lesions are generally smooth edged. In this thesis, a convex-hull based method has been used to detect shape irregularity. The convex hull of a contour

is the smallest polygon that encloses it. A contour convexity defect between two contour convex hull key points occurs when the contour deviates inwards away from the convex hull. Thus a lesion with notched, ragged edges should have more convexity defects than a smooth un-notched lesion. A count of significant defects (SDC) is used as a feature. Determining the convex hull of a contour is accomplished in [168] using the implementation of Sklansky's algorithm. Defects with width greater than 5% of the convex hull length and depth greater than 20% of the defect width are considered as significant defects.

- **5.2.1.2 Colour Features**

Skin lesions are difficult to classify because of their short colour ranges as compared to other real-world images. Malignant melanoma is a kind of skin cancer that has some characteristic colour groups like: black, blue-grey, red, light brown, dark brown and skin colour. These colours appear on images and can depend on cancer progress stage, lesion depth and blood vessels. These colours are also used in melanoma diagnosis scales like ABCD. Calorimetric features quantitatively describe concepts such as presence of specific colours, their distribution and granularity, the irregularity of the pigmentation on the border of the lesion etc.

In order to quantify the colours present in a lesion, two statistics (mean and standard deviation) over the channels of different colour spaces and several colour asymmetry, histogram distance, and centroidal distance features can be calculated. The colour spaces used include RGB, rgb (normalized RGB), HSV, I1/2/3 (Ohta space), l1/2/3 and CIE L*u*v*. All of these except for the l1/2/3 space are well-known in the literature [432]. The l1/2/3 space is a colour space model described in [419, 433, 434] . The nonlinear transformation from RGB to l1/2/3 is given by:

$$l1 = \frac{(R-G)^2}{((R-G)^2+(R-B)^2+(G-B)^2)} \tag{5.2}$$

$$l2 = \frac{(R-B)^2}{((R-G)^2+(R-B)^2+(G-B)^2)} \tag{5.3}$$

$$l3 = \frac{(G-B)^2}{((R-G)^2+(R-B)^2+(G-B)^2)} \tag{5.4}$$

The mean and standard deviation values calculated over a particular channel quantify the average colour and the colour variegation in that channel, respectively. Plain RGB colour plane average and variance responses for pixels within the lesion can be calculated using built in functions of MATLAB. Intensity, Hue, Saturation Colour Space average and variance responses for pixels within the lesion are given as

$$I = \frac{R+G+B}{3} \tag{5.5}$$

$$S = 1 - \frac{3}{R+G+B}\,[\min(R,G,B)] \tag{5.6}$$

$$W = \arccos\{\ \frac{R-\frac{1}{2}(G+B)}{[(R-G)^2+(R-B)(G-B)^{\frac{1}{2}}}\ \} \tag{5.7}$$

H=W (if G>B) or H=2π − W (if G<B) or H=0 (if G=B)

Spherical coordinates LAB average and variance responses for pixels within the lesion are given as

$$L = \sqrt{R^2 + G^2 + B^2} \tag{5.8}$$

$$AngleA = \cos^{-1}[\frac{B}{L}] \tag{5.9}$$

$$AngleB = \cos^{-1}[\frac{R}{L\sin(AngleA)}] \tag{5.10}$$

When the lesion is divided into three regions, one hundred eight color features can be calculated: (6 color spaces)×(3 channels in each color space)×(2 statistics: mean and standard deviation)×(3 regions {lesion, inner periphery, outer periphery}).

The ratios and differences of the 2 statistics over the 3 regions can also be calculated: (outer/inner), (outer/lesion), (inner/lesion), (outer–inner), (outer–lesion), and (inner–lesion). The motivation for calculating the ratio and differences is two-fold. First, the colour characteristics of the three regions signify valuable diagnostic information. For example, a sharp transition from the inner periphery to the outer periphery (or vice versa) indicates malignancy. So, in addition to features calculated over the three regions, the differences and ratios might provide additional information about the transition between these regions.

Colour distribution in the lesion

For the diagnosis of melanoma asymmetry in colour distribution is a very important indicator of malignancy. In order to calculate it a binary drawing a white filled

copy of the selected contour on a black image the same size as the source creates lesion mask image. The means and standard deviations of the RBG values of the subset of pixels from the source image that correspond to white pixels in the mask are calculated.

The lesion is divided into 2 halves, left and right, at its central vertical axis. The difference between the colour features (RBG mean and standard deviation) of the lesion halves gives an indication of the evenness or unevenness of the lesion colour.

In addition to shape and colour, texture information is another important and efficient measure to estimate the structure, roughness, smoothness, or irregularity present in the skin lesion, which can help us in distinguishing between cancerous and non-cancerous lesions. Although, the human visual system can easily classify different textures in a general real worldview, it is not that easy to quantitatively characterize the texture pattern when it comes to diagnosis. Hence, texture analysis is regarded as one of the most difficult challenges in the diagnosis process especially for doctors with less experience.

Image texture can be broadly defined as a certain pattern repeated in a local area of an image [319]. It is generally agreed that texture exhibits periodicity of some basic patterns. The texture identification and classification have been proved to be critical in image segmentation, content-based image retrieval and MRI-, or X-ray- based diagnosis and prognosis. Research analysis shows that computer aided image analysis can help in developing better automated diagnostic models, that can save a lot of human effort and reduce the chances of false diagnosis caused by human error or incompetency. For years, the scientific community has been trying to figure out an effective method of feature extraction for revealing texture details and the succeeding classifications. The feature extraction methods used in this thesis for extracting textural details can be divided into following categories. 1) Features based on statistical analysis (first order and second order statistics) and autoregressive modelling 2) features based on spectral analysis including wavelet transform, Gabor wavelet and fuzzy mutual information based wavelet transform.

5.2.2 Features based on Statistical Analysis

The statistical analysis techniques used in the thesis for feature extraction include histogram intensity analysis, Grey level co-occurrence matrix, Grey-Tone Difference Matrix and autoregressive Modelling.

- **5.2.2.1 Intensity Histogram Features**

The characteristic of the histogram is closely related with the characteristic of image such as brightness and contrast. Histogram based features include:

$$\text{Mean } \mu = \sum_{i=1}^{N} p_i x_i \tag{5.11}$$

$$\text{Variance } \sigma^2 = \sum_{i=1}^{N} p_i \ (x_i - \mu)^2 \tag{5.12}$$

$$\text{Standard deviation } \sigma = \sqrt{\sigma^2} \tag{5.13}$$

$$\text{Skewness} = \sum_{i=1}^{N} \frac{(x_i - \mu)^3}{(N-1)(\sigma)^3} \tag{5.14}$$

$$\text{Kurtosis} = \sum_{i=1}^{N} \frac{(x_i - \mu)^4}{(N-1)(\sigma)^4} \tag{5.15}$$

$$\text{Energy} = \sum_{i=1}^{N} (p_i \ . p_i) \tag{5.16}$$

$$\text{Entropy} = -\sum_{i=1}^{N} (p_i \ . \log(p_i)) \tag{5.17}$$

The class mark of i^{th} intensity level is denoted by x_i , frequency of i^{th} level is denoted by f_i and the relative frequency of the i^{th} level is denoted by $p_i = \frac{f_i}{N}$, and N is the number of grey levels. Mean reveals the general brightness of an image. Standard deviation or variance tells the contrast of an image. Image with good contrast have high variance. Standard Deviations (SD) also characterizes the cluster. Skew measures how asymmetrical (unbalance) are the distribution of the gray level. Image with object in contrast background have high variance but low skew distribution i.e. one peak at each side of mean can be of a benign lesion. Energy measurement is also closely related to skewness. Highly skew distribution usually gives high-energy measurement. Entropy provides the average number of bits to code each gray level. It has inverse relationship with skew and energy measurement. Highly skew distribution tends to yield low entropy. Within ROI (region of interest) in the segmented skin image a histogram distribution of the image is computed. Then six features described above are calculated.

- **5.2.2.2 GLCM based features**

The grey level co-occurrence matrix (GLCM) is a one of the powerful tool for image feature extraction by mapping the grey level co-occurrence probabilities based on spatial relations of pixels in different angular directions. GLCM-based texture description is one

of the most well-known and widely used methods in the literature [246] . In developing the GLCM for texture representation, three fundamental parameters need to be defined: 1. Quantization levels of the image, 2. The Orientation, and 3. Displacement values of the measurements.

The GLCM features characterize the texture of an image by calculating how often pairs of pixel occur in an image with specific values and in a specified spatial relationship [245]. This calculation is used to create GLCM, and then statistical measures are extracted from this matrix. The matrix element p (i, j | d, ɵ) contains the second order statistical probability values for changes between gray levels 'i' and 'j' at a particular displacement distance d and at a particular angle (ɵ).

The co-occurrence probabilities provide a second-order method for generating texture features. These probabilities represent the conditional joint probabilities of all pair wise combinations of grey levels in the spatial window of interest given two parameters: inter-pixel distance (d) and orientation (θ) [246]. The probability measure can be defined as: $\{ C(i,j) \mid (d,\theta)\}$, where $C(i,j)$ (the co-occurrence probability between grey levels i and j) is defined as:

$$C(i,j) = \frac{p(i,j)}{\sum_{i,j=1}^{Ng} p(i,j)} \tag{5.18}$$

where $p(i,j)$ represents the number of occurrences of grey levels i and j within the given window, given a certain (d,θ) pair; and N_g is the quantized number of grey levels and is equal to the total number of rows and columns of the GLCM matrix. The sum in the denominator thus represents the total number of grey level pairs (i, j) within the window.

The mean (μ_x, μ_y) and standard deviations (σ_x, σ_y) for the rows and columns of the matrix are in the equations (5.19)-(5.22):

$$\mu_x = \sum_i \sum_j i.C(i,j) \tag{5.19}$$

$$\mu_y = \sum_i \sum_j j.C(i,j) \tag{5.20}$$

$$\sigma_x = \sum_i \sum_j (i-\mu_x)^2 .C(i,j) \tag{5.21}$$

$$\sigma_y = \sum_i \sum_j \left(i-\mu_y\right)^2 .C(i,j) \tag{5.22}$$

$p_x(i)$ is i^{th} entry in the marginal-probability matrix obtained by summing the rows of $C(i,j) = \sum_{j=1}^{N_g} P(i,j)$ Where N_g Number of distinct gray levels in the quantized image .

$$p_y(j) = \sum_{i=1}^{N_g} C(i,j) \tag{5.23}$$

$$p_{X+Y}(K) = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} C(i,j) \quad k = 2,3,\dots 2N_g \quad i+j=k \tag{5.24}$$

$$p_{X-Y}(K) = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} C(i,j) \quad k = 0,1,\dots . N_g - 1 \quad |i-j| = k \tag{5.25}$$

Although many statistics exist that can be derived from GLCM [245] but it is suggest to used thefollowing 16 grey level shift invariant statistics in order to obtain a texture characterization of skin images that is robust to linear shifts in the illumination intensity. Mathematical formulae of GLCM based statistics is provided in Table 5.1.

Table 5.1 GLCM based Features

Feature	Mathematical Equation
Autocorrelation	$\sum_{i=0}^{N_g-1}\sum_{j=0}^{N_g-1}(ixj)C(i,j)$
Contrast	$\sum_{n=0}^{N_g-1} n^2 \left\{\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} p(i,j)\right\}, \lvert i-j\rvert = n$
Uniformity	$\sum C(i,j)^2$
Correlation	$\sum_{i=0}^{N_g-1}\sum_{j=0}^{N_g-1}\frac{(ixj)C(i,j) - \{\mu_x \times \mu_y\}}{\sigma_x \times \sigma_y}$
Cluster Prominence	$\sum_{i=0}^{N_g-1}\sum_{j=0}^{N_g-1}\left(i+j-\mu_x-\mu_y\right)^4 \times C(i,j)$
Cluster shade	$\sum_{i=0}^{N_g-1}\sum_{j=0}^{N_g-1}\left(i+j-\mu_x-\mu_y\right)^3 \times C(i,j)$
Dissimilarity	$\mathrm{d} = \sum_{\mathrm{i,j}=0}^{N_g-1} C(i,j)\lvert \mathrm{i}-\mathrm{j}\rvert$

Energy	$\sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} C(i,j)^2$
Entropy	$-\sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} C(i,j) \times \log(C(i,j))$
Homogeneity	$\frac{\sum_{i=0}^{N_g-1} \sum_{i=0}^{N_g-1} (ij) C(i,j) - \mu_x \mu_y}{\sigma_x \sigma_y}$
Maximum probability	$\max\{C_{ij}\}$ for all (i, j)
Difference variance	variance of p_{x-y}
Difference entropy	$-\sum_{i=0}^{N_g-1} p_{x-y}(i) \log(p_{x-y}(i))$
Inverse difference normalized	$\sum_{i,j=1}^{N_g} \frac{C(i,j)}{1 + \lvert i-j \rvert^2 / {N_g}^2}$
Inverse difference moment normalized	$\sum_{i,j=1}^{N_g} \frac{C(i,j)}{1 + (i-j)^2 / {N_g}^2}$
Information measure of correlation	$\frac{-\sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} C(i,j) \times log\big(C(i,j)\big) + \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} C(i,j) \times log\{p_x(i)\, p_y}{\max\{entropy\ of\ p_x, entropy\ of\ p_y\}}$

Some investigations are performed to produce recommendations on how to set necessary parameters including: N_g, d,θ (N_g : quantized number of grey levels, d: inter-pixel distance, θ: orientation) and window size. It is accepted that larger window sizes will, in theory, provide more accurate classifications. Clausi et.al [246] tested the role of

N_g and θ and showed that above a certain threshold, for some statistics, with an increasing N_g, the discrimination power of the statistics remain the same for two of them (dissimilarity and contrast) while decreasing for the rest. Another problem with high N_g values is the high computational cost in both the calculation of the GLCM and the statistics.

To obtain statistical confidence in the estimation of the joint probability distribution, the normalized GLCM should be reasonably dense. For example, at full dynamic range (N_g = 256 grey levels for 8-bit images), since very few grey level pairs are repeated, the entropy statistic attains similar values for different texture patterns. In the approach presented by Celebi et.al[208], the images were uniformly quantized to 64 grey levels in order to avoid having a sparse matrix. But the choice of 64 grey levels in this study was arbitrary, though Clausi [246] has demonstrated that this value should not be too low (for example below 24). Another advantage of using a low N_g value is the reduction of the effects of noise in the image.

In order to obtain rotation invariant features, Celebi et. al [208] computed the normalized GLCM for each of the four orientations ({0◦, 45◦, 90◦, 135◦}) and the statistics calculated from these matrices were averaged. Based on the concepts proposed by Haralick [245] The co-occurrence matrices were computed here for 0, 45, 90 and 135 degrees and at distance 1.

A set of 4 values for each of the 16 measure given in Table 1 were obtained. The mean and range of each of these 16 measures, averaged over the four values, provided us with a set of 32 features. It was found that in this set of features some of the features are strongly correlated, so later feature selection procedure was applied to select the most differentiating features for the respective classification process.

- **5.2.2.3 GLRLM based features**

The Gray level Run length Matrix (GLRLM) method is used for extracting higher order statistical texture features. Let G be the number of gray levels, R be the number of different run lengths in the image data set, and RT is the total number of runs in the image given by $\sum_{i=1}^{G}\sum_{j=1}^{R} p(i,j|\theta)$ and Np be the number of pixels in the image given by$\sum_{i=1}^{G}\sum_{j=1}^{R} j.\, p(i,j|\theta)$. GLRLM is a two dimensional matrix having (GxR) elements in which each element p(i,j|θ) gives the total number of occurrences of runs having length j

of gray level i, in a given direction θ. GLRLM based features used in this thesis are provided in the Table 5.2. Galloway [435] introduced the five statistical texture features to be extracted from the GLRL matrices. Chu et al. [436]introduced two additional features i.e. LGRE and HGRE. These features use the gray level of the runs and help to distinguish textures that are similar according to their SRE and LRE features but differ in gray level distribution of the runs. Another four feature extraction functions following the idea of joint statistical measure of gray level and run length were described in [437].

Table 5.2 GLRLM based Features

Feature	Mathematical Equation
Short Runs Emphasis (SRE)	$\sum_{i=1}^{G}\sum_{j=1}^{R}\frac{p(i,j\|\theta)}{j^2}\Big/_{R_T}$
Long Runs Emphasis (LRE)	$\sum_{i=1}^{G}\sum_{j=1}^{R}j^2 p(i,j\|\theta)\Big/_{R_T}$
Gray Level Non-uniformity (GLN)	$\sum_{i=1}^{G}(\sum_{j=1}^{R}p(i,j\|\theta))^2\Big/_{R_T}$
Run Length Non-uniformity (RLN)	$\sum_{j=1}^{R}(\sum_{i=1}^{G}p(i,j\|\theta))^2\Big/_{R_T}$
Run Percentage (RP)	$\frac{R_T}{N_P}$
Low Gray Level Runs Emphasis (LGRE)	$\sum_{i=1}^{G}\sum_{j=1}^{R}\frac{p(i,j\|\theta)}{i^2}\Big/_{R_T}$
High Gray Levels Runs Emphasis (HGRE)	$\sum_{i=1}^{G}\sum_{j=1}^{R}i^2 p(i,j\|\theta)\Big/_{R_T}$
Short Run Low Gray-Level Emphasis (SRLGE)	$\sum_{i=1}^{G}\sum_{j=1}^{R}\frac{p(i,j\|\theta)}{i^2.j^2}\Big/_{R_T}$
Short Run High Gray-Level Emphasis (SRHGE)	$\sum_{i=1}^{G}\sum_{j=1}^{R}\frac{i^2.p(i,j\|\theta)}{j^2}\Big/_{R_T}$

Long Run Low Gray-Level Emphasis (LRLGE)	$\sum_{i=1}^{G}\sum_{j=1}^{R}\frac{j^2.p(i,j\|\theta)}{i^2}\Big/_{R_T}$
Long Run High Gray-Level Emphasis (LRHGE)	$\sum_{i=1}^{G}\sum_{j=1}^{R}i^2.j^2.p(i,j\|\theta)\Big/_{R_T}$

- **5.2.2.4 GTDM based features**

Texture analysis done using statistical techniques like GLCM are good for micro texture analysis, which is one of the requirements of our application. However, these techniques do not cater well for macro textures which are also an important factor for differentiating between cancerous and non-cancerous lesions [211]. It is a fair observation that human perception mechanism, works well for various types of textures. The properties that dermatologists use to discriminate between different textural patterns in skin lesions also include aspects of coarseness of skin, contrast in shades of skin, complexity of structure, busyness (fineness), shape of mole, directionality, and texture strength. Therefore, for general applicability of various existing texture measures, and also for improved performance in automatic classification of skin lesions on the basis of textural differences, it is relevant that measures used in the research reflect or represent to some extent some of the aforementioned textural properties.

In the present approach, an attempt is made to use computational measures corresponding to some textural properties that can be present in skin images, so as to ensure applicability of more general textural measures perceived by a human eye for differentiating skin lesions at the macro level. The idea is based on GTDM (Gray tone difference matrix) suggested by Amadasun [438] in an attempt to define texture measures correlated with human perception of textures. Five perceptual attributes of texture, namely: coarseness, contrast, busyness, complexity, and strength of texture, were computed. It is observed that the degree to which a given texture possesses any particular property mentioned above is considerably dependent on two factors, i.e., spatial changes in intensity and/or dynamic range of intensity.

In a normal digital image, information about spatial changes in intensity can be obtained usually by looking at the difference between the gray tone of each image pixel and the gray tones of its surrounding neighbours. Therefore, central to the development of the reported features was the computation of a one-dimensional (1-D) matrix for an

image, referred in literature as Gray Tone Difference Matrix (GTDM). A GTDM matrix is a column vector containing G elements. Its entries are calculated based on the difference between intensity level of a pixel and average intensity computed over a square, while sliding window centered at the pixel. Suppose the image intensity level $I_L(x,y)$ at location (x,y) is I, i=0,1,… G-1. The average intensity over a window centered at (x,y) will be $\bar{I}_l = \bar{I}(x,y) = \frac{1}{W-1}\sum_{m=-d}^{d}\sum_{n=-d}^{d} f(x+m, y+n)\ (m,n) \neq (0,0)$, where d specifies the window size and $W= (2d+1)^2$. The i^{th} entry of GTDM is $s(i)= \sum(i - \bar{I}_l)$ for all pixel having intensity level I, otherwise, s(i)=0.

Five different features derived from GTDM are computed, to quantitatively describe perceptual texture properties namely, coarseness, contrast, busyness, complexity and texture strength, that can be related to skin lesion texture analysis. The details of these features are as follows:

Coarseness

Coarseness is one of the most fundamental properties of texture and sometimes; it is used to imply texture in a narrow sense. It is a common observation that in a pattern referred as coarse texture, the primitives or basic patterns making up the texture are usually large. **Thus,** such a texture tends to possess a high degree of local uniformity in intensity, even over a fairly large area. Or, it can be said that the spatial rate of change in intensity is slight. Therefore, the intensities of neighbouring pixels would tend to be similar; thus there would be small differences between the gray tones of pixels and the average gray tones of their neighbourhoods. Hence the summation of such differences computed over all image pixels would give an indication of the level of spatial rate of change in intensity, and thereby (in an inverse manner) show the level of coarseness of the texture. This summation is the same as adding up the entries in the GTDM. However, in the summation process each entry is weighted by the probability of occurrence of the corresponding intensity value.

Therefore the computational measure for coarseness is given as:

$$f_{cos} = [\epsilon + \sum_{i=0}^{G} p_i s(i)]^{-1} \tag{5.26}$$

Where G is the highest gray-tone value present in the image and ϵ is a small number to prevent f_{cos} becoming infinite. For an NxN image, p_i is the probability of occurrence of gray-tone value i , and is given by

$$p_i = {}^{N_i}/_{n^2} \quad \text{where n= N-2d} \tag{5.27}$$

Where N_i is the set of all pixels having gray tone I (except in the peripheral regions of width d). The f_{cos} is the reciprocal of normalized sum of the deviations of pixel intensities from their neighbourhood average intensities of pixel intensities from their neighbourhood average intensities. Large values represent areas where gray-tone differences are small, i.e. coarse texture. Coarse textures in a skin lesion image (image should consist of mole along with the surrounding skin area) is mostly an indication of less chances of cancer, as coarseness shows that the texture of mole is somehow similar to the surrounding skin and no drastic random pattern changes exist on the lesion.

Contrast

Perceptually, an image is said to have a high level of contrast if areas of different intensity levels are clearly visible in the image. Thus high contrast can indicate that the intensity difference between neighbouring regions is large. The spatial frequency of the changes in intensity (i.e., the amount of local intensity variations in the image) will affect the contrast of an image. But, it should also be kept in mind that when the dynamic range of gray scale is large or when the image is stretched it can also effect the measurement of contrast. For instance, a small pigmented pattern will appear to have a higher contrast than a coarse pigmented pattern (when image is stretched) for the same gray scale range. Thus, these factors are kept into consideration in the following equation, when computing the contrast present in the image.

$$f_{con} = \left[\frac{1}{N_t(N_t-1)} \sum_{i=0}^{G}\sum_{j=0}^{G} p_i p_j (i-j)^2\right]\left[\frac{1}{n^2}\sum_{i=0}^{G} s(i)\right] \tag{5.28}$$

Where N_t is the total number of different gray levels present in the image.

$$N_t = \sum_{i=0}^{G} Q_i \quad \text{where } Q_i = \begin{cases} 1 & if\ p_i \neq 0 \\ 0 & otherwise \end{cases} \tag{5.29}$$

The $\boldsymbol{f_{con}}$ is a product of two terms. The first quantity is the average weighted squared difference between the different gray tone values taken in pairs, and is used to reflect the dynamic range of gray scale; the weighting factor is a product of the probabilities of the two gray-tone values under consideration. The second term is the average difference between pixel gray tones and the average gray tone of their neighbourhoods; this quantity increases with the amount of local variation in intensity.

Busyness

In a busy texture spatial frequency of intensity changes is very high, and there are rapid changes of intensity from one pixel to its neighbour. If these changes are very small in magnitude, then in that case, they may not be visually noticeable and a high level of local uniformity in intensity may be perceived. While on the other side, if the spatial frequency of changes in intensity is very low, a high degree of local uniformity may still be perceived, even if the magnitude of the changes is large. It should be noted that while the spatial frequency of intensity changes reflects the level of busyness, the magnitude of these changes depends upon the dynamic range of gray scale, and thus it also relates to contrast.

Therefore, a suppression of the contrast aspect from the information about spatial rate of change in intensity may indicate the degree of busyness of a texture. The following computational measure is used:

$$f_{bus} = \left[\sum_{i=0}^{G} p_i s(i)\right] \Big/ \left[\sum_{i=0}^{G} \sum_{j=0}^{G} i p_i - j p_j\right], \quad p_i \neq 0, p_j \neq 0 \tag{5.30}$$

The numerator is essentially a measure of the spatial rate of change in intensity, while the denominator is a summation of the magnitude of differences between different gray-tone values present in the image. Each value is weighted by its probability of occurrence. The denominator suppresses the effect of contrast variations. Thus, the expression would tend to emphasize the frequency of spatial changes in intensity values, i.e., busyness.

Complexity

Complexity refers to the visual information content of a texture and a texture is usually considered complex if the information content is high. This occurs when many patterns or primitives are present in the texture, and more so when the primitives also have different average intensities. Again a skin lesion texture with a large number of sharp edges and/or lines or with pigmented patterns may be considered as complex. All these depend upon the spatial period of pattern repetition and also on the dynamic range of gray scale. Thus complexity is partially correlated with busyness and contrast. A texture in which there are very rapid spatial changes in intensity is more likely to have higher complexity measure than a texture that has a high degree of local uniformity in intensity. However, rapid spatial changes from one intensity level to another is may be a result of presence of a large number of different levels of intensity (gray-tone values), but with a

low probability of each individual value occurring. Therefore, the sizes of primitives and/or probabilities of occurrence of gray-tone values tend to have inverse relationship with complexity. The equation used for computing complexity is as follows:

$$f_{com} = \sum_{i=0}^{G} \sum_{j=0}^{G} \left\{ \frac{|i-j|}{n^2(p_i+p_j)} \right\} \{p_i\, s(i) + p_j s(j)\},\ p_i \neq 0, p_j \neq 0 \qquad (5.31)$$

The f_{com} is a sum of normalized differences between the intensity values taken in pairs. These differences are weighted by the sum of the (probability-weighted) entries in the GTDM corresponding to the two intensity values under consideration. The normalizing factor, $n^2(p_i + p_j)$, which would take high values for coarse textures and small values for busy or fine textures, is used to show the inverse relationship between complexity and the sizes of primitives and/or the probabilities of intensity values. The absolute differences between the different intensity values are used to convey the influence of contrast variations on complexity, while the weighting factor reflects the rapidity or otherwise with which spatial changes in tonal values occur. High values of f_{com} should indicate a high degree of information content or existence of random various patterns in a skin lesion.

Texture Strength

The term texture strength is a difficult concept to define in a concise manner. However, a texture is generally referred to as strong when the patterns that comprise it are easily definable and clearly visible. Such textures present a high degree of visual feel. But the ease with which distinctions can be made depends to a considerable extent upon the sizes of the structures present and the differences between their average intensities. For instance it may be possible to distinguish between coarse textures even with small differences between their average intensities. However for such distinctions to be made in busy textures there must be wide differences between their intensities. Thus, in part, texture strength may be correlated with coarseness and contrast. The following computational approximation is used for this property (5.32)

$$f_{str} = \frac{\left[\sum_{i=0}^{G} \sum_{j=0}^{G} (p_i+p_j)(i-j)^2\right]}{\left[\epsilon + \sum_{i=0}^{G} s(i)\right]},\ \ p_i \neq 0, p_j \neq 0 \qquad (5.32)$$

This expression involves two terms. The numerator is a factor stressing the differences between intensity levels, and therefore may reflect intensity differences between adjacent patterns, particularly as the intensities are weighted by the sum of their

probabilities of occurrence; these probabilities would tend to be high for large patterns. *While,* the denominator can convey information about the size of texture pattern, as it is a sum of the difference between a pixel gray tone and the average gray tone in its neighbourhood over all pixels. Its value would be small for coarse textures (since neighbouring pixels generally have similar intensities in such textures), and high for busy or fine textures. Hence a high value of f_{str} would correspond to a strong texture and tend to emphasize the boldness or indicative of existence of distinct patterns in the image under consideration. For mathematical formulae see in Table 5.3 for quick reference.

Table 5.3 GTDM based Features

Feature	Mathematical Equation
Coarseness	$\left(\varepsilon + \sum_{i=0}^{G-1} p_i\, s(i)\right)^{-1}$
Contrast	$\left[\frac{1}{N_t(N_t-1)} \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} p_i p_j (i-j)^2\right] \left[\frac{1}{n} \sum_{i=0}^{G} s(i)\right]$
Business	$\frac{\sum_{i=0}^{G-1} p_{i\, s(i)}}{\sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \lvert i p_i - j p_j \rvert} \quad p_i \neq 0\,, p_j \neq 0$
Complexity	$\sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \frac{\lvert i-j \rvert}{n(p_i + p_j)} \left[p_i s(i) + p_j\, s(j)\right] \quad p_i \neq 0\,, p_j \neq 0$
Texture Strength	$\frac{\sum_{i=0}^{G-1} \sum_{j=0}^{G-1} (p_i + p_j)(i-j)^2}{\varepsilon + \sum_{i=0}^{G-1} s(i)} \quad p_i \neq 0\,, p_j \neq 0$

- **5.2.2.5 Autoregressive Modelling based features**

Autoregressive modeling is an all pole modeling which is widely used to get a robust spectral estimation of one dimensional signal [439]. Many methods have been proposed to find out autoregressive parameters. Yule-walker method is the most widely used method to estimate autoregressive coefficients. In order to determine the coefficients, Yule-Walker method uses Levinson-Durbin algorithm to minimize error.

This estimation method minimizes square of forward prediction error and finds out autoregressive parameters by solving autocorrelation function (5.33) as expressed in [440].

$$\sigma^2 = \frac{1}{N} \sum_{n=-\infty}^{\infty} \left| x(n) - \sum_{k=1}^{p} a(k)x(n-k) \right|^2 \quad (5.33)$$

Where x(n) is input and a(k) demonstrate autoregressive parameters. As images are 2-dimensional signals, so for doing autoregressive modeling, skin lesion images were converted into corresponding vectors. Autoregressive parameters modeling the image are calculated by Yule-Walker method. Estimated autoregressive parameters which are poles of 1-dimensional signals are used as features vectors extracted from images.

5.2.3 Feature extraction based on transform based approach

In case of image analysis for skin cancer diagnosis, the problem is that there are no firm clinical inspection rules for making a good judgment based classification system and the system requires a lot of subjective judgment on the part of the practitioner. One technique that does not require subjective judgment on feature extraction is frequency-domain based method. Past researches in various fields have proved that frequency domain analysis methods are more effective than spatial-based ones and others like model-based ones. Among frequency-domain based methodologies, Wavelet transform is investigated more thoroughly than others within the category. One advantage of wavelet transform is that it enables people to observe signals in different scales, under different frequencies, which is also its superiority over Fourier transform. The continuous advances in wavelet transform theory and application have provided more conveniences for researchers. Through wavelet transform, more details can be revealed and texture can be better characterized.

Typically, the majority of a dermoscopic image does not contain any particular relevant structure, and since pattern detection is a computer intensive task, the areas that contain any structure can be outlined by use of wavelet transform where large residuals of the wavelet compression is used as scoring for structure. Variability in various features like texture, colour, local changes like granularity is a factor the helps in separating malignant melanoma from benign nevi, therefore the best approach at feature extraction would retain as much of the data variability as possible.

The noteworthy thing about Wavelet analysis is that it looks at variability within a signal for example colour indexes in an image. Since images are only composed of colour values, changes in texture, granularity, and colour are all represented by the same value system [166]. Wavelet analysis looks at these changes over different scales which should detect whole lesion changes such as texture, colour, and local changes like granularity [184]. This section consists of details of feature extracted using 1) wavelet transform 2) Gabor wavelet based feature extraction. 3) As wavelet packet transform works better for image processing applications an extension of wavelet packet transform named Fuzzy Mutual Information based Wavelet-Packet Algorithm is also presented for extracting features for skin cancer images.

- **5.2.3.1 Feature extraction based on Wavelet transform**

The wavelet transform [441] is used to analyse the signal at different frequency bands with different resolutions by decomposing the signal into a coarse approximation and detail coefficients. The coefficients represent different frequency sub-bands. For discrete signals, the discrete wavelet transform (DWT) can be obtained by the discretization of time, translation and scale parameters. The wavelet $\Psi(x)$ is a linear combination of the scaling and translating of the scaling function$\varphi(x)$, which satisfies a two-scale differential equation, using Equations. (5.34) and (5.35):

$$\varphi(x) = \sqrt{2}\sum_k h(k)\varphi(2x-k) \tag{5.34}$$

$$\Psi(x) = \sqrt{2}\sum_k g(k)\varphi(2x-k) \tag{5.35}$$

where h is the low-pass filter and g is the high-pass filter. These two filters are quadrature mirror filters, which satisfy $g(k) = (-1)^k h(1-k)$.

To apply the wavelet transform to image processing, two-dimensional (2-D) wavelets are required from the vector product of $\Psi(x)$ and $\varphi(x)$ which are defined using Equations (5.36-5.39).

$$\varphi(x,y) = \varphi(x)\varphi(y) \tag{5.36}$$

$$\Psi^H(x,y) = \Psi(x)\varphi(y) \tag{5.37}$$

$$\Psi^V(x,y) = \varphi(x)\Psi(y) \tag{5.38}$$

$$\Psi^D(x,y) = \Psi(x)\Psi(y) \tag{5.39}$$

where $\varphi(x, y)$ is a 2-D scaling function, Ψ^H , Ψ^V and Ψ^D are three 2-D wavelets, and H ,V , and D represent the horizontal, vertical and diagonal directions, respectively.

Different mother wavelets have different supporting lengths and regularities, therefore the results of texture feature extraction highly depend on the selection of the mother wavelet and the number of levels. The wavelets generally used consist of Haar wavelet, Daubechies wavelets, biorthogonal wavelets, Coiflets wavelets, symlet wavelets. Several studies have compared the effect of various wavelets on the extraction of texture features [442-444]. Daubechies wavelets [445] has a family of orthogonal wavelets which define a discrete wavelet transform and characterized by a maximal number of vanishing moments. With each wavelet type, there is a scaling function (also called mother wavelet) which generates an orthogonal multi-resolution analysis. The Daubechies wavelet is a compactly supported orthogonal wavelet which is significant in the wavelet field and is used in the present study.

The number of decomposed levels determines the amount of details during textural extraction. The first level has the finest texture, and the fineness decreases as the number of decomposed levels increases. Synthesizing the texture at various levels can reflect an image's texture. However, the computational cost increases when more levels are decomposed. In the present study, when the image was decomposed from the histogram of coefficients, it was found that the image achieved better equalization at level 4. Therefore, after considering the computational costs, these images were decomposed at four levels. Three features are extracted at each level and in each direction (horizontal, vertical, and diagonal), i.e., the mean, variance and energy, which are defined using Equations (5.40-5.42) where C_{ij} is the coefficient matrix at point (i,j).

$$\text{Mean: } M = \frac{1}{i \times j} \sum_{i=0} \sum_{j=0} C_{ij} \tag{5.40}$$

$$\text{Variance: } V = \frac{1}{i \times j} \sum_{i=0} \sum_{j=0} (C_{ij} - M)^2 \tag{5.41}$$

$$\text{Energy: } e = \frac{1}{i \times j} \sum_{i=0} \sum_{j=0} \left| C_{ij} \right|^2 \tag{5.42}$$

- **5.2.3.2 Feature extraction based on Gabor wavelet**

A 2D Gabor function [446] is a Gaussian modulated sinusoid and it can be illustrated as following equation (5.43) as given in [447]

$$\Psi_{i,k}(m,n) = \frac{1}{2\pi\sigma_m\sigma_n}\exp(-\frac{1}{2}\left(\frac{m^2}{\sigma_m{}^2}+\frac{n^2}{\sigma_n{}^2}\right)+2\,\pi j\omega m) \quad (5.43)$$

Where, ω is the frequency of the sinusoid and σ_m and σ_n are the standard deviations of the Gaussian envelope. 2D Gabor wavelet are obtained by dilation and rotation of the mother Gabor wavelet $\Psi(m,n)$ using

$$\Psi_{i,k}(m,n) = a^{-1}\Psi[a^{-1}(mcos\theta + nsin\,\theta), a^{-1}\,(-msin\,\theta + ncos\theta], a > 1 \quad (5.44)$$

Where a^{-1} is a scale factor, l and k are integer, the orientation θ is given by $\theta = k\pi/K$, and K is the number of orientations. The parameter σ_m and σ_n are calculated according to the design strategy proposed by Majunath and Ma [446]. Given an image I(m,n), its Gabor wavelet transform is obtained as

$$x_{l,k}(m,n) = I(m,n) * \Psi_{l,k}(m,n) \text{ for l=1,2,…S and k=1,2,…K} \quad (5.45)$$

Where * represents a convolution operator. The parameters S and K are number of scales and number of orientations respectively. The mean and standard deviation are used as features and are given by

$$\mu_{l,k}(m,n) = \frac{1}{M^2}\sum_{m=1}^{M}\sum_{n=1}^{N}\left|x_{l,k}(m,n)\right| \quad (5.46)$$

$$\sigma_{l,k} = \left(\frac{1}{M^2}\sum_{m=1}^{M}\sum_{n=1}^{N}(\left|x_{l,k}(m,n)\right| - \mu_{l,k})^2\right)^{1/2} \quad (5.47)$$

The feature vector is then constructed using $\mu_{l,k}$ and $\sigma_{l,k}$ as feature components, for K=6 orientations and S=4 resulting in a feature vector of length 48.

- **5.2.3.3 Fuzzy MI-based Wavelet-Packet Algorithm for Feature Extraction**

In case of skin cancer images, it was quite obvious that details can have better discriminative capacities and can make the classification of two classes of subtle differences possible. Thus, extraction of information for details was also quite critical for better classification. From the frequency` s point of view, details are often contained in the middle and high frequency band. It was found that although wavelet transform and Gabor wavelet can reveal more details as compared to spatial-domain based methods to enhance the classification accuracy, it was still not up to the mark because details of first-level decomposition are not decomposed further, thus chances of loss of details are quite possible.

In order to overcome the weakness of wavelet transform, intuitively it comes to more promising algorithm-like Wavelet Packet Transform. Although wavelet packet transform and wavelet transform bear the same basic decomposition principles. However they are quite different in terms of the decomposition structures. WPT decomposes not only the approximation of the previous decomposition, but the detail components as well. The fact is that much of the information content needed for extracting image texture features is contained in the middle-high frequency, intuitively, WPT are expected to be more promising.

The wavelet-packets transform was initially introduced by Coifman et al. [448] by generalizing the relation between multi resolution approximations and wavelets. The wavelet packet transform may be thought of as a tree of subspaces, with $\Omega_{0,0}$ representing the original signal space, i.e., the root node of the tree. In general, node can be represented as $\Omega_{j,k}$, with j denoting the scale and k denoting the sub band index within the scale. Thus decomposition done into two orthogonal subspaces: an approximation space $\Omega_{j,k} \rightarrow \Omega_{j+1,2k}$ plus a detail space $\Omega_{j,k} \rightarrow \Omega_{j+1,2k+1}$. The major difference to the wavelet transform is that, for the subsequent decomposition levels, the wavelet packet transform not only decomposes the approximation coefficients, but also the detail coefficients.

This process is repeated J times, where $J \leq \log 2N$, and N is the number of samples in the original signal. This in turn results in $J \times N$ coefficients. Thus, at resolution level j, where $j = 1, 2, \ldots, J$, the tree has N coefficients divided into $2j$ coefficient blocks. This iterative process generates a binary wavelet-packet tree-like structure with multiple levels of decomposition.

After constructing the wavelet decomposition tree, the process of wavelet packet based feature extraction can be divided into two main steps. 1) Feature construction step where the coefficients generated at each of the wavelet packet tree subspace are utilized to obtain variables/properties that can represent the classes at hand. 2) Bases selection step for the identification of the best bases form which the constructed features can highly discriminate between the samples belonging to different classes under test.

The following approach is proposed for feature extraction based on wavelet packet transform:

1) For each original image vector, perform a full WPT decomposition to the maximum level J (taken as 4 here). For all j = 0, 1,. ., J and k = 0, 1, . . , 2j − 1,

2) Construct features according to relation $E_{\Omega j.k} = \log(\frac{\sum_n (w^T_{j,k,n}\, x)^2}{N/2^j})$.

Where $E_{\Omega j,k}$ is the normalized logarithmic energy of the wavelet packet coefficients extracted for the subspace $\Omega_{j,k}$ with $\Omega_{j,k}$ is the decomposition subspace with j denoting scale and k denoting sub-band index within the scale. $w_{j,k}$ x is the coefficient evaluated at the subspace $\Omega_{j,k}$ and $N/2^j$ is the number of coefficients in that specific subspace.

3) After obtaining n features f_i (i = 1,2,..n) from the decomposition tree, construct associated fuzzy sets and compute fuzzy entropies and mutual information. For this fuzzy entropy and mutual information based approach is used as suggested by Rami [449].

4) Evaluate classification ability of n No. of features using fuzzy-set based criterion Fi where

Fi = I(C; fi)/H (fi) for i = 1, 2, . . . , n.

Note: I (f;C) = H(f) – H(f |C)where H(f) is marginal entropy of f and H(f |C) is conditional entropy of f and C. For this calculate $F(\Omega_{j,k})$ according to the above equation, where $\Omega_{j,k}$ is the subspace representing each of the features.

5) Determine the optimal WPT decomposition X, being the one that corresponds to the maximum value of F, Starting with X= ∅.

6) Sort the subspaces by F in the descending order, Ω ={ Ω(1), Ω(2),… Ω(l)} , move first element in Ω to X.

7) ∀Ω(k) ∈ Ω , if Ω(k) is a father or child of $\Omega(j_1)$, remove Ω(k) from Ω.

8) if Ω=∅, stop, Otherwise go to step 6 and continue.

The set X is the final FMIWPT-based decomposition and this algorithm is applied to optimize the WPT tree of each skin image vector in the dataset to form corresponding feature vector.

5.3 Feature Selection

The aim of feature selection stage is to select a more relevant subset of features from the bigger feature sets. It is important in order to improve the prediction performance of the classification process by removing the features that are redundant or can adversely affect the performance of the classifier. It also helps in providing faster and more cost-effective classification performance by reducing the amount of redundant information to be handled by the classifier.

In this section, the details of proposed features selection method is provided that is based on adaptive differential evolution (ADE). The performance of proposed Adaptive differential Evolution algorithm is also compared with some algorithms based on Genetic Algorithm methods (GA), Ant Colony Optimization (ACO), and Particle Swarm Optimization (PSO), for skin cancer dataset analysis and various other types of cancer dataset as well.

5.3.1 Background of Proposed Feature Selection Method

The importance of feature selection phase has led to the development of a variety of techniques for selecting an optimal subset of features from a larger set of possible features. In chapter 3, a brief overview of the feature selection methods found in literature was provided.

Once a feature selection criterion is selected, a procedure must be developed which will find the subset of useful features. Directly evaluating all the subsets of features (2N) for a given data becomes an NP-hard problem as the number of features grows. Hence a suboptimal procedure must be used which can remove redundant data with tractable computations. In any feature subset selection method, there are some factors that need to be considered, the most significant are: the evaluation measure and the search strategy to find that subset. In other words, the search for the optimal feature subset requires an evaluation measure to estimate the goodness of subsets and a search strategy to generate candidate feature subsets.

As discussed in chapter 3, on the basis of evaluation measures, feature selection algorithms are divided into filters, wrappers and embedded approaches [257]. Filter approach is computationally faster. Wrapper and embedded methods are computationally more expensive than filters; however they are usually more accurate. On the other hand, although, wrapper and embedded approach can provide smaller subset of features with higher accuracy, but it is argued that filter approaches have more generalization tendency as it is independent of any learning algorithm. However in our proposed method, it was tried to encounter the effect of learning algorithm on the generalization tendency of the overall model, as well as do the feature selection while the model is being created. Therefore, the hybrid approach is used (i.e. merge of wrapper and embedded feature selection concept). Along with that the data is divided into training and testing phase, to ensure better generalization capability of the final model.

Search is a key topic in the study of feature selection methods. An efficient search strategy is needed to explore the feature space. Searching for the optimal subset, which can achieve the best performance according to the defined evaluation measure, is a quite challenging task. Various search methods that differ in their optimality and computational cost have been developed to search the solution space. These methods include: Exhaustive Search, Branch and bound method, sequential search algorithms and Heuristic/stochastic search algorithms. A brief description of all these methods are provided in chapter 3 for reference, but the main focus in this research is on evolutionary computation and population based search procedure like Differential evolution [271], as the proposed feature selection method in this research is also based on adaptive version of differential evolution algorithm which also comes under the category of stochastic/population based search algorithms.

- **5.3.1.1 Differential Evolution**

Differential evolution (DE) algorithm, proposed by Storn and Price [304, 305] emerged as a very competitive form of evolutionary computing method. It is a simple but powerful stochastic, population-based, evolutionary search algorithm for solving global optimization problems.

DE based algorithms has been frequently used to tackle different kind of optimization problems and its effectiveness and efficiency has been successfully demonstrated in many application fields [306-308] including recent applications in pattern recognition [309-312] which is the concerned area of this research.

DE has proven to outperform many other optimization algorithms in terms of convergence speed and robustness over many common benchmark problems and real world applications [308, 313, 314]. The strength of the algorithm lies in the fact that it has easier implementation and comparatively lesser parameter tuning is required. It provides speed of finding the optimal or suboptimal points of the search space and it is quite robust, i.e. it produces nearly same results over repeated runs. The algorithm uses only primitive mathematical operators and is conceptually very simple. In addition to that the performance does not deteriorate severely with the growth of the search space dimensions. These issues perhaps have a great role in the popularity of the algorithms within the domain of machine intelligence and pattern recognition. However, it should be noted that the control parameters and learning strategies involved in DE are highly dependent on the problems under consideration.

Basic DE algorithm is an evolutionary algorithm, used for optimization. DE also consists of mutation, crossover and selection phase. Note that the mutation scheme used in basic DE is DE/rand/1/bin. DE is started with an initial population vector, which is randomly generated with no preliminary knowledge about the solution space. Let $\vec{X}\_i_G$, i=1,2,…Np; are the solution vector, where i denote the population and G denote the generation to which the population belongs. The mutation, crossover, and selection operators for basic DE are defined as follows:

Mutation: For each target vector $\vec{X}\_i_G$ mutant vector $\vec{V}\_i_G$ is defined as per DE/rand/1 scheme in the original DE algorithm

$$\vec{V}\_i_G = \overrightarrow{X_{r1}\_i_G} + F \cdot (\overrightarrow{X_{r2}\_i_G} - \overrightarrow{X_{r3}\_i_G}) \tag{5.48}$$

To achieve good performance on a specific problem by using the original DE algorithm, one can try all available (usually 5) learning strategies in the mutation phase. These strategies include DE/rand/1, DE/best/1, DE/current to best/1, DE/best/2, De/rand/2.

$$\text{DE/rand/1 } \vec{V}_{i_G} = \overrightarrow{X_{r1_{i_G}}} + F.\left(\overrightarrow{X_{r2_{i_G}}} - \overrightarrow{X_{r3_{i_G}}}\right) \tag{5.49}$$

$$\text{DE/best/1 } \vec{V}_{i_G} = \overrightarrow{X_{rbest,G}} + F.\left(\overrightarrow{X_{r1_{i_G}}} - \overrightarrow{X_{r2_{i_G}}}\right) \tag{5.50}$$

$$\text{DE/current to best/1 } \vec{V}_{i_G} = \vec{X}_{i_G} + F.\left(\overrightarrow{X_{rbest,G}} - \vec{X}_{i_G}\right) + F.\left(\overrightarrow{X_{r1_{i_G}}} - \overrightarrow{X_{r2_{i_G}}}\right) \tag{5.51}$$

$$\text{DE/best/2 } \vec{V}_{i_G} = \overrightarrow{X_{rbest,G}} + F.\left(\overrightarrow{X_{r1_{i_G}}} - \overrightarrow{X_{r2_{i_G}}}\right) + F.\left(\overrightarrow{X_{r3_{i_G}}} - \overrightarrow{X_{r4_{i_G}}}\right) \tag{5.52}$$

$$\text{DE/rand/2 } \vec{V}_{i_G} = \overrightarrow{X_{r1_{i_G}}} + F.\left(\overrightarrow{X_{r2_{i_G}}} - \overrightarrow{X_{r3_{i_G}}}\right) + F.\left(\overrightarrow{X_{r4_{i_G}}} - \overrightarrow{X_{r5_{i_G}}}\right) \tag{5.53}$$

Where r_1,r_2,r_3,r_4,r_5 are randomly and mutually different integers generated in the range [1,NP], and it should also be different from the current trial vector's index i. $\overrightarrow{X_{rbest,G}}$ is the individual vector that achieved the best value for fitness function in the population at generation G. F is a real and constant factor having value suggested between [0,2] and it basically controls the amplification of differential variation $(\overrightarrow{X_{r2}\_i_G} - \overrightarrow{X_{r3}\_i_G})$

Crossover: Crossover phase is meant to increase the diversity of perturbed parameter vectors and trial vector is $\vec{U}\_i_G = [u_{1,i,G}, u_{2,i,G}, u_{3,i,G}, \ldots, u_{D,i,G}]$ where uj_i_G is defined as

$$uj_{i_G} = \begin{cases} vj_{i_G} & if\ (randi,j[0,1] \le Cr\ or\ j = j_{rand} \\ xj_{i_G} & otherwise \end{cases} \tag{5.54}$$

Where j= 1,2,…D (D= dimension of problem) , $rand_j \in [0,1]$; Cr is the crossover constant that takes value in the range [0,1] and k ∈ 1,2,… D; is the randomly chosen index.

Selection: It is the phase that decide which vector $\vec{X}\_i_G$ (target) or $\vec{U}\_i_G$ (trial) should be a member of next generation G+1. If vector $\vec{U}\_i_G$ yields a better performance in terms of objective function value than $\vec{X}\_i_G$ then $\vec{X}\_i_{G+1}$ is set to $\vec{U}\_i_G$; otherwise, the old value of $\vec{X}\_i_G$ is retained.

In order to use differential evolution for better feature selection, Rami [450] present a differential evolution based feature selection method by utilizing a combination of differential evolution (DE) optimization method and a repair mechanism based on feature distribution measures. This method, abbreviated as DEFS, utilizes the DE float number optimizer in the combinatorial optimization problem of feature selection. In order to make the solutions generated by the float-optimizer suitable for feature selection, a roulette wheel structure is constructed and supplied with the probabilities of features distribution. These probabilities are constructed during iterations by identifying the features that contribute to the most promising solutions. The DEFS was used to search for optimal subsets of features in datasets with varying dimensionality and showed quite promising result. The concept put forward in Rami's work form the basis of selection phase of the proposed adaptive differential evolution based feature selection method.

5.3.2 Proposed Adaptive Differential Evolution based Feature Selection

It was found during the research analysis that although use of DE as an optimization algorithm has reached an impressive state [313], but, there are still many open problems, and new challenging application areas are continually emerging in pattern recognition area that need the algorithm to be more adaptive to suite the data. Few adaptive versions of differnetail evolution are presented for general mathematical optimization problems [451, 452]. However, there is still a need for improving the search performance of DE to face challenges from the modern application areas of pattern recongnition.

The DE family consists of more than five distinct mutation strategies including DE/rand/1, DE/best/1, DE/current to best/1, DE/best/2, DE/rand/2 and two prominent

crossover schemes i.e. exponential and binomial. The rate of convergence of DE as well as its accuracy can be improved largely by applying different mutation and selection strategies. It can be found that each of these mutation and crossover operations may be effective over certain problems but poorly perform over the others; works like [453] provide the claims, counter-claims, and comparisons among these evolutionary operators in DE.

Another major issue faced while using a population based search technique like DE is the setting/ tuning of control parameters associated with it. Apart from the mutation strategy and the cross over operation, the performance of DE primarily depends on the intrinsic control parameters like scale factor (F), crossover rate (Cr), and population size (NP). Population size is normally user specified. However, an adaptive adjustment of the two key parameters namely the scale factor F and the crossover rate Cr can considerably effect the performance of DE. This is a fairly time consuming task but is also the most crucial task for the success of the algorithm in most applications. In order to make the performance of DE more robust, several research efforts can be found in literature [308, 454] regarding adaptive tuning of control parameters F and Cr. Similarly some objective functions are very sensitive to the proper choice of the parameter settings in DE. Therefore, researchers naturally started to consider techniques to automatically find an optimal set of control parameters for DE [455] [456].

During the study for best feature selection method for our application, it was found that DE can prove to be a good choice but it needs have more dynamic and adaptive tuning capability is a must for applying it to the analysis phase of more complex pattern recognition applications like that of skin cancer detection. Hence the following DE based feature selection and parameter optimization algorithm is proposed which can adaptively adjust the control parameter of DE algorithm for the problem under consideration, select the mutation strategy more dynamically a and optimize the parameters and select the feature subset, without degrading the classification accuracy.

The proposed algorithm suggests adaptive approach in the following three algorithmic components, where one or more of which can be integrated with the DE family of algorithms to improve their performances on complicated fitness landscapes.

1) First, a scheme is suggested to update the values of F and Cr in each generation, guided by the knowledge of their successful values that were able to generate better trial vectors in the previous generation. Finally, the adaptation schemes for the parameters include a certain degree of randomization that is likely to promote the explorative behaviour in a controlled manner. Varying the values of control parameters in successive generations provides more randomness to the algorithm which in turn may help in obtaining an explicit balance between exploration and exploitation as the algorithm proceeds.

2) Second, a less greedy and more explorative variant of the DE/current-to-best/1 mutation strategy is proposed. It is called DE/current-to-group_best/2. Unlike DE/current-to-best/1 that always uses the fittest (hence best) member of the entire current population to perturb a target (parent) vector, DE/current-to-group_best/2 forms a group, corresponding to each target vector, by randomly selecting population members to form a group whose size is, for example, n% of the total population size. The best member of this dynamic group, along with the best for the entire current best is used to perturb the target vector. Apart from that, for lessening the computational burden a method to probabilistically select one out of several available learning strategies is suggested in order to avoid the use of more complicated mutation strategy when it is not required.

3) Third, instead of the conventional binomial crossover scheme of DE a more exploitative crossover scheme, referred to here as "p-best crossover" [451]is used here. Under this scheme, a mutant vector is allowed to exchange its components not with the parent at the same index but with a randomly selected member from the p top-ranked individuals from the current generation through uniform (binomial) crossover. Note that it is not a new crossover operator but the usual binomial crossover with a biased parent selection strategy.

In what follows the details for each step of the proposed adaptive DE based feature selection algorithm are illustrated to make it suitable for tackling the optimization problems in our specific application and will later provide the results of analysis of the performance of the proposed algorithm for well-known datasets as well.

1. Initially the generation number is set as G = 0 and a population of *NP* individuals is randomly initialized say $Pop_G = \{\vec{X}\_1_G, ..., \vec{X}\_NP_G\}$ where $\vec{X}\_i_G$ = [x1_i_G, x2_i_G, x3_i_G, ..., xD_i_G] , with i = [1, 2, ...,NP] and D is the number of parameter to be optimized

including features and setting parameter for the classifier, Pop_G is population for G generation. Each individual parameter is uniformly distributed in the range $[\vec{X}_{min}, \vec{X}_{max}]$, where $\vec{X}_{min} = \{x1_{min}, x2_{min}, ..., xD_{min}\}$ and $\vec{X}_{max} = \{x1_{max}, x2_{max}, ..., xD_{max}\}$

2. At every generation the mutation and cross over control parameter are generated independently for each target vector using following relations.

$F_i = \text{Cauchy}(F_m, 0.1)$ with $F_m = (w_F . F_m) + ((1 - w_F) F_m\_best)$ and $w_F = 0.8 + 0.2 \times \text{rand}(0,1)$ Note F_m is initialized with value of 0.5 while Cauchy distribution prevent premature convergence due to its far wider tail property.

$Cr_i = \text{Gaussian}(Cr_m, 0.1)$ with $Cr_m = (w_{Cr} . Cr_m)) + ((1 - w_{Cr}) Cr_m\_best)$ and $w_{Cr} = 0.9 + 0.1 \times \text{rand}(0,1)$. Note Cr_m is initialized with vlaue of 0.6 and Gaussian distribution is used as opposite to Cauchy distribution, its short tail property help in keeping the value of Cr within unity [451] which is required here. The weight terms in the above equations are meant to control the average life span of successful F and Cr values. Small random perturbations to the weight terms are very effective in improving the performance of the corresponding DE algorithm for various different datasets analysed.

In the original DE, the 3 critical control parameters Cr, F and NP are closely related to the problem under consideration. In the proposed algorithm NP is kept as a user-specified value as in the original DE, so as to deal with problems with different dimensionalities while the value F and Cr are selected based on the above proposed formulations.

Earlier theoretical studies on DE [457, 458] have indicated that the scale factor F has a big role in controlling the population diversity and the explorative power of DE and it is more related to the convergence speed. According to our initial experiments, the choice of F has a larger flexibility, although most of the time the values between (0, 1] are preferred in literature. However, here F_m is used as a location parameter of Cauchy distribution, which diversifies the values of F more as compared to the traditional normal distribution. Our motivation for dynamically adjusting the control parameters using Cauchy and Gaussian distribution is inspired by the work of Minhazul et al. [451] however, it was chosen to not use the memorization of all successful scale factors in the current generation, as used by them as a trade-off to avoid excessive computational cost. The essence of F_m_best is that it memorizes the most successful scale factor in the current

generation. It enhances the chance of creating better donor vectors by tracking the best value found so far. On the other hand, the fact that the Cauchy distribution has a far wider tail than traditional Gaussian distribution [459], it is beneficial when the global optima is far away from the current search point as the values of F taken from the tail region give sufficient perturbation so that premature convergence can be avoided. **By using the suggested way, the scale factor over multiple populations is self adaptively determined and is found that a proper scale factor is dynamic over time (during a run) as well as problem dependent to cater for different data sets.**

The control parameter Cr, also plays an essential role in the DE algorithm and is much more sensitive to the problem's property and complexity. The proper choice of Cr can help to achieve good performance under several learning strategies while a wrong choice may cause performance deterioration under any learning strategy. Also, it was noted that the good Cr parameter value with which the algorithm can perform consistently well on a complex problem like ours, usually falls within a small range. Therefore, the learning experience is accumulated within a certain generation so that the algorithm can dynamically adapt the value of Cr to a suitable range.

The adaptation of Cr_m is also based on the record of recent successful crossover probability Cr_{m_best} and uses it to guide the generation of new Cr_i. This helps in generating better individuals as offspring, which are more likely to survive. The reason for avoiding the Cauchy distribution and selecting Gaussian distribution is that the long tail property of the former is not needed in case of the crossover probability adaptation. If the Cauchy distribution were used, the long tail property of the Cauchy distribution would lead to higher values of Cr, which would eventually get truncated to unity and would become independent of the Cauchy distribution. However, the short tail property of the Gaussian distribution provides the opportunity to generate most of the Cr values within unity.

According to the analyses and empirical results presented in [454] and [460], DE has a limited amount of search moves, and the effectiveness of such moves explicitly depend on how the scale factor values (F and Cr) are being adjusted over generations adaptive to the problem under consideration. In this regard, it was confirmed through experimental analysis for our specific problem and comparisons on standard data sets as

well, that the adaptations of the DE scheme proposed here results in improved performance.

3. While the termination criterion (maximum number of iterations) is not satisfied, for i = 1 to NP the following steps are followed for each individual

Step 1 – Mutation

The performance of DE algorithm is highly dependent on the mutation strategy. However, the most suitable strategy and the corresponding control parameters for a specific problem may require a huge amount of computation time. In addition to that it may be found that during different evolution stages, different strategies and corresponding parameter settings might be preferred. Therefore, the adaptive DE algorithm is developed that can automatically adapt the learning strategies during evolution.

The idea behind our proposed learning strategy adaptation consists of two parts. First a new mutation strategy is proposed named as DE/current-to-group_best/2, that works better than other mutation strategies in various applications. Secondly, a way to probabilistically select one out of several available learning strategies and apply to the current population is suggested. This procedure helps to determine the probability of applying suggested learning strategy. In our current implementation, two learning strategies are selected as candidates, i.e. "DE/rand/l" and DE/current-to-group_best/2 strategy. The reason for our choice for these strategies is as follows.

The oldest of the DE mutation schemes is DE/rand/1/bin, is said to be the most successful and widely used scheme in the literature [308]and demonstrates good diversity. However, [453]indicate that current to best/2 can give some advantages over DE/rand/1. Compared to DE/rand/*1*, greedy strategies like DE/current-to-best/*k* shows good convergence by guiding the evolutionary search with the best solution so far discovered. However, as a result of such exploitative tendency, in many cases, it compromises the diversity and global exploration abilities within a relatively small number of generations. Thus the chances of getting trapped to some locally optimal point in the search space are increased. Taking into consideration these facts and to overcome the limitations of fast but less reliable convergence the use of a less greedy and more explorative strategy that is DE/current-to-group_best/2 is proposed in this algorithm.

Unlike DE/current-to-best/1 that always uses the fittest member of the entire current population to perturb a target (parent) vector, DE/current-to-group_best/2

forms a dynamic group, corresponding to each target vector, by randomly selecting population members to form a group whose size is, for example, *n*% of the total population size. The best member of this dynamic group is used to perturb the target vector. This scheme is expressed as follows.

$$\vec{V}\_i_G = \overrightarrow{X_{rbest,G}} + F.\left(\overrightarrow{X_{group\_best,G}} - \vec{X}\_i_G\right) + F.\left(\overrightarrow{X_{r1}}\_i_G - \overrightarrow{X_{r2}}\_i_G\right) \qquad (5.55)$$

Where $\overrightarrow{X_{group\_best,G}}$ is the best of n% vectors randomly chosen from the current population, $\overrightarrow{X_{rbest,G}}$ is the best of entire current population, while $\overrightarrow{X_{r1}}\_i_G$ and $\overrightarrow{X_{r2}}\_i_G$ are two distinct vectors picked randomly from current population, that are distinct from the target vector or $\overrightarrow{X_{group\_best,G}}$ or $\overrightarrow{X_{rbest,G}}$. Under this scheme, the target solutions are not always attracted toward the same best position found so far by the entire population, and this feature is helpful in avoiding premature convergence at local optima. It is seen that keeping the group size equal to 15% of the population size provides very good results on majority of the tested benchmarks.

After mentioning the chosen and proposed mutation strategies, it is need to make the system adapt to the appropriate learning strategy at different stages of evolution. Since here two candidate mutation strategies are used, assuming that the probability of applying strategy "DE/rand/l" to each individual in the current population is p_1 , the probability of applying another strategy should be $p_2 = 1\text{-}p_1$. The initial probabilities are set to be equal i.e. $p_1 = p_2 = 0.5$, to give equal priority to both the strategies in the initial population. For the population of size NP, a vector of size NP is generated randomly with uniform distribution in the range [0, 1] for each element. If the 1th element value of the vector is smaller than or equal to the probability p_1, the strategy "DE/rand/l" will be applied to the jP individual in the current population. Otherwise the proposed strategy DE/current-to-group_best/2 will be applied. After evaluation of all newly generated trial vectors, the number of trial vectors that successfully entered the next generation while generated by the strategy "DE/rand/l" and the strategy "DE/current-to-group_best/2" are recorded as ns1, and ns2, respectively, and the numbers of trial vectors discarded are recorded as nd1 and nd2. Those two numbers are accumulated within a specified number of generations (50 in our experiments), called the "learning period". Then, the probability p1 and p2 are updated as:

$$p_2 = \frac{\text{ns2.(ns1+nd1)}}{\text{ns2.(ns1+nd1)+ns1.(ns2+nd2)}} \text{ And } p_1 = 1 - p_2 \qquad (5.56)$$

The above expression calculates and updates the probability of using respective mutation strategy based on the success rate of its generated trial vectors during the learning period. Also all the counters ns1, ns2, nd1 and nd2 get reset for the next learning period, in order to avoid the possible side-effect accumulated in the previous learning stage. **This adaptation procedure has the capability to gradually evolve the most suitable learning strategy at different learning stages for the problem under consideration.**

Step 2 – Cross over

Crossover is the process where the donor vector mixes its components with the target vector *Xi,G* to form the trial vector $U_{i,G} = [u_{1,i,G}, u_{2,i,G}, u_{3,i,G}, \ldots, u_{D,i,G}]$. There are two kinds of methods used by the DE family of algorithms for crossover 1. *Exponential* (or two-point modulo) and 2. *Binomial* (or uniform). Here only the binomial crossover is discussed as the proposed method uses variant of this scheme. Standard DE algorithm employ binomial crossover on each of the D variable as follows for building trial vector. The number of parameters inherited from the donor has a (nearly) binomial distribution. The scheme may be outlined as

$$uj\_i_G = \begin{cases} vj\_i_G & if\ (randi,j[0,1] \le Cr_i\ or\ j = j_{rand} \\ xj\_i_G & otherwise \end{cases} \quad (5.57)$$

where, as before, *randi,j*[0, 1] is a uniformly distributed random number, which is called anew for each *j*th component of the *i*th parameter vector. Here $j_{rand} \in [1,2,..D]$ is a randomly selected index to ensure that $\vec{U}\_i_G$ gets at least some component from $\vec{V}\_i_G$.

However, the proposed algorithm here uses *p*-best crossover operation [451] that is meant to incorporate a greedy parent selection strategy with the conventional binomial crossover scheme of DE. Here, for each donor vector, a vector is randomly selected from the *p* top-ranking vectors (according to their objective function values) in the current population, and then, normal binomial crossover is performed as per (5.57) between the donor vector and the randomly selected *p*-best vector to generate the trial vector at the same index. Parameter *p* is linearly reduced with generations in the following way:

$$p = ceil[\frac{NP}{2}\ (1 - \frac{G-1}{G_{max}})] \quad (5.58)$$

where *G* is the current generation number, G_{max} is the maximum number of generations, $G = [1, 2, \ldots, G_{max}]$, and *ceil*(*y*) function returns the lowest integer greater

than its argument *y*. The reduction routine of *p is meant to* favour exploration at the beginning of the search and exploitation during the later stages by gradually downsizing the elitist portion of the population, with a randomly selected member from where the component mixing of the donor vector is allowed for formation of the trial vector. This helps in getting a more efficient search process.

Step 3 – Selection

Evaluate the trial vector $\vec{U}\_i_G$ with the fitness/objective function f = error of classifier (which is needed to be minimized) to select whether the target or the trial vector survives to the next generation.

if f ($\vec{U}\_i_G$) ≤ f ($\vec{X}\_i_G$), this means that the new trial vector provides better performance, then $\vec{X}\_i_{G+1} = \vec{U}\_i_G$, else $\vec{X}\_i_{G+1} = \vec{X}\_i_G$, i.e. the target the vector is retained.

It must be noted here that as DE is a real number optimizer, two dimensions can settle at the same feature coordinates after rounding off. So before going to the next generation the roulette wheel weighing scheme [271] is utilized in order to overcome the problem of duplicate features.

For this a cost weighting is implemented where the probabilities of individual features are calculated from the distribution factor that is associated with each feature. The distribution factor of feature f_i within the current generation G is calculated as follows:

$$FD_{j,g} = a_1 \times \left(\frac{PD_j}{PD_j + ND_j}\right) + \frac{NF-DNF}{NF} \times \left(1 - \frac{(PD_j + ND_j)}{\max(PD_j + ND_j)}\right) \tag{5.59}$$

where NF is the total number of features and DNF number of desired features. PD_j and ND_j is the number of times feature f_i has been used in the good subsets and less competitive subsets respectively. Whereas, a_1 is the constant that reflects the importance of features in PD. Here $\frac{PD_j}{PD_j + ND_j}$ factor shows the degree to which feature f_i contributes in forming good subsets and the second term help in favouring exploration as this term will get close to 1 when the overall use of a particular feature is too low. Thus, based on this ranking the duplicated features of the trial vector are replaced by the next available top ranked features.

Secondly, for supressing the domination of certain features on the distribution factor, the relative difference in distribution factor is calculated using (5.60) [461].

$$T = ((FD_{G+1} - FD_G) \times FD_{G+1}) + FD_G \quad (5.60)$$

Where $FD_G = FD_G / \max(FD_G)$ and $FD_{G+1} = FD_{G+1} / \max(FD_{G+1})$. This will give higher weigh to features that are making noticeable improvement in the current generation as compared to previous one. It also helps in intentionally keeping features that are found to be highly relevant for a particular application, even if they do not show noticeable improvement.

These Steps are repeated until the algorithm reach a pre-decided generation number or stopping criterion is attained.

- **5.3.2.1 Computational Steps of ATDEFS Algorithm**

1. Initialize the population

2. Set the control parameters.

3. While the termination criterion (maximum number of iterations) is not satisfied

Do

for i = 1 to NP //do for each individual

3. Perform Mutation

3.1 Perform DE/rand/1 based mutation with a probability p_1,

or

3.2 Perform DE/group_best/2 mutation with a probability $1\text{-}p_1$.

4. Crossover operation based on p-best approach.

5. Evaluate the population with the objective function.

6. Perform roulette wheel based Selection.

end if

end for

7. Repeat from step 2 to 6 until G_{max}

5.3.3 Experimental Analysis

- **5.3.3.1 Experimental Model**

The model used for experimental analysis is presented in Figure 5.1. Table 5.4 shows the data sets used in the experiments. These are standard medical data sets used for benchmarking; however, the values of various attributes vary in different ranges. Therefore, firstly, the data sets under consideration are linearly scaled to the range [-1, +1] or [0, 1] to avoid the domination of features in greater numeric ranges on the ones with smaller numeric ranges. The optimization of feature subsets and control parameters is done based on the adaptive differential evolution algorithm explained in the previous section. After the evolutionary operations are done the trail vector provides the selected feature subset and the optimized SVM parameters. Using the selected feature sets the data divided into training and testing sets are fed into classification stage where the advised SVM classifier tuned on the basis of optimized parameters is used. The details for advised support vector machine algorithm are provided in the next chapter. Each trail vector is evaluated to see which provides less classification error or consequently more accuracy of the classifier. If the termination criterion (maximum number of generations) is satisfied, the process ends, otherwise proceed to the next generation and the evolution process is repeated.

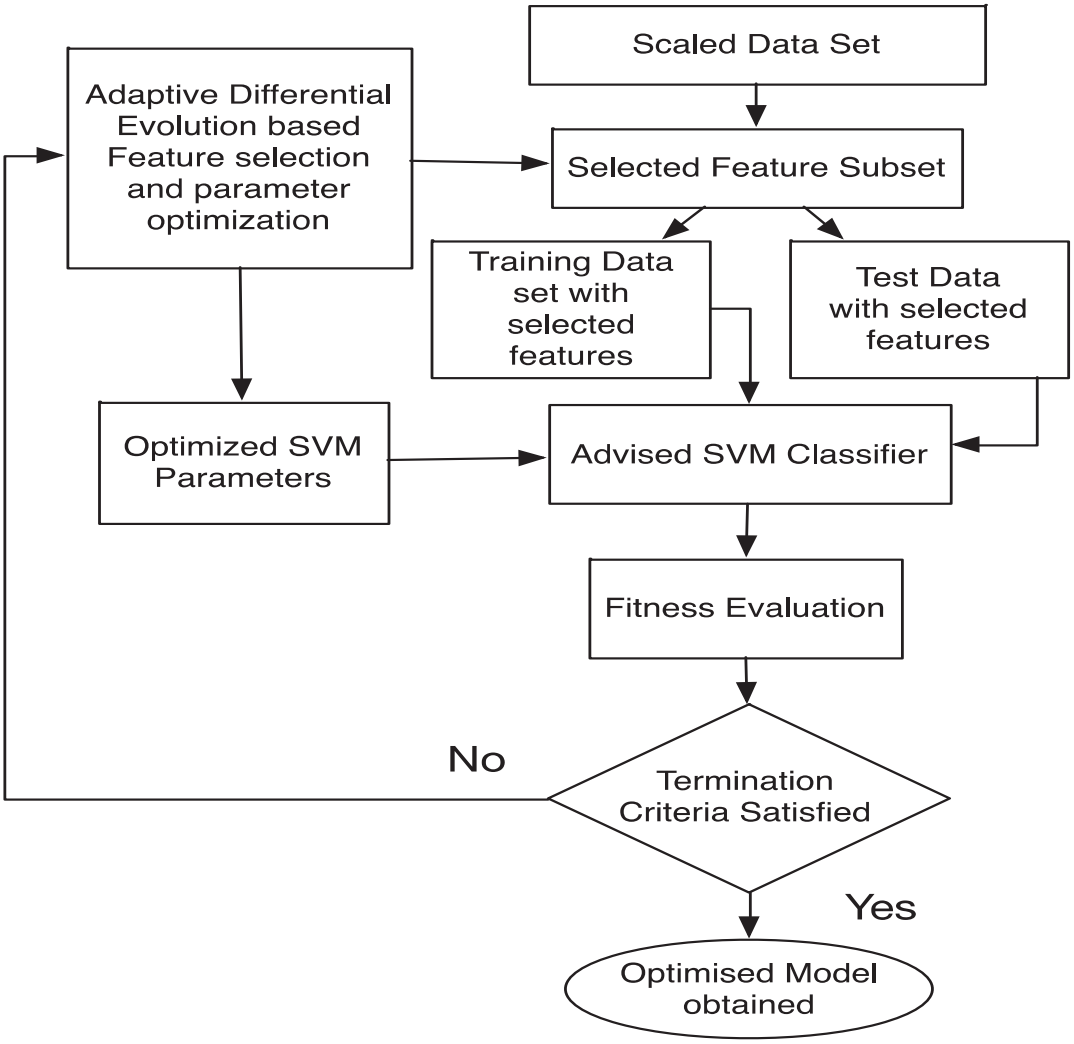


Figure 5.1 Experimental Learning Model

- **5.3.3.2 Test Datasets**

In order to analyse the effectiveness of the proposed model, 10 medical datasets with varying dimensionalities are utilized and the classification accuracies are calculated. See Table 5.4. The first 5 datasets (9_Tumor, 11 Tumor, 14 Tumor, Brain Tumo1, Prostate Tumor) used are available online from http://www.gems-system.org, colon dataset is obtained from http://research.janelia.org, dataset for breast cancer, Lung Cancer and dermatology is available from https://archive.ics.uci.edu/ml/machine-learning-databases and the skin cancer dataset is the data set based on skin cancer images collected from Sydney Melanoma Diagnostic Centre, Royal Prince Alfred Hospital. The details of the attributes used for skin cancer dataset is provided in the earlier half of the chapter.

Table 5.4 Description of Datasets used in experiments

Data Sets	No of attributes	No of Classes	No of instances
9_Tumor	5726	9	73
11_Tumor	12533	11	174
14_Tumor	15009	26	308
Brain Tumor1	5920	5	90
Prostate Tumor	10509	2	102
Colon	2000	2	62
Lung Cancer	56	3	32
Breast cancer	30	2	569
Dermatology	33	6	366
Skin Cancer	281	2	270

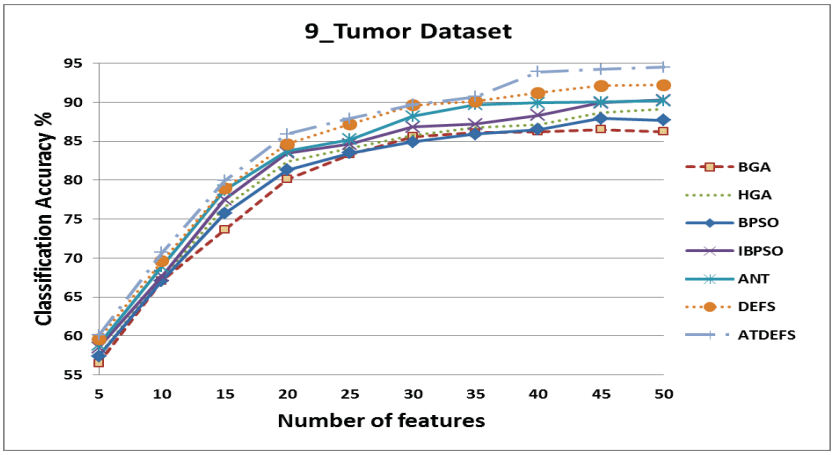
- **5.3.3.3 Comparative analysis**

For testing the effect of feature selection method and the number of selected features on the overall performance of the model, the performance of the proposed model is compared with the ones based on well-established binary Genetic Algorithm BGA [284], Binary PSO (BPSO) [289], improved BPSO [462]and hybrid GA [299], Ant colony based feature selection [300] and differential evolution based feature selection DEFS [271].

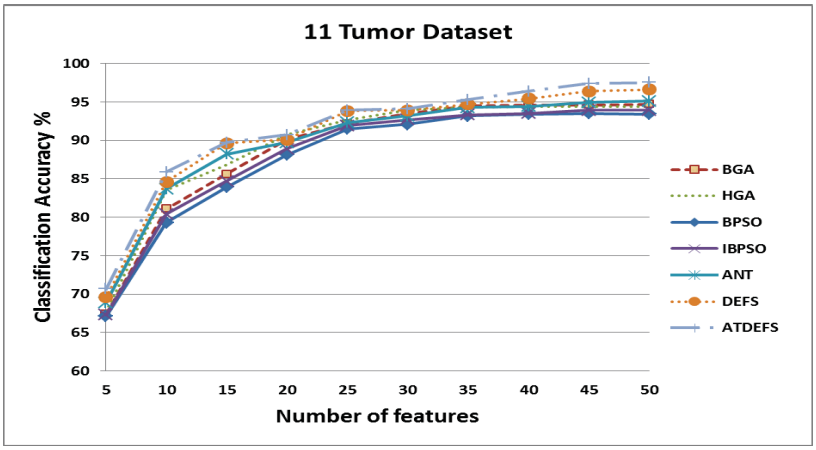
For BGA probability of mutation = 0.02 and probability of crossover was chosen as 0.5 after running several tests. This is used to make sure to have the number of ’1’s in the strings matching a predefined number of desired features. For BPSO the inertia weight was made to decrease linearly from 0.9 to 0.4 while the maximum velocity was set to be

clipped within 20% of the corresponding variable; and acceleration constants were set to 2.0. Both of BGA and BPSO utilize binary strings representing a feature subset with ones and zeros to indicate the selection and neglecting of features respectively. Improved binary particles warm (IBPSO) was implemented according to the algorithm described in [462]. Hybrid genetic search algorithm (HGA) was implemented as proposed in [299] to search for subsets of fixed sizes. It should be noted HGA is computationally very expensive for larger datasets, as the number of subsets to be formed and evaluated increases with the number of features in the dataset. Ant colony based feature selection used here is based on one proposed in [300] and will be referred here as ANT. While the DEFS method used for comparison is based on algorithm proposed in [450]that uses a combination of differential evolution optimization method and a repair mechanism based on feature distribution measures. This method has been compared with a range of methods proving its powerful performance [271] and forms the basis of the proposed method as well.

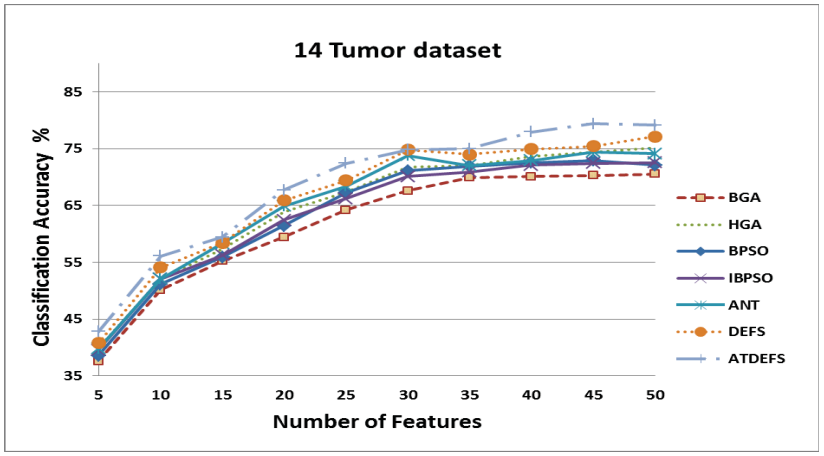
All methods were made to start from the same initial population with the population size set to 50 and terminated at the same number of iterations set to 100. 10 fold cross validation techniques is used and average classification accuracy for 20 runs is presented. Since the appropriate size of the most predictive feature subset is unknown, therefore, the desired number of features was varied up to 50 features for the first 5 datasets; however as for the last 5 datasets not a noticeable increase in classification accuracy was noted after 25 features so there was no need to continue to 50 features. The obtained results are shown in Figure 5.2 (a-j).



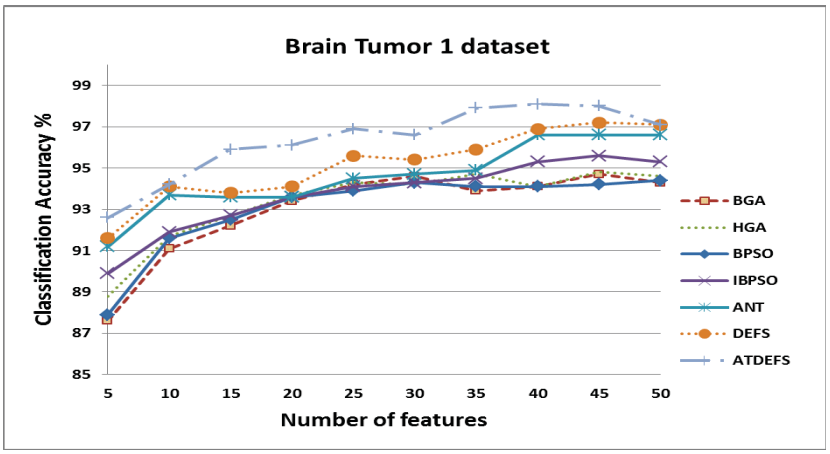
(a)


11 Tumor Dataset
Classification Accuracy %
100
95
90
85
80
75
70
65
60
5
10
15
20
25
30
35
40
45
50
Number of features
BGA
HGA
BPSO
IBPSO
ANT
DEFS
ATDEFS

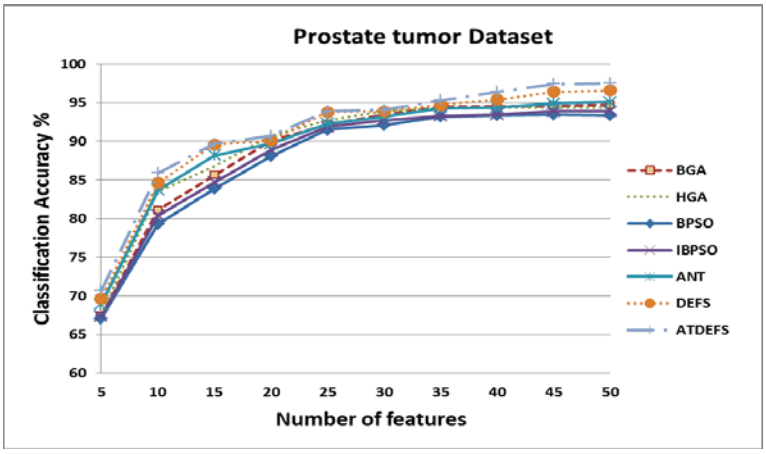
(b)


14 Tumor dataset
Classification Accuracy %
85
75
65
55
45
35
5
10
15
20
25
30
35
40
45
50
Number of Features
BGA
HGA
BPSO
IBPSO
ANT
DEFS
ATDEFS

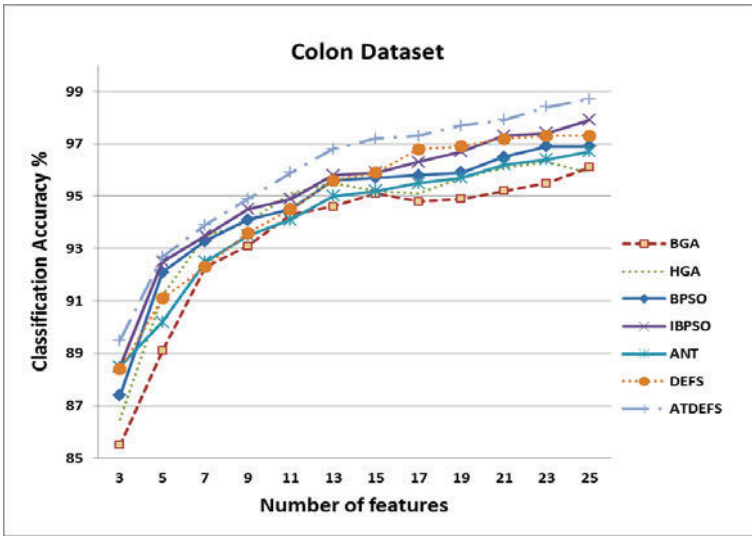
(c)


Brain Tumor 1 dataset
Classification Accuracy %
99
97
95
93
91
89
87
85
5
10
15
20
25
30
35
40
45
50
Number of features
BGA
HGA
BPSO
IBPSO
ANT
DEFS
ATDEFS

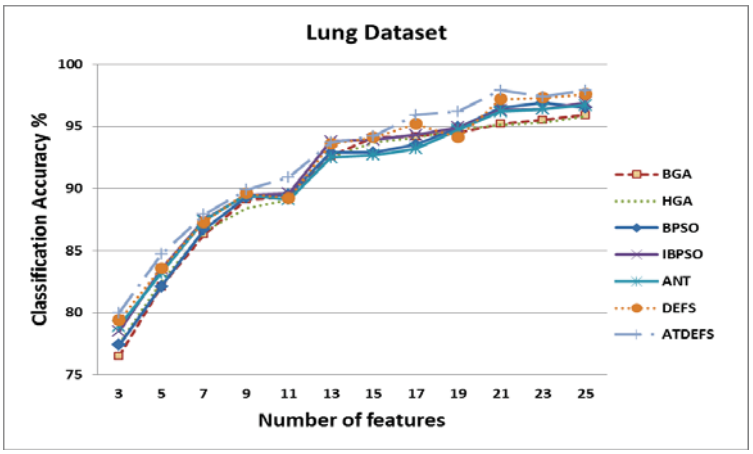
(d)


Prostate tumor Dataset
Classification Accuracy %
100
95
90
85
80
75
70
65
60
5
10
15
20
25
30
35
40
45
50
Number of features
BGA
HGA
BPSO
IBPSO
ANT
DEFS
ATDEFS

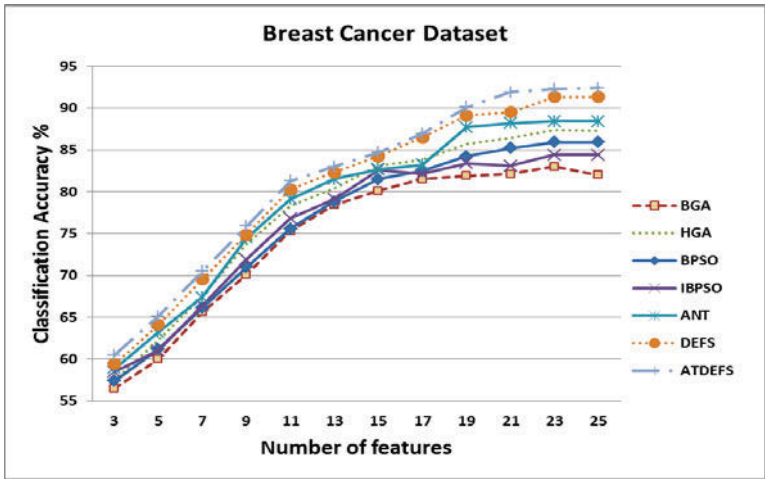
(e)


Colon Dataset
Classification Accuracy %
99
97
95
93
91
89
87
85
3
5
7
9
11
13
15
17
19
21
23
25
Number of features
BGA
HGA
BPSO
IBPSO
ANT
DEFS
ATDEFS

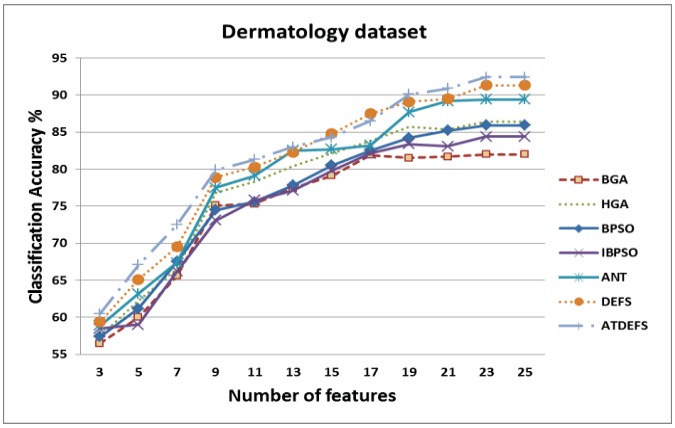
(f)


Lung Dataset
Classification Accuracy %
100
95
90
85
80
75
3
5
7
9
11
13
15
17
19
21
23
25
Number of features
BGA
HGA
BPSO
IBPSO
ANT
DEFS
ATDEFS

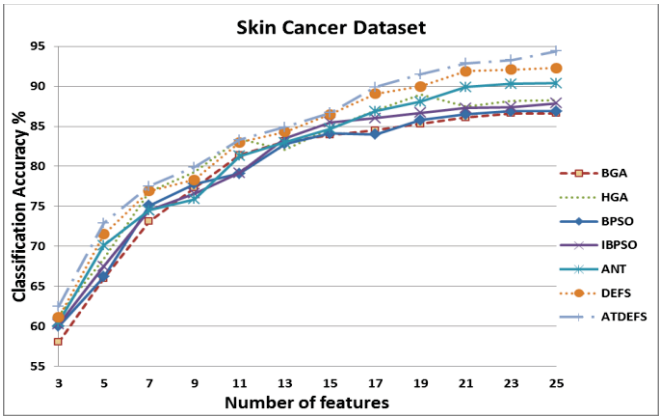
(g)



(h)



(i)



(j)

Figure 5.2 Average Classification Accuracies across 20 runs for each subset size across different datasets

In order to analyse the results one can start by looking at the performance of each of the BSPO, IBPSO2, BGA, HGA, ANT, DEFS and ADFES. It can be observed that on the basis of performance the results can be categorized into two parts. First category can

be of DEFS and ADEFS, with ADEFS showing improved performance for almost all datasets. Second category is of BGA, HGA, BPSO and IBPSO. For almost all datasets the performance of BGA, HGA, BPSO and IBPSO methods is close with HGA showing relatively better performance and IBPSO competing in case of 9 tumor, brain tumor 1 and colon dataset. The performance of BGA on the other hand is close to that of BPSO in some datasets and a bit worse in others. While the performance of the ANT method can be seen floating between these two categories. Despite the "on average" good performance of the ANT method in comparison to BGA,HGA BPSO and IBPSO, the relatively high computational and memory requirements of the ANT method can represent a major drawback.

The performance of DEFS proved to be better than that of BPSO, IBPSO, BGA and HGA methods on all considered datasets. On the other hand, the performance of DEFS is shown to be competing with that of the ANT method across the different datasets, with DEFS showing slightly better results. One main difference between the ANT and DEFS in spite of the very similar performance is that the DEFS highly reduced the computational cost that is required by the ANT method. There are many factors that highly increase the computational cost of the ANT method, like computational cost for estimating the mutual information and the corresponding memory requirements to store these matrices. Due to the sequential nature of the ANT-based feature selection method and the ant colony optimization methods in general, the computational time is usually much higher than that of DEFS, BPSO, IBPSO2, BGA and HGA, where each one of those algorithms has an embedded parallel component.

It can be seen that the proposed ADEFS method attained comparable or better classification accuracies as compared to other methods for comparatively lesser number of selected features for almost all datasets. Particularly, it achieved a classification accuracy of around 95% for the skin cancer dataset, which is the main dataset under consideration in this thesis. This shows that if parameter tuning and feature selection is done simultaneously, it can improve the classification accuracies of the learning models. In addition to that, it can also help in minimizing the use of redundant/irrelevant features in the final optimized model as it achieve better accuracy by using even smaller feature subsets, which will reduce the computational complexity on the final optimized model and also reduces the chances of having over fitted models.

5.4 Summary

This chapter provides the details of the feature extraction methods used for extracting the various quantitative features that can help in differentiation of melanoma and benign lesions during the analysis phase. An automated feature selection method is also proposed for coming up with the reduced number of features that works best in improving the accuracy of the training phase while maintaining the generalization capability of the classifier.

Experimental analysis shows that the proposed feature selection model works well and provides an optimal feature set with higher classification rate when compared with some other popular methods used in literature. The model is more adaptive and can work for various different types of datasets and help in choosing more relevant features that can help in classification and reducing the number of feature used in the final learning model. In future, this algorithm can be used for various other applications of optimization and feature selection.

It must be mentioned here that the feature sets suggested in the first part of the chapter are fed in various combinations to the classifier to record their contribution in identifying benign and melanoma in chapter 6. This helped in giving appropriate weight to different type of features on the basis of their performance in identifying melanoma and benign and to develop a more efficient classification mechanism.

Chapter 6

Thesis Contributions to Classification Methods

6.1. Introduction

This chapter provides details for the thesis contributions to the classification stage of the computer aided automated diagnostic system. The research objectives covered in this chapter include the development of more efficient algorithm for classification of skin lesions.

The chapter consists of two main parts. In the first part, parallelized classification model based on advised weighted support vector machine is presented that is proposed for developing an efficient learning model for classification of melanoma and benign lesions. The details and experimental/comparative analysis of the proposed learning model is provided. Getting significant amount of labelled data for the training phase is a major issue specially when it comes to histopathological image analysis and it often results in limited performance of classifiers in the testing phase. Thus in the second part, a parallelized learning model based on semi advised SVM and deep belief network is demonstrated that can be trained using significant amount of unlabelled data by making efficient use of the limited labelled data. The proposed model is tested for both dermoscopic and histopathological image analysis for skin cancer and the corresponding findings are presented.

Parts of work presented in this chapter are also published as peer reviewed articles that include 1 article in Journal of Computer Science and Communication [426], 1 book chapter in Lecture notes in Computer Science [427]and 2 other international conferences [463, 464].

6.2 Classification model based on labeled training data

Classification of the lesion as cancer or non-cancer is the final step. Classification is a significant area of research in a variety of fields including pattern recognition, artificial intelligence and vision analysis. Classification phase is one of the most important and final

stage of the computer aided diagnosis process. The supervised classification using labelled dataset includes two stages, namely, training and testing. At the training stage, a classifier is trained on a set of training feature vectors with an assigned label. At the testing stage, the classifier is applied to a set of test feature vectors that have been reserved for testing purposes (and are different from training feature vectors), and performance of the classifier is evaluated.

Predictive models are used in a variety of medical domains for diagnostic and prognostic tasks. These models are built from ‘‘experience’’, which constitutes data acquired from actual cases. The data can be pre-processed and expressed in a set of rules, such as it is often the case in knowledge-based expert systems, or serve as training data for statistical and machine learning models. Among the options in the latter category, the most popular models in medicine are logistic regression (LR), artificial neural networks (ANN), K nearest neighbour, decision trees, discriminant analysis and support vector machines. These models have their origins in two different communities (statistics and computer science), but share many similarities. This thesis proposes advised weighted Support vector machine algorithm that will be used in the classification phase.

The overall automated diagnostic system based on proposed methods for segmentation, analysis and classification of skin lesions is presented in the Figure 6.1. It is aimed to find the exact boundaries of a lesion automatically and also to provide an efficient estimate of the probability of a disease. The system bases its automatic diagnosis in three phases, detection of region of interest (as provided in chapter 4), calculation and analysis of lesion differentiating features (as provided in chapter 5) and final diagnosis phase using parallelized classification model aided with expertise based weight (EW) phase explained in the next section.

Section 6.2.1 provides brief background of SVM followed by details of the advised weighted SVM algorithm in section 6.2.2. Section 6.2.3 presents the details and experimental analysis of the Expertise Weight (EW) generation phase and parallelized classification model, followed by comparative performance analysis with some other popular classification algorithms used for the same application.

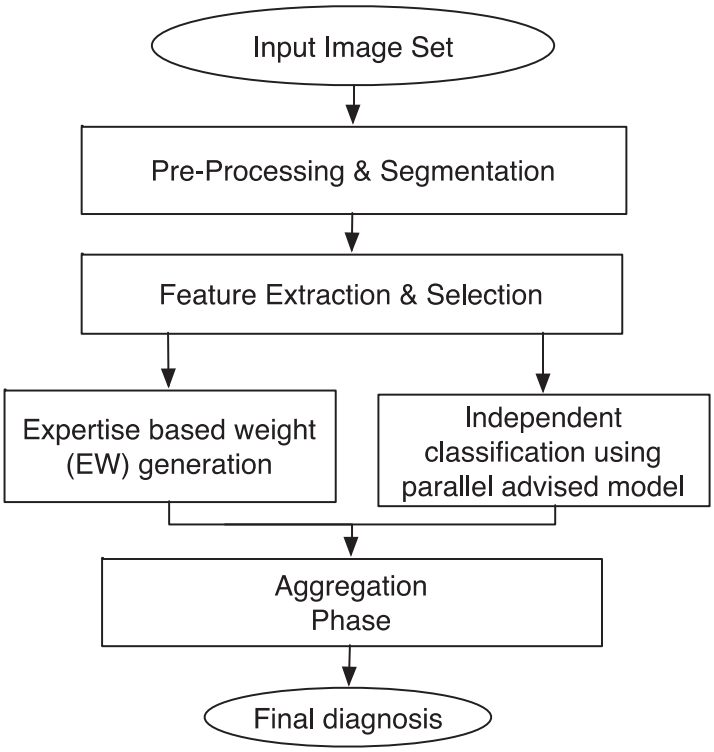


Figure 6.1 Proposed Computer Aided Diagnostic Model

6.2.1. Support Vector Machine

In this section most important aspects about support vector machines that form the background of proposed method are reviewed, more detailed discussion on general SVM is already provided in chapter 3.

The support vector machine (SVM) is a powerful classification method proposed by Vapnik [350]. Different types of SVM have been proposed in literature for use for various applications. While this method basically discriminated between two classes, it can still be used for multi class problems. SVM can find a good decision boundary between two classes, where the margin between the decision boundary and both classes have been maximised. Consider a binary classification, using a training set of N samples $(x_1, y_1), \dots, (x_i, y_i), \dots, (x_N, y_N) \in \mathcal{R}^n \times \{\pm 1\}$, where x_i is the input vector corresponding to the ith sample that is labelled by y_i depending on its class. SVM aims at separating the binary labeled training data with a hyper-plane that is at maximum distance from them. This is known as the maximum margin hyperplane. Figure 6.2 shows the basic idea of the SVM, graphically. The pair (w, b) defines the hyperplane with the equation $< w, x > + b = 0$. This hyperplane can separate the train data linearly if

$$y_i(w.x_i + b) \geq 1\,,\ i = 1, \ldots, N \tag{6.1}$$

The distance of each training data x_i from the hyperplane is given by

$$d_i = \frac{w.x_i+b}{\|w\|}, \tag{6.2}$$

and combining inequality (7.1) and (7.2), for all x_i will result in

$$y_i d_i \geq \frac{1}{\|w\|}. \tag{6.3}$$

Therefore, $\frac{1}{\|w\|}$ is the lower bound of the distance between the training data x_i and the separating hyperplane. The maximum margin of the hyperplane can be considered as the solution to the problem of maximising the $\frac{1}{\|w\|}$ subject to the constraint (6.1), or by solving the following problem

$$\text{Minimize} \quad z = \frac{1}{2} w.w \tag{6.4}$$

$$\text{s.t.} \quad y_i(w.x_i + b) \geq 1\,, \quad i = 1, \ldots, N.$$

Consider $(\alpha_1, \alpha_2, \ldots, \alpha_N)$ as the N non-negative Lagrange multipliers associated with the constraints (6.1), and without considering a few steps, the resulting decision function is given by [465],

$$f(x) = \text{sign}\left(\sum_{\alpha_i>0} y_i \alpha_i < x, x_i > + b\right), \tag{6.5}$$

Note that the non-zero α_i is those for which the constraints (6.1) are satisfied by the equality sign. This has an important consequence. Since most of α_i is usually zero the vector w is a linear combination of a relatively small percentage of the training data x_i. These points are called Support Vectors (SV) because they are the closest points to the separating hyperplane and the only points needed to determine the hyperplane. Support Vectors are the training patterns that lie on the boundaries of the margin. In reality, SVM only uses a small subset of the training samples SVs for the classification.

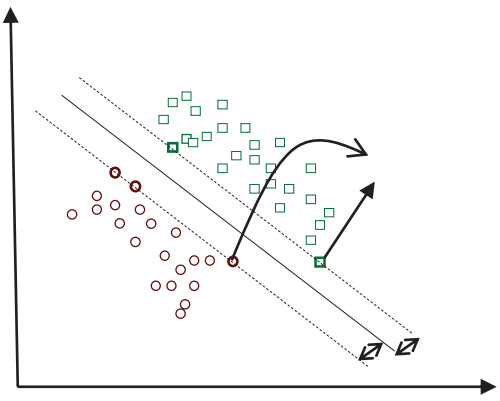


Figure 6.2 Basic ideas of support vector machines

There is also another type of support vectors that consists of the training data that are beyond their corresponding margins. These support vectors are regarded as misclassified data [466].

If the training data are not linearly separable, the problem of searching for a separating hyperplane is meaningless (there may be no separating hyperplane to start with). Fortunately, the previous analysis can be generalised by introducing N non-negative variables $(\xi_1, \xi_2, \dots, \xi_N)$ such that,

$$y_i(w.x_i + b) \geq 1 - \xi_i\,, \quad i = 1, \dots, N. \tag{6.6}$$

The purpose of the variables ξ_i is to enable a small number of misclassified points. If the data x_i satisfies inequality (6.1), then, ξ_i is zero and (6.6) reduces to (6.1). Instead, if the data x_i does not satisfy inequality (6.1), the extra term $-\xi_i$ is added to the right hand side of (6.1) to obtain inequality (6.6).

It should be noted that by introducing this tolerance parameter actually some training data were ignored in order to have a linearly separating hyperplane. The generalized separating hyperplane is then regarded as the solution to,

$$\begin{aligned} &\text{Minimize} \quad z = \frac{1}{2} w.w + C\sum_{i=1}^{N} \xi_i \\ &\text{s.t.} \quad\quad y_i(w.x_i + b) \geq 1 - \xi_i, \quad i = 1, \dots, N. \end{aligned} \tag{6.7}$$

The purpose of the $C\sum_{i=1}^{N} \xi_i$, is to keep the number of misclassified points under control. Note that this term leads to a more robust solution. The penalty parameter C can be regarded as a regularisation parameter. The above problem tends to maximise the minimum distance 1/w for small C, and minimise the number of misclassified points for large C. For intermediate values of C the solution of the problem (6.7) trades errors for a larger margin. In this case, the decision function is given by,

$$\begin{gathered} f(x) = \text{sign}\left(\sum_{\alpha_i > 0} y_i \alpha_i < x, x_i > + b\right), \\ 0 \leq \alpha_i \leq C, \quad i = 1, \dots, N \end{gathered} \tag{6.8}$$

In order to use the SVM to produce non-linear decision functions, the training data is projected to a higher dimensional inner product space F, called feature space, using a non-linear map $\phi(x)$: $\mathcal{R}^n \rightarrow \mathcal{R}^d$. The optimal linear hyperplane is computed in the feature space. Nevertheless, by using kernels it is possible to make all the necessary operations in the input space by using $k(x_i, x_j) = < \phi(x_i), \phi(x_j) >$ as $k(x_i, x_j)$ is an inner product in the feature space. The decision function can be written in terms of these kernels as follows:

$$f(x) = \text{sign}\left(\sum_{\alpha_i>0} y_i \alpha_i k(x, x_i) + b\right). \quad (6.9)$$

Also, the decision value for each x of the test set which can get a negative or positive value depends on the position of the x and the hyperplane, which is defined as equation (6.10).

$$h(x) = \sum_{\alpha_i>0} y_i \alpha_i k(x, x_i) + b \quad (6.10)$$

There are 3 common kernel functions in SVM:

Polynomial kernel: $K(x_i x_j) = (x_i x_j + 1)^q$

RBF kernel : $K(x_i x_j) = e^{-\gamma |x_i - x_j|^2}$

Sigmoid kernel: $K(x_i x_j) = \tanh(\gamma x_i^T x_j + c)$

Here q, γ, c are kernel parameters.

6.2.2 Advised weighted Support Vector Machine

The classic SVM ignores the training data that has not been separated linearly by the kernels during the training phase. This occurs through the introduction of the tolerance parameters in the objective function and constraints. Thus, if data that is similar or identical to this misclassified data appears in the test set, it will be classified wrongly. This is because the data which is close to the misclassified data is uncertain. This misclassification is not reasonable and it can be handled if the available data and information in the training phase has not been ignored by the SVM algorithm.

Several modified versions of SVM are presented in literatures that are aimed to increase general classification efficiency as well as performance for special cases [467] [468] [469]. Similarly, incorporating prior knowledge in SVM few methods are presented [470, 471]. They include kernel methods that select or create new kernel; methods that generate new data or modify the way data is used as well as some optimization methods that add constraints to original problem or try to define new formulation for the problem in hand. Use of advice sets based on expert knowledge is one other approach but it comes with the increase of computational cost due to the added parameters.

In this study a non-iterative self-advising approach for SVM is adapted that extracts subsequent knowledge from training phase without adding extra parameters. The misclassified data can come from two potential sources 1) outliers 2) data that have not been separated correctly. Advised Weighted-SVM deals with the ignorance of SVM from the knowledge that can be acquired from misclassified

data by generating advice weights based on the distance of misclassified training data from the correctly classified training data, and through use of these weights together with decision values of SVM in the test phase. These weights also help the algorithm to eliminate the outlier data.

The advised weighted SVM algorithm is illustrated in Figure 6.3. The details of algorithm are as follows:

1. The classifying hyperplane is found by using decision function$f(x) = \text{sign}(\sum_{\alpha_{i>0}} y_i \alpha_i \, k(x, x_i) + b)$, here xi is the input vector corresponding to the ith sample and labelled by yi depending on its class and αi is the nonnegative Lagrange multiplier that is inconsistence with standard SVM training.

Note that in order to use SVM to produce non-linear decision functions as the data is comprised of nonlinearly separable cases, radial basis function kernel $K(x_i x_j) = e^{-\gamma |x_i - x_j|^2}$ is used to make all necessary operations in the input space. The reason to choose RBF kernel is that unlike the linear kernel, it can handle the cases when the relation between class labels and attributes is nonlinear. On the other hand, sigmoid kernel behaves like RBF for certain parameters and has number of hyper-parameters which influence the complexity of model. While, polynomial kernel has more hyper parameter than RBF kernel and adds more to the complexity of model as compared to any increase in classification accuracy.

2. The data samples that are misclassified in the initial training phase are identified. The misclassified data sets (MD) in the training phase is determined using (6.11)

$$MD = \cup_{i=1}^{N} x_i \; | y_i \neq \text{sign}(\sum_{\alpha_j > 0} y_i \, \alpha_j \, k(x_i, x_j) + b) \qquad (6.11)$$

The MD set can be null, but experimental results revealed that the occurrence of misclassified data in training phase is a common occurrence. It must also be noted that any method that tries to get benefit from misclassified data, must also have some control on the impact of outlier data. It is observed that when the misclassified data is comprised of resembling samples, the use of misclassified data actually improved the classification accuracy more.

3. If the MD is null, go to the testing phase, else compute neighbourhood length (NL) for each member of MD. NL is given as (6.12)

$$NL(x_i) = \text{minimum}_{x_j} \left(\|x_i - x_j\| \mid y_i \neq y_j\right) \tag{6.12}$$

Where xj , j=1, …., N are the training data that do not belong to the MD set. Here the training data is mapped to a higher dimension, the distance between xi and xj is computed according to the following equation with reference to the related RBF kernel.

$$\|\theta(x_i) - \theta(x_j)\| = (k(x_i, x_i) + k(x_j, x_j) - 2k((x_i, x_j))^{0.5} \tag{6.13}$$

4. For each sample x_k from the test set advised weight $AW(x_k)$ is computed. Where AW is computed as (6.14), These AWs represent how close the test data is to the misclassified data.

$$\begin{cases} 0 & \forall x_i \in MD, \|x_k - x_i\| > NL(x_i) \text{ or } MD = NUL, \\ \sum 1 - \frac{\sum_{x_i} \|x_k - x_i\|}{\sum_{x_i} NL(x_i)} & x_i \in MD, \|x_k - x_i\| \leq NL(x_i) \end{cases} \tag{6.14}$$

5. The absolute value of the SVM decision values for each x_k from the test set are calculated and scaled to [0, 1].

6. For each xk from the test set, If (AW (x_k) < decision value (x_k) then $y_k = \text{sign}\left(\sum_{\alpha_j>0} y_j \alpha_j k(x_k, x_j) + b\right)$ which is in consistence with normal SVM labelling, otherwise $y_k = y_i \mid (\|x_k - x_i\| \leq NL(x_i) \text{and } x_i \in MD)$

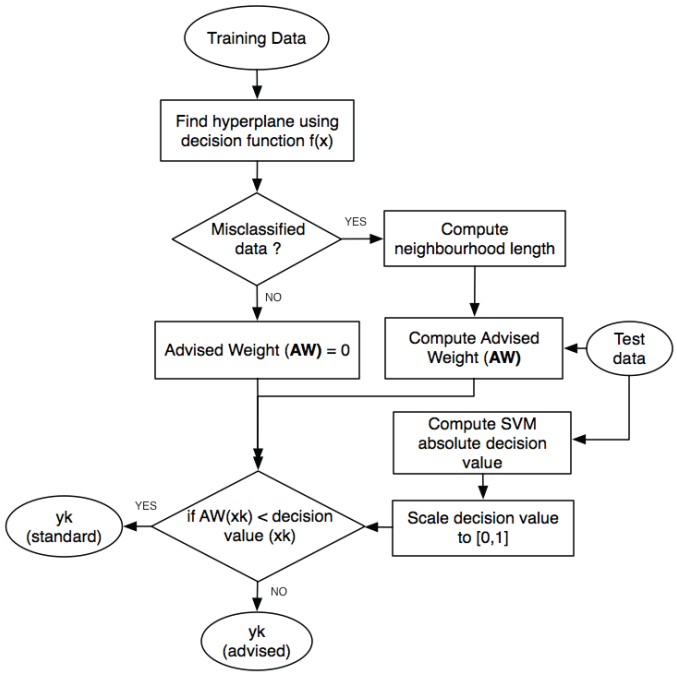


Figure 6.3 Advised Weight Support Vector Machine

6.2.3 Analysis of proposed classification model and discussion

The analysis section comprises of two main part; 1) Performance of each feature set in distinguishing cancer vs. non-cancer is presented, that is further used for calculating the expertise weight. 2) Analysis of proposed expertise weighted parallelized classification model and demonstration of 10 folds cross validation of the diagnosis results of proposed model. 3) Finally, average classification performance of the proposed model is compared with those of other popular classification methods used for skin cancer diagnosis.

- **6.2.3.1 Calculation of Expertise Weight (EW)**

The features details presented in chapter 5 are suggested here for feeding into the parallel classifier model. The constructed feature sets 1 to 5 are used for feeding the classifier. (Figure 6.4) The average benign and melanoma identification capability of each feature extraction method and the proposed set of features can be seen from Figure 6.5.

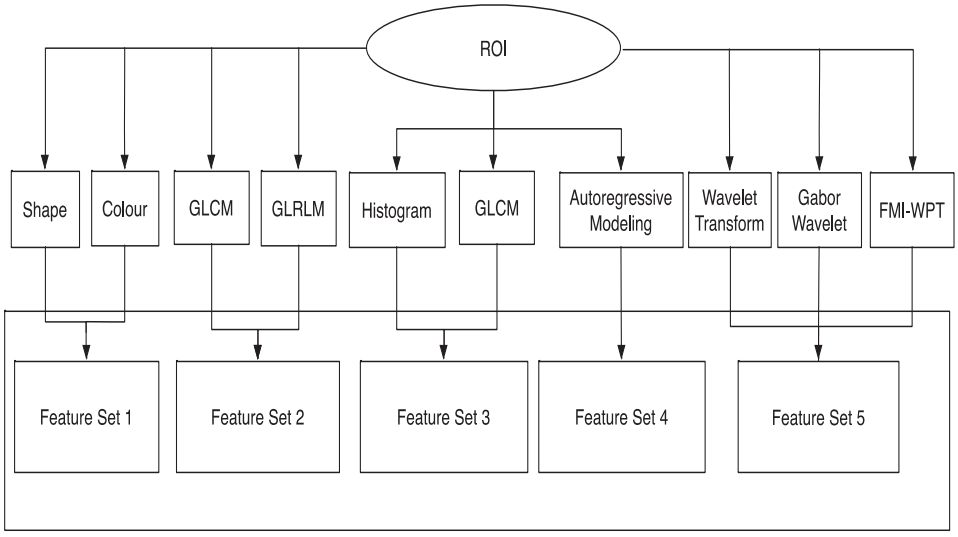


Figure 6.4 Feature extraction sets used in experiments

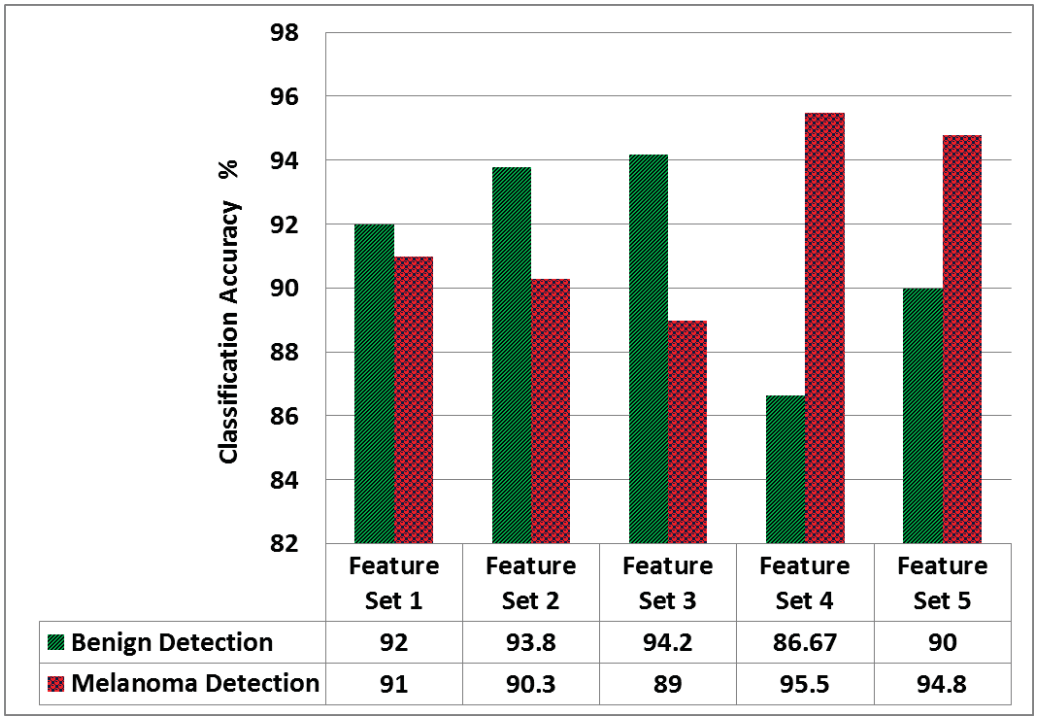


Figure 6.5 Diagnosis expertise analysis of suggested feature sets

As it can be noticed clearly that use of feature set 1 (shape and Colour) provided close performance for benign and melanoma. However, feature set 2 comprised of GLCM and GLRLM and feature set 3 comprised of GTDM and histogram based features resulted in better Benign Detection i.e. higher specificity $\left(\frac{\text{True Negative}}{\text{True Negative+False Positive}}\right)$ but lower melanoma detection i.e. sensitivity $\left(\frac{\text{True Positive}}{\text{True Positive+False Negative}}\right)$. It can also be noted the proposed feature set 2 (combination of GLCM and GLRLM) provided the best specificity among all the other combinations. That is why the classifier based on these features will get more expertise weight (EW) if it declares the lesion as benign as compared to if it declares the lesion as melanoma. On the other hand, use of Feature set 4 comprised of autoregressive modelling based features provided the highest sensitivity among all the feature groups however its quite low in specificity, similarly feature set 5 comprised of Wavelet transform, Gabor wavelet, and FMI-WPT based features lead to good sensitivity but relatively lower specificity as well. That is the main reason that the classifier based on these features will get more expertise weight if it declares the lesion as melanoma than as if declares the lesion as benign. Note that the expertise weight is based on the

experimental validation of the efficiency of the feature set in finding cancer or non-cancer.

In order to calculate the expertise weight for each feature set, each feature set based classifier is trained by a training dataset (155 Benign and 155 melanoma lesions) and is verified by the validation samples containing different lesion data set (155 Benign and 155 melanoma lesions). If the total specificity of the validation samples of classifier k is denoted by SP_n, weight of that classifier for detecting benign can be calculated as follows. (6.15)

$$EW_{n\_b} = {SP_n}\Big/{\sum_{n=1}^{p} SP_n} \tag{6.15}$$

where p is the number of classifier based on each feature set, here p=5 as five different feature sets are used. Similarly, if the total sensitivity of the validation samples of classifier k is denoted by SE_n, weight of that classifier for detecting melanoma can be calculated as follows. (6.16)

$$EW_{n\_m} = {SE_n}\Big/{\sum_{n=1}^{p} SE_n} \tag{6.16}$$

Important observation made is that because of wide variety of skin cancer lesions different feature sets classified the lesions in different ways. Some of the features were more prominent in the benign lesions so the classification based on those features was successful but the same classifier struggled to classify melanoma when used the same set of features and vice versa. For example, the features derived on the basis of intensity variations were more efficient in detecting benign lesions and these lesions have more clear intensity differences between skin and surrounding area. On the other hand, the wavelet and autoregressive modelling based features were more successful in the classification of melanoma images as precise details within the lesion can be analysed quantitatively.

The observations showed that the generalization capability of classifier decreases with the increasing number of features due to the curse of dimensionality factor. Thus, it was not reasonable to use 281 features in one go. On the other hand, as all these features covered different details, cutting down the number of features just to avoid dimensionality issues is also not a reasonable choice. So, a parallelized classifier model is suggested that makes use of all the features independently and efficiently.

- **6.2.3.2 Analysis of parallelized classification model**

The block diagram of parallelized classification model is shown in Figure 6.6. Note that training dataset used here is different from the training and validation dataset used for expertise weight calculation. This will help in analysing the generalization performance of the overall model without being over fitted.

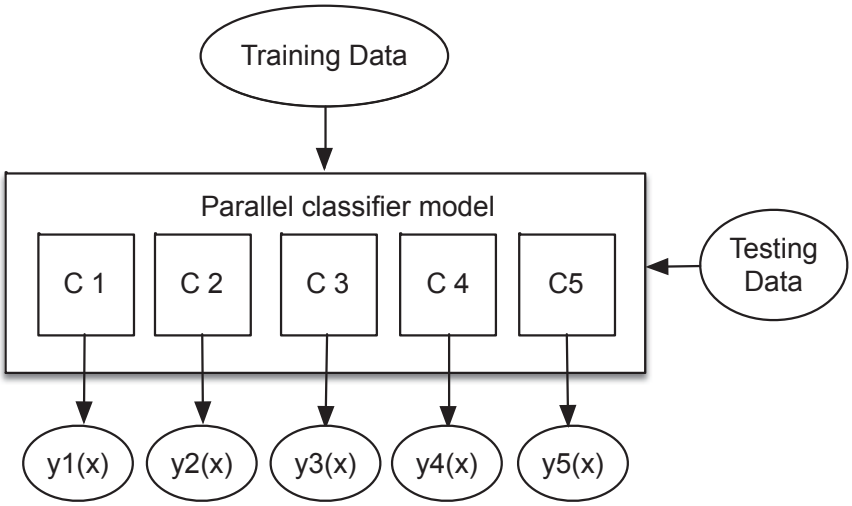


Figure 6.6 Parallelized Classification model

Once the suggested output of each classifier is obtained for the test data, the final diagnosis of each classifier Gn(x) while taking into consideration the expertise weight, as follows Figure 6.7.

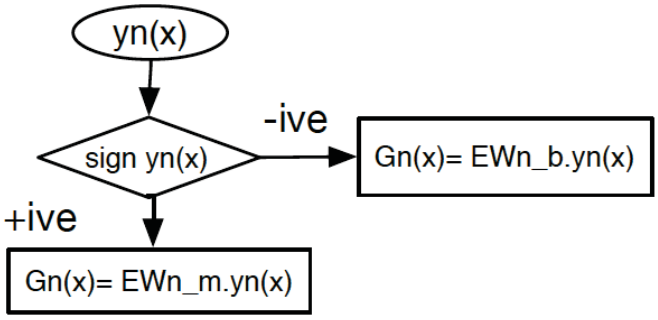


Figure 6.7 Diagnosis results based on Expertise weight

Finally, the overall diagnosis results G(x) is obtained by aggregating the diagnosis results of each independent classifier using the following relation (6.17), where p=5, is the number of individual classifiers in the proposed model.

$$G(X) = \sum_{n=1}^{p} G_n(x) \tag{6.17}$$

Figure 6.8 shows the classification results of the overall proposed model using 10 fold -cross validation. It can be seen that the diagnostic model achieved an average

specificity of 94.1% and average sensitivity of 94.8 % with max specificity in one set goes up to 94.7% and maximum sensitivity reaching 95.6%.

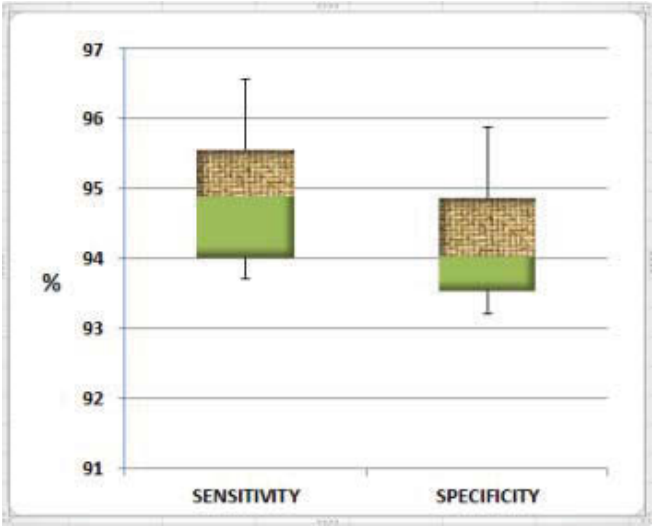


Figure 6.8 10-fold cross validation results for the proposed model with y-axis representing % sensitivity and % specificity (green parts shows sensitivity and specificity variation below average and the yellow pattern shows the above variation range

- **6.2.3.3 Experimental Results for Parallelized Classification Model**

In order to justify the performance of the proposed method, classification results of the proposed method are also compared with the diagnosis based on some of the most popular methods used in literature so far for skin cancer diagnosis [70, 168, 169, 322]. It include individual models using KNN, Logistic regression, Artificial neural network (ANN), C4.5, standard with k set to 2, C4.5 and standard support vector machine and ensemble models of Standard SVM and Advised weighted SVM using majority voting and EW-based weighted averaging. In this study, k for KNN set to 2, for C4.5 (a popular decision tree builder) standard pruning factors are used with confidence factor of 0.25. A two layer feed forward network with sigmoid activation function is designed with 22 input neurons, 20 hidden neurons and 2 output neurons and scaled conjugate gradient algorithm is used for training this artificial neural network (ANN). For the standard SVM model, Gaussian kernel function is used and penalty parameter C and Kernel parameters are determined using grid search method.

Experimental data set used for comparative analysis was comprised of 500 lesions (250 melanoma and 250 benign). The classification task was to distinguish between melanoma and benign. Three commonly used evaluation criteria are used: Sensitivity (True positive / (True positive and false negative)), Specificity (True negative/ (True

negative + False Positive)) and total accuracy ((True positive+ True negative)/Total number of lesions). Classification results in Table 6.1 clearly indicate that the proposed model provided improved performance in differentiating melanoma and benign lesions.

Table 6.1. Comparative Analysis of Classification results

Category	Model	Aggregation Strategy	Specificity	Sensitivity	Accuracy
Single	LogR		83.8	87	85.5
	KNN		83.2	85.1	84.2
	ANN		89.1	90.3	89.6
	C4.5		82.6	85.2	83.4
	SVM		90.9	89.6	90.3
	A-W-SVM		91.6	92.2	91.9
Ensemble	SVM	Majority Voting	91.4	92.0	91.7
	A-W-SVM		92.3	92.9	92.6
	SVM	**Expertise based weight**	91.9	92.5	92.2
	A-W-SVM		**94.1**	**94.8**	**94.4**

The above-mentioned system worked well for the digital dermoscopic images, but the performance for histopathalogical images was not good enough due to the limited availability of training data and the complexity of the images, therefore in the following section a more advanced learning model is proposed that can use both labelled and unlabelled data to acquire sufficient amount of information that can help in better classification performance.

6.3 Semi Supervised learning model for Classification

The need to use learning techniques in any area in order to improve performance requires a proper choice of the learning algorithm and of their statistical validation. Computer aided skin cancer diagnostic is a difficult problem given the relative paucity of labeled lesion data and consequently the low quality of training data available. Secondly, a trade-off between model complexity and generalization ability of the final classification model are some areas that are lacking attention in the research area of automated skin cancer diagnosis.

Classification using insufficient number of labeled data is a common hard problem [470, 472]. Unfortunately, this is also a case in developing skin cancer diagnosis models as obtaining the labeled data from experts is not trivial, cost money or is time consuming. To solve this issue, semi supervised learning is proposed, which uses large amount of unlabeled data together with labeled data to build better learners, is an important aspect needing consideration. Typical semi-supervised methods used for different applications include Expectation-maximization (EM) algorithm with generative mixture models [473], Transductive support vector machines[474] , graph-based methods [475], and co-training [476] etc.

Most of semi-supervised techniques use shallow architecture in order to model the problem [470, 477, 478]. However, for several applications, work done in machine learning has highlighted the circumstances, which favour deep learning architectures, such as multilayer neural nets, over shallow architectures, such as support vector machines [479, 480]. This is because deep learning architecture, composed of multiple levels of non-linear operations, is expected to perform better in semi supervised learning. This is because of its ability to better model hard artificial intelligence tasks [481].

However, deep learning architectures are generally more difficult to train as compared to the shallow ones. They involve difficult nonlinear optimizations and many heuristics. The challenges of deep learning explain the early and continued appeal of SVMs, which learn nonlinear classifiers via the kernel trick. Unlike deep learning architectures, SVMs are often trained by solving a simple problem in quadratic programming. But, machine learning is a giant interconnected web of ideas and as the field continues, researchers working to generalize across models and finding new relationships in disparate concepts and models [482-484].

Based on the study of semi-supervised learning and deep learning architecture, a novel semi supervised classification model that combines the benefits of deep belief networks and support vector machine with a self-advising capability. It addresses the skin lesion classification problem with unlabeled data by making efficient use of limited labeled data. In addition to that, bootstrapping [485] is being used to increase the generalization performance of the proposed learning model and to prevent over fitting.

The proposed classification model is shown in Figure 6.9. Usually a single training set $\{(x_i,y_i)|i= 1,2,3,\ldots L\}$, where L is the number of data samples, is used if a

single classifier is being used. But in proposed model, 3 training samples sets are needed to construct the model with 3 independent classifiers as the proposed classification model is shown in figure 6.9. It includes deep belief net, Semi advised support vector machine with RBF (radial basis function) kernel and with polynomial kernel. From the statistical point, it is needed to make the training sample sets different as much as possible in order to obtain higher improvement of the aggregation result. For doing this, the bootstrap technique is used to build 3 replicate training data sets by randomly resampling, along with replacement, from the given overall training data set repeatedly. Each replicate training set is used to train one candidate classifier. This technique is aimed to reduce over fitting of the overall model.

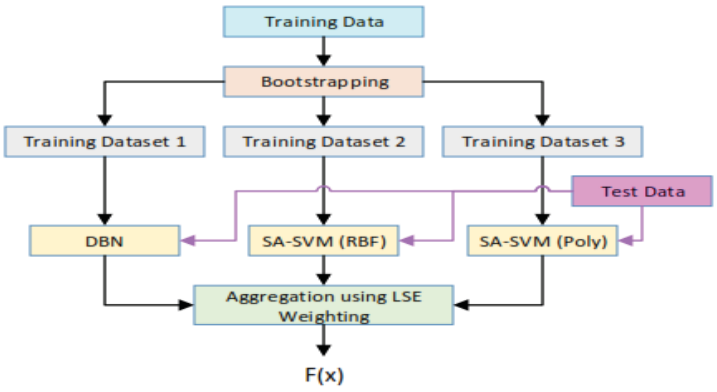


Figure 6.9 Ensemble Learning Model

For aggregating the classifiers least square error based weighting is used. The weights of the classifier in the overall model are determined in proportional to their accuracies of classification [485]. Let f_k (k= 1,2,3) be a decision function of the kth candidate that is being trained by a replicate training data set, here k=3 is the number of classifiers used. The weighing vector w was obtained by w= A-1 y , where A= $(f_i(x_j))$ kxL , and y= (y_i)1xL. Then the final decision of the classification model $F_{final}(x)$ for a given test vector x is given as

$$F_{final}(x) = \text{sign}\ (w.\ [f_i(x)kxL]) \qquad (6.18)$$

The theoretical detail of the classifiers used is as follows.

6.3.1 Semi Advised-SVM

SVM is a well-known machine learning method proposed by Vapnik [350] and is used for various applications in machine learning. The basic idea of SVM is to construct

a maximized separating hyper plane that can separate data in the feature space. The classic SVM usually ignores the training data that has not been separated linearly by the kernels during the training phase. Therefore, if data that is similar or identical to this misclassified data appears in the test set, it will be classified wrongly by the classifier. This misclassification is not reasonable and it can be handled if the available data and information has not been ignored by SVM algorithm in the training phase.

In this chapter, a non-iterative self-advising approach [471] for SVM is adapted that extracts subsequent knowledge from training phase and does not ignore misclassified data. The misclassified data can usually come from two potential sources 1) outliers 2) data that have not been linearly separated by using any of the kernels. SA-SVM deals with the ignorance of SVM from the knowledge that can be acquired from misclassified data by generating advice weights based on using the misclassified training data, and use these weights together with decision values of SVM in the test phase. These weights also help the algorithm in eliminating the outlier data as well. The details of SA-SVM algorithm are as follows:

Given the dataset is divided into two subsets: a labelled data set $D_L = \{(x_i, y_i)\}_{i=1}^{L}$ and an unlabelled data set $D_{UL} = \{(x_i)\}_{i=L+1}^{L+UL}$

Step 1. The unlabelled data set D_{UL} is equally divided into n subsets D_{UL_1}, D_{UL_2}, ... , D_{UL_n} , then D_L is taken as the initial training set T_s and initialize i = 1 , where i denotes the ith loop of the algorithm.

Step 2: Normal SVM classifier is trained using the labelled data set and the classifying hyperplane is found by using decision function $f(x) = \text{sign}(\sum_{\alpha_{i>0}} y_i \alpha_i k(x, x_i) + b)$, where xi is the input vector corresponding to the ith sample and is labelled by yi depending on its class and αi is the nonnegative Lagrange multiplier that is inconsistence with standard SVM training.

As the data is comprised of nonlinearly separable cases so SVM to produce non-linear decision functions and radial basis function kernel $K(x_i x_j) = e^{-\gamma |x_i - x_j|^2}$ or polynomial kernel is used to make all necessary operations in the input space.

Step 3: The misclassified data sets (MD) in the training phase is determined using following relationship. (6.19)

$$MD = \bigcup_{i=1}^{N} x_i \; | y_i \neq \text{sign}(\sum_{\alpha_j>0} y_i \; \alpha_j \, k(x_i, x_j) + b) \tag{6.19}$$

The MD set can be null, but most experiments showed that the occurrence of misclassified data in training phase is a common occurrence. It must also be noted that any method that tries to get benefit from misclassified data, must also have some control on the impact of outlier data. It was observed that when the misclassified data is comprised of resembling samples, the use of misclassified data actually improved the classification accuracy more as it can lead to the variations required in the final separating hyperplane.

If the MD is null, go to the next step, else compute neighbourhood length (NL) for each member of MD using the following mathematical relation. (6.20)

$$\mathrm{NL}(x_i) = \mathrm{minimum}_{x_j} \left(\left\| x_i - x_j \right\| \mid y_i \neq y_j \right) \tag{6.20}$$

Where x_j , j=1, …., N are the training data that do not belong to the MD set. Here as the training data is mapped to a higher dimension, the distance between x_i and x_j is computed according to the following equation with reference to the related RBF kernel.

$$\left\| \theta(x_i) - \theta(x_j) \right\| = \left(k(x_i, x_i) + k(x_j, x_j) - 2k((x_i, x_j)) \right)^{0.5} \tag{6.21}$$

Step 4. The labels for data samples in DUL_1 are estimated using the current classifier, then the most confidently classified elements are determined according to the distance between the element and the separating boundary. The criteria is formulated as $|x \cdot w - b\} \geq Th$, where constant $Th > 0$ is the distance threshold. If the distance between the element and the separating boundary is larger than Th, it is taken as a confident element. The most confidently classified elements with their predicted labels are represented as set R and are added together with their predicted labels, to the training set T_s, that is, $T_s = T_s \cup R$. The remaining elements of D_{UL_1} are denoted as unlabelled query UL_Qi.

For each sample x_k from the unlabelled Query set UL_Qi advised weight $AW(x_k)$ is computed using following mathematical relationship. These AWs represent how close the data is to the misclassified data from the labelled set.

$$\begin{cases} 0 & \forall x_i \in \mathrm{MD}, \left\| x_k - x_i \right\| > \mathrm{NL}(x_i) \text{ or } \mathrm{MD} = \mathrm{NUL}, \\ \sum 1 - \frac{\sum_{x_i} \left\| x_k - x_i \right\|}{\sum_{x_i} \mathrm{NL}(x_i)} & x_i \in \mathrm{MD}, \left\| x_k - x_i \right\| \leq \mathrm{NL}(x_i) \end{cases} \tag{6.22}$$

The absolute value of the SVM decision values for each x_k from the unlabelled Query set set are calculated and scaled to [0, 1]. For each x_k from the unlabelled Query

set set, If (AW (x_k) < decision value (x_k) then $y_k = \text{sign}\left(\sum_{\alpha_j>0} y_j \alpha_j\, k(x_k, x_j) + b\right)$ which is in consistence with normal SVM labelling, otherwise $y_k = y_i \mid (\|x_k - x_i\| \leq \text{NL}(x_i) \text{and } x_i \in \text{MD})$. After getting the labels of UL_Qi add UL_Qi with predicted labels to T, that is, T = T ∪ Qi.

Additionally, since the size of the training set is enlarged during the training procedure, the penalty parameter C of the semi-advised SVM should adapt to this change. Thus, the empirical formula for C introduced in [486] is used as: $C_i = \frac{(1-\lambda)}{\lambda(TR_i+UL_i)}$, where i denotes the i^{th} loop. TR_i is the size of training set of the ith loop. UL_i is the size of unlabeled data set. The value of λ is taken as 0.01 in our experimental data analysis.

Step 5. i = i + 1

Step 6. If i equals n, terminate; otherwise, go back to Step 2.

This method allows efficient use of labelled data along with consolidating the unlabelled data in smaller chunks in the training process. In this way the algorithm can learn better by updating the separation hyperplane in the light of the information coming from unlabelled data.

6.3.2 Deep Belief Network

Deep belief network is used to aid the semi supervised SVM in dealing with the unlabelled data. Deep belief net has the capability to learn undefined features in terms of weights settings and the training of the proposed deep belief net uses both labelled and unlabelled data. The mathematical explanation of how it is done is as follows.

Let X be a set of samples given as X= [x1, x2, …. xS] where x is a sample data, and the whole set X contains S samples including SL labelled samples and S_{UL} unlabelled samples, F is the number of features for each sample.

Let Y be a label matrix given as Y= [y1, y2, …. , y^{S_L}] that corresponds to S_L labelled samples. Where y is a label vector for a sample data and total number of classes is C, which is two in our case.

The component yij can be defined as follows. (6.23)

$$y^i_j = \begin{cases} 1 \text{ if } x^i \in j^{th} \text{ class} \\ -1 \text{ if } x^i \notin j^{th} \text{ class} \end{cases} \quad (6.23)$$

Where j is the index for vector y_i , the size of y_i is the same as the number of classes of the input data.

In order to obtain a mapping relation X → Y using the S_L labelled data samples and S_{UL} unlabelled data samples a semi supervised learning algorithm is developed using deep architecture techniques. This can provide a respective y when a new sample x comes. The deep architecture consists of a fully interconnected directed belief nets with one input layer L_0, N hidden layers L_1, L_2, … L_N, and one label layer at the top. The input layer L_0 has F units, equal to the number of features of sample data x. The label layer has C units, equal to the number of classes of label vector y. There number of units for hidden layers is predefined in this case according to experience or intuition. The seeking of the mapping function X → Y, here is transformed to the problem of finding the parameter space W= $\{w_1, w_2, \dots w_N\}$ for the deep architecture.

The training of the deep belief net can be divided into two stages:

(1) Deep belief net is constructed by greedy layer-wise unsupervised learning using SUL unlabelled data samples together with S_L labelled data samples.

(2) The parameter space of W obtained through unsupervised learning model is further fine-tuned using the exponential loss function to maximize the separability among S_{UL} labelled data.

The mathematical details of both stages are explained below.

- **Layer-Wise Unsupervised Learning**

The overall deep belief net is constructed layer by layer using Restricted Boltzmann Machines (RBM) as building blocks. RBM is a two-layer recurrent neural network in which stochastic visible binary inputs are connected to stochastic invisible binary outputs using symmetrically weighted connections. The details can be found in [487] [480].

The energy of the state (L^v, L^h) is given as follows. (6.24)

$$E(L^v, L^h; \theta) = -\sum_{q=1}^{F_v}\sum_{r=1}^{F_h} {w_{qr}}^h {L^v}_q {L^h}_r - \sum_{q=1}^{F_v} {b^v}_q {L^v}_q - \sum_{r=1}^{F_h} {a^h}_r {L^h}_r \quad (6.24)$$

Where $\theta = (w, b, a)$ are the model parameters, ${w_{qr}}^h$ is the symmetric interaction term between unit q in the visible layer L^v and unit r in the hidden layer L^h. ${b^v}_q$ is the qth bias of visible layer L^v and similarly, ${a^h}_r$ is the rth bias of hidden layer L^h. The probability assigned to the visible vector L^v is given as follows. (6.25)

$$P(L^v;\ \theta) = \frac{1}{Z(\theta)} \sum_{L^v} e^{-E(L^v, L^h, \theta)} \tag{6.25}$$

$$Z(\theta) = \sum_{L^v} \sum_{L^h} e^{-E(L^v, L^h, \theta)} \tag{6.26}$$

Where $Z(\theta)$ denotes the normalizing constant. The conditional distributions over L^v and L^h are given as follows.

$$p\left(L^h \middle| L^v\right) = \prod_r p\left(L_r^h \middle| L^v\right) \tag{6.27}$$

$$p\left(L^v \middle| L^h\right) = \prod_q p\left(L_q^v \middle| L^h\right) \tag{6.28}$$

The probability of tuning on unit q is a logistic function of the states of L^v and ${w_{qr}}^h$

$$p\left(L_r^h = 1 \middle| L^v\right) = \text{sigm}({a^h}_r + \sum_s {w_{qr}}^h L^v) \tag{6.29}$$

Where the logistic function is:

$$\text{sigm}(x) = \frac{1}{1+ e^{-x}} \tag{6.30}$$

The derivative of the log-likelihood with respect to the model parameter which can be obtained using Contrastive Divergence method proposed by Hinton [488].

$$\frac{\partial \log p(L^v)}{\partial {w_{qr}}^h} = \langle L^v L^h \rangle_{P_o} - \langle L^v L^h \rangle_{P_T} \tag{6.31}$$

Where $\langle . \rangle_{P_o}$ denotes an expectation with respect to the data distribution and $\langle . \rangle_{P_T}$ denotes a distribution of samples from running the Gibbs sampler initialized at the data, for T full steps. Setting T= ∞ recovers maximum likelihood learning, although T is typically set to 1.

Based on wh, the layer L^h can be obtained as follows with the sample input x from layer L0:

$$L_r^h(x) = \text{sigm}\left(\sum_{q=1}^{F_v} {w_{qr}}^h {L^v}_q(x) + {a^h}_r\right) \quad r= 1,\ldots F_h \tag{6.32}$$

The above discussion is based on one sample data x. In the deep belief network, used here, the deep architecture is constructed using all labelled data with unlabelled data by putting them one by one from layer L_0.

- **Fine-tuning Supervised Learning**

After unsupervised learning, the deep network is obtained with initial parameter space W. This section explains the process used to fine-tune W to maximize the separability

among SL labelled data. A loss function is used as following to make the data points far from each other if they are in the different classes:

$$f(L^N(x), y) = \frac{1}{C} \sum_{j=1}^{C} e^{-L_j^N(x) y_j} \quad (6.33)$$

It is intended to minimize the loss of the whole labelled dataset by solving the optimization problem below:

$$\arg_w \min F = \sum_{i=1}^{S_L} f(L^N(x^i), y^i) \quad (6.34)$$

As the samples come from two categories and the top hidden layer has two units, i.e. C = 2 and DN = 2. The discriminability of deep architecture is not powerful for the data distribution before supervised learning [480]. To minimize F, computed gradient, and use back propagation through the whole DDBN until satisfying the pre-defined requirement. DDBN constructs a wide margin between labelled samples from two classes after supervised fine-tuning, which can help in better classification.

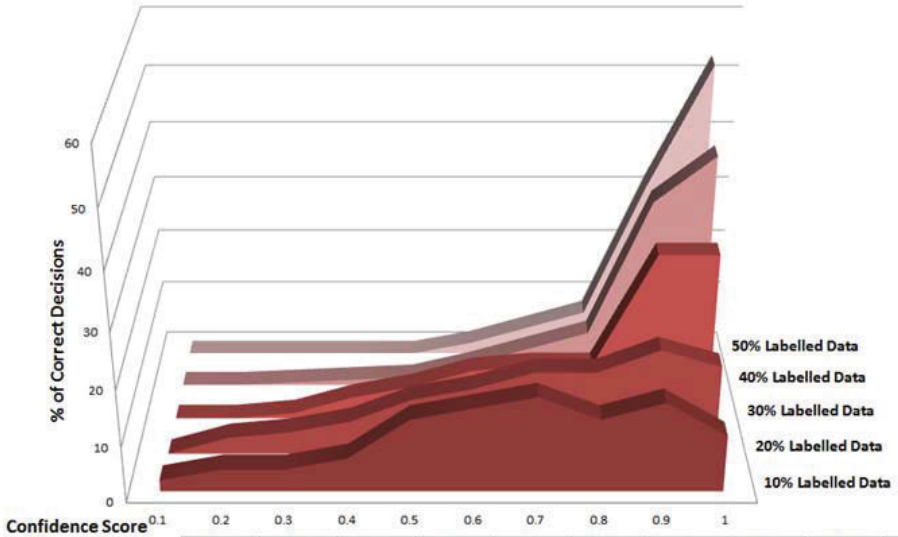
6.3.3 Experimental Results for Semi Supervised Learning Model

Two datasets were used in the experiments, dataset 1 is based on dermoscopic or clinical view lesion images and dataset 2 is based on histopathological images obtained from the biopsy samples of skin cancer patients. Most of the images in the datasets came from Sydney Melanoma Diagnostic Centre, Royal Prince Alfred Hospital. Dataset one comprise of 300 labelled images and 500 unlabelled images. While the Dataset 2 consists of 160 images including 60 labelled and 100 unlabelled samples.

Initially just the proposed semi advised support vector machine is used for the classification of both datasets. 10 fold cross validation rule is used for dividing the data into training and testing sets to validate the performance of the overall model for both datasets. In order to extract more information than discrete class membership alone, the metric of confidence score is used. The confidence score for the classifier is obtained by calculating the normalized maximum probabilities corresponding to the selected class. Figure 6.10 shows the distribution of confidence scores (horizontal axis) with vertical axis representing percentage of total correct decisions for the detection of melanoma lesions during testing phase based on dataset 1. Figure 6.11 shows the distribution of confidence scores (horizontal axis) with vertical axis representing percentage of total correct decisions for the detection of lesions during testing phase based on dataset 1. While figure 6.12 shows the distribution of confidence scores (horizontal axis) with vertical axis representing percentage of total correct decisions for the identification of

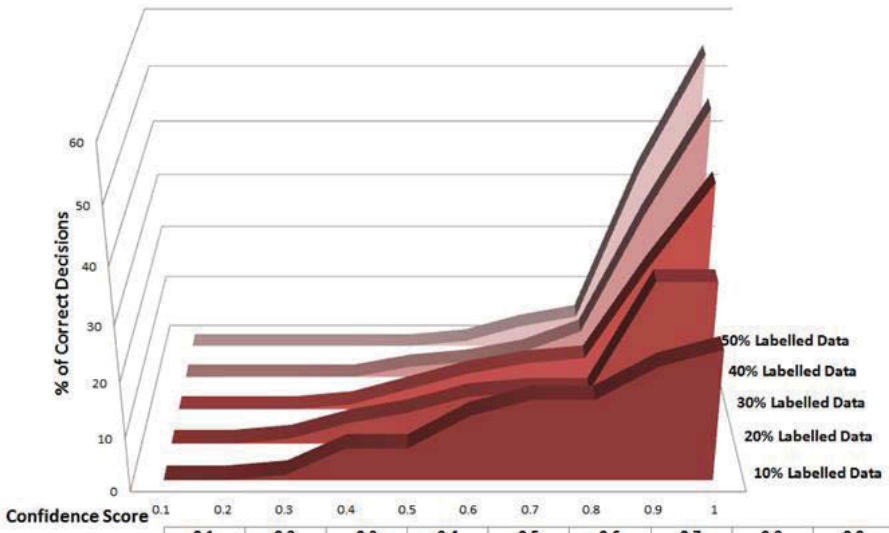
cancerous regions in the corresponding histopathological image during testing phase based on dataset 2.

The distribution characteristics of a confidence metric may differ with application and the type of classifier. However, for the present case as a semi supervised classifier is used, confidence score was strongly indicative of the accuracy. For example, a decision with 50% confidence would have 50% chances of being correct. It can be observed that for dataset1 85% of the melanoma lesion detections were made with within confidence score range of 0.9-1 and 89% of the benign lesion detection were made within confidence score range of 0.9-1. On the other hand 83% of the cancerous regions detections were made correctly within the confidence range of 0.7-1, which is quite a significant achievement keeping an eye on the complexity of details present in the histopathological images.



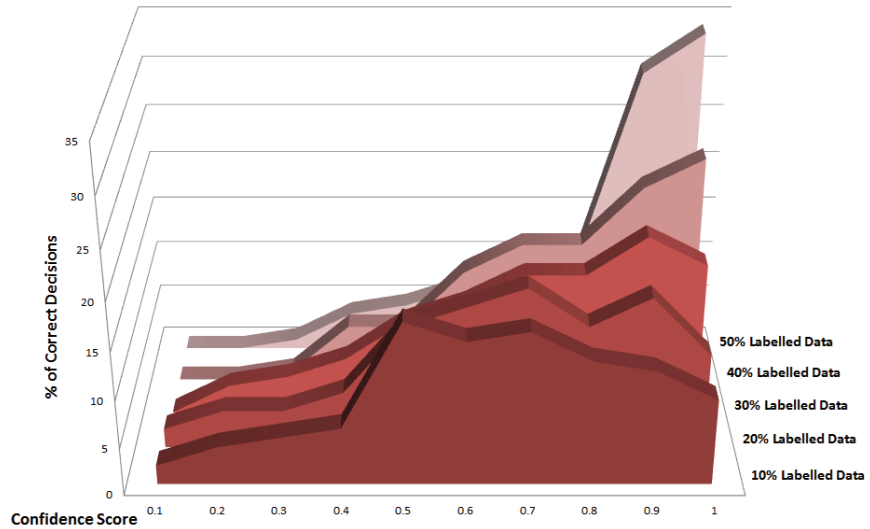
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
10% Labelled Data	2	4	4	6	13	15	17	13	16	10
20% Labelled Data	0	3	4	6	10	12	15	15	19	16
30% Labelled Data	0	0	1	4	6	9	10	10	30	30
40% Labelled Data	0	0	0.5	1	1.5	4	7	10	34	42
50% Labelled Data	0	0	0	0	0	2	5	8	32	53

Figure 6.10. Confidence score distribution Vs total correct decisions % for melanoma lesion identification



	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
10% Labelled Data	0	0	1	6	6	12	15	15	21	24
20% Labelled Data	0	0	1	4	6	9	10	10	30	30
30% Labelled Data	0	0	0	1	4	7	9	10	27	42
40% Labelled Data	0	0	0	0	2	3	5	9	31	50
50% Labelled Data	0	0	0	0	0	1	4	6	34	55

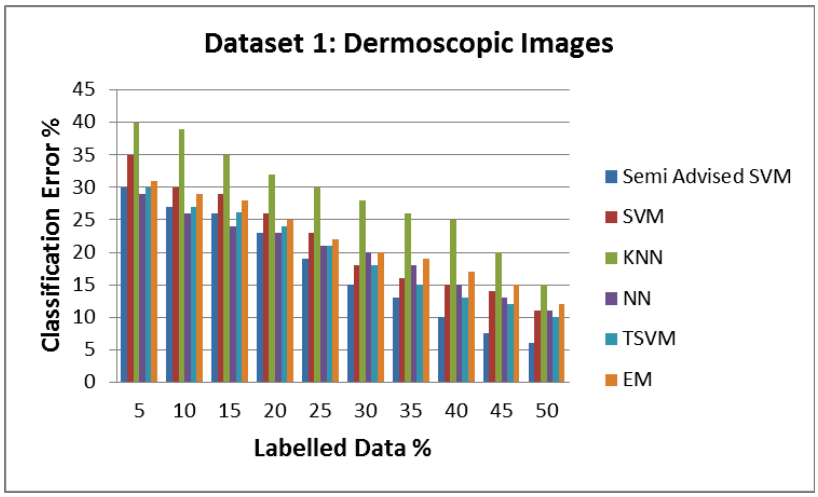
Figure 6.11. Confidence score distribution Vs total correct decisions % for benign lesion identification



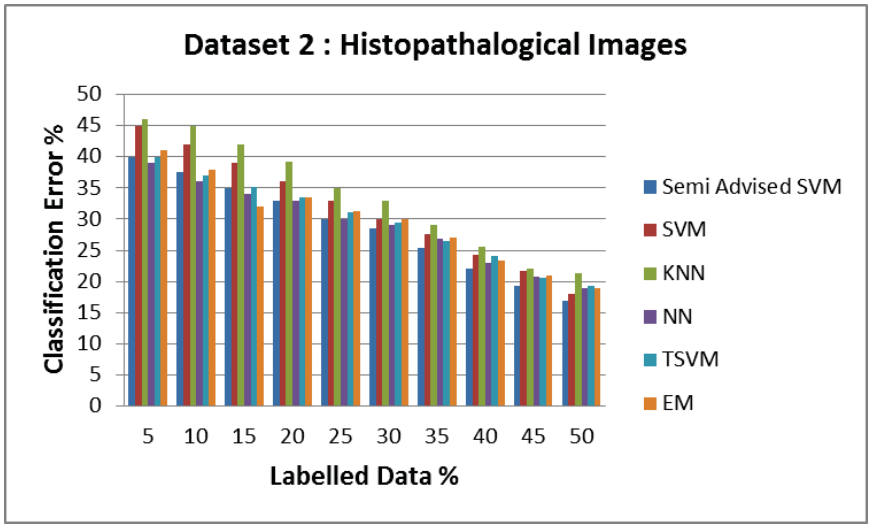
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
10% Labelled Data	2	4	5	6	17	15	16	13	12	9
20% Labelled Data	2	4	4	6	13	15	17	13	16	10
30% Labelled Data	0	3	4	6	10	12	15	15	19	16
40% Labelled Data	0	0	1	6	6	12	15	15	21	24
50% Labelled Data	0	0	1	4	5	7	9	10	30	34

Figure 6.12. Confidence score distribution Vs total correct decisions % for cancerous region identification in histopathological images

The classification performance of the semi advised SVM based classification algorithm is compared with that of the normal SVM, artificial neural network, KNN, Transductive SVM and Expectation-maximization (EM). Figure 6.13, 6.14 shows the average classification error rate of different classifiers with respect to the change in the ratio of labeled and unlabeled data samples used for training phase. In consistent with a lot of finding in different other applications [473, 480][489] it was observed that by increasing the number of labeled data in the training phase helps in reducing the classification error. However, it can be seen that the proposed learning model make the most efficient use of unlabeled data along with the limited labeled data as compared to other methods. The classification error reduced to around 16.5 % when the learning model used 50% of labeled and 50 % of unlabeled data for Histopathological images (with selected number of features increased to 30) which is quite reasonable keeping an eye on the complexity of histopathological image analysis. While the classification error was reduced to 6% for Dermoscopic images (with selected number of features kept to 25) when the learning model used 50% of labeled and 50 % of unlabeled data.



(a)



(b)

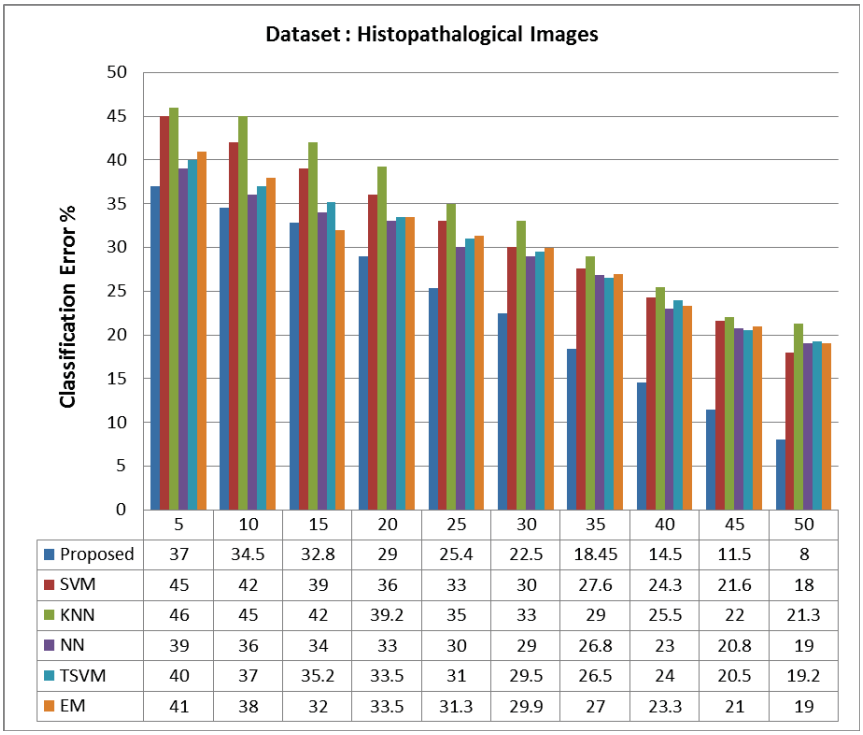
Figure 6.13 Performance comparison of semi advised SVM for dermoscopic and Histopathalogical image

For dermoscopic images classification accuracy was around 94% which is quite close to the one obtained using knowledge based learning model proposed above where the training phase was comprised of just labeled dataset, and it shows that the proposed semi advised SVM can achieve approximately same classification accuracy with using just 50% of the labeled data. On the other hand, for histopathological images the classification accuracy was around 84.5% which is not very high as compared to that achieved for dermoscopic images. Therefore, the ensemble learning model presented in Figure 6.9 is proposed to use for more complex datasets like that of histopathalogical images to improve the classification performance.

For experimental analysis of this model for histopathalogical images, 100 data samples were selected for the training set, out of which 50 were unlabeled data samples and rest 50 were labeled. For bootstrapping, the data was randomly resampled with replacement from the training data set. Each member classifier was trained independently over the replicated training data set and aggregated them via LSE weighting. Bootstrapping can counteract the issues of limited data availability to some extent in the cases like histopathological image datasets for skin cancer where limited number of data is available that is labeled with the help of expert dermatologists.

The classification model was tested using test data set consisting of 50 data samples. Out of which 25 were originally diagnosed as melanoma and 25 were benign. The comparative average classification results obtained through the proposed model

with increasing number of labeled data percentage is presented in Figure 6.14. The classification accuracy of around 92% is achieved with the help of proposed model for histopathalogical images which is significantly high performance for classification of histopathalogical images.



	5	10	15	20	25	30	35	40	45	50
Proposed	37	34.5	32.8	29	25.4	22.5	18.45	14.5	11.5	8
SVM	45	42	39	36	33	30	27.6	24.3	21.6	18
KNN	46	45	42	39.2	35	33	29	25.5	22	21.3
NN	39	36	34	33	30	29	26.8	23	20.8	19
TSVM	40	37	35.2	33.5	31	29.5	26.5	24	20.5	19.2
EM	41	38	32	33.5	31.3	29.9	27	23.3	21	19

Figure 6.14 Performance comparison of learning model proposed for histopathalogical image classification

6.4 Summary

This chapter presents the details of the research carried out related to the classification phase of the automated diagnostic system for skin cancer. An expert weight generation system is introduced based on the differentiating capabilities of different feature sets. And a parallelized classification model based on advised SVM aided with expert weight generation system is proposed for differentiating melanoma and benign lesions. For dealing with the issue of limited labeled data availability specially for histopathalogical images, a novel learning model based on deep belief neural net and semi advised SVM is proposed which can make efficient use of labeled data along with unlabeled data for the

training phase. The proposed diagnostic model showed reasonably improved performance when compared with several state of the art methods that are used so far for skin cancer diagnosis. The proposed model is tested for skin lesion classification but it can be applied on other applications as well where limited amount of labeled data is available.

The algorithms proposed in this thesis are aimed at contributing to the development of efficient automated diagnostic tools for skin cancer, to help GPs in early diagnosis of skin cancer and provide a second opinion to experts during biopsy decision making. Now to what extent the combination of human and machine-based diagnoses would affect the decision-making process and help the doctors by improving the detection of early melanoma and/or decreasing unnecessary surgery depends on detailed well-designed, randomized clinical trials in the field.

Chapter 7

Conclusion and Future Research

7.1. Introduction

This thesis has established algorithms for developing automated system for skin cancer detection and diagnosis based on soft computing approaches.

The algorithms developed in this thesis are related to four main phases of computer aided diagnostic system, i.e. image segmentation, recommended feature generation, dimensionality reduction/feature selection, and machine learning methods that have been used for classification or cancerous area recognition. Many experiments have been conducted and several algorithms are developed that can be used in skin cancer studies or some can even be used in wider range of applications. The conclusions of these experiments and their results are summarised as follows.

7.2 Conclusion

The thesis is commenced with providing the medical background for skin cancer diagnosis to set some foundations for automated diagnostic system. Afterwards a detailed review and analysis is provided for the existing works presented in literature for the development of automated diagnostic systems for skin cancer.

The main contribution of the thesis in the detection phase is the development of novel methods for segmentation of digital/dermoscopic and histopathological images for

skin cancer. The proposed method achieved an average True Detection Rate of around 93% for digital/dermoscopic images and approximately 88% for histopathological images, which is a significantly improved performance when compared with some the most frequently state-of-the-art used in the area for this application.

After segmenting the images, features are generated based on physical, statistical as well as spectral analysis of the skin images. The details of the suggested feature extraction methods and their relevant significance is provided in the thesis with respect to their differentiating capability for melanoma and benign skin images.

In the feature selection stage, a new adaptive differential evolution based feature selection method is proposed. The method was applied to the skin cancer dataset and its significance is also proved by testing its performance for another 9 standard datasets for different kinds of cancer. The comparative analysis done with a number of well-known evolutionary algorithms based feature selection methods showed reasonably improved performance of the proposed algorithm.

For the classification phase, two different types of classification models are proposed for digital/dermoscopic images classification and histopathological image analysis. Firstly, an expertise weight based advised SVM model is presented that reached a classification accuracy of around 95% for digital/dermoscopic skin image. Limited availability of training data especially for the case of histopathological images is a major limitation in the development of more generalized learning models. Therefore, a semi advised SVM algorithm is proposed that is used in parallel with a deep learning based algorithm for the analysis and classification and the model can efficiently use both labelled and unlabelled training data. It achieved better results than traditional SVM, KNN, Neural Networks, Transductive SVM as well as Expectation Maximization

algorithms. It reached an accuracy of 94% for digital/dermoscopic images and around 92% for histopathlogical images which is a significant contribution keeping an eye on the complexity of histopathological images.

7.3 Future Works

In this thesis dermoscopic/digital and histopathological images were used, but other image acquisition methods e.g. Laser based enhanced diagnosis, optical Coherence Tomography, magnetic resonance imaging etc. can also be considered as the future trend in skin studies. Similarly, detection of melanoma with some under skin imaging methods can be an interesting area to extend the skin cancer research to help the medical specialist in detecting melanoma before it becomes visible.

Of the methods that developed in this thesis, there are a number of directions in which they could be extended. For Example,

Image segmentation can have a huge impact on the final result; the proposed method can be applied on other type of images with slight modifications based on the application. Similarly, while a significant effort is made for coming up with the proposed segmentation algorithm but it worth further investigation for a more intelligent segmentation algorithm for skin cancer.

Ranking of features based on their capability of detecting benign or melanoma lesions can be considered by attention to their performance presented in this thesis. Adding more criteria and using multi criteria decision making algorithms can also be considered in future.

In this thesis the classification models are proposed for binary classification, so extending this approach for multi classification problems can be considered in future works.

The proposed automated diagnostic system can be extended to an online diagnostic system by making use of cloud based computation, parallel high speed processing and data sharing capability. This can also help in combining the cancer related research resources like DERMOFIT (research project conducted in University of Edinburg) where the research is carried out for developing an interactive skin cancer image database indexing tool, cancer council libraries and various other resources and make a more centralized system that can be made accessible for patients, researchers and medical experts for knowledge sharing. At the same time the useful data collected through such resources can be used to strengthen the learning capability of the online automated diagnostic system that can be developed by making use of machine learning algorithms proposed in this thesis.

7.4 Summary

This chapter was aimed to briefly summarize the overall thesis contributions and provide some future directions for the application and extension of the methods presented in this thesis.

References

1. *"Cancer Facts & Figures 2002"*, by *American Cancer Society,* 2002: Atlanta, Georgia.
2. M.Binder, M.Schwarz, A.Winkler, "*Epiluminescence Microscopy. A useful tool for the diagnosis of pigmented lesions for formally trained dermatologists"*, Arch Dermatol, Vol. 131, pp. 286-291 (1995).

3. John W Kelly, Josephine M Yeatman, Cheryl Regalia, Grahame Mason and Amanda P Henham, "*A high incidence of melanoma found in patients with multiple dysplastic naevi by photographic surveillance"*, The Medical Journal of Australia, Vol. 167, pp. 191-194 (1997).

4. *American Cancer Society. Cancer Facts & Figures 2014*. Available from: http://www.cancer.org/acs/groups/content/@research/documents/webcontent/acspc-042151.pdf. Accessed January 5, 2015

5. M.P. Curado et al., *Cancer Incidence in five continents.* International Agency for Research on Cancer, (2007)

6. Cancer in australia, an overview 2014, by *Australian Institute of Health and Welfare 2014.*: Canberra.

7. *Cancer in Australia statistics*. 2016; Available from: (http://canceraustralia.gov.au).

8. V. Criscione, M.Weinstock, "*Melanoma Thickness Trends in the United States, 1988-2006"*, Journal of Investigative Dermatology, (2009).

9. H.M. Douglass, R.McGee, S.Williams, "*Are young adults checking their skin for melanoma?,"*Australian and New Zealand Journal of Public Health, Vol. 22: pp. 562-567 (1998).

10. A. Andrion, C. Magnani, P G Betta, A. Donna, F Mollo, M Scelsi, P Bernardi, M Botta, B Terracini, "*Malignant mesothelioma of the pleura: interobserver variability,"*Journal of clinical pathology, Vol. 48(9), pp. 856-860 (1995).

11. M. Preti, et al., "*Inter-observer variation in histopathological diagnosis and grading of vulvar intraepithelial neoplasia: results of an European collaborative study,"*An International Journal of Obstetrics & Gynaecology, Vol. 107(5), pp. 594-599 (2000).

12. B.H. Porras, CJ Cockerell, "*Cutaneous malignant melanoma: Classification and clinical diagnosis"*, Seminars in Cutaneous Medicine and Surgery, Vol. 16(2): pp. 88-96 (1997).

13. John W Kelly, J.M.Yeatman, Cheryl Regalia, Grahame Mason and Amanda P Henham, "*A high incidence of melanoma found in patients with multiple dysplastic naevi by photographic surveillance"*, Medical journal of Australia, Vol. 167(4), pp. 191-194 (1997).

14. Rhodes AR, Weistock MA, Fitzpatrick TB et. al., "*Risk factors for cutaneous melanoma: A practical method of recognizing predisposed individuals",*The Journal of the American Medical Association, Vol. 258(21), pp. 3146-3154 (1987).

15. W.H. Clark, Elder DE. Guerry et.al, "*Model Predicting Survival in Stage I Melanoma Based on Tumor Progression",* Journal of the National Cancer Institute, Vol. 81(24), pp. 1893-1904 (1989).

16. M.N. Sirott et al., "*Prognostic factors in patients with metastatic malignant melanoma: A multivariate analysis"*, Cancer, Vol.72(10), pp. 3091-3098 (1993).

17. Stidham KR, Johnson JL, Seigler HF, "*Survival superiority of females with melanoma: A multivariate analysis of 6383 patients exploring the significance of*

gender in prognostic outcome" , Archives of Surgery, Vol.129(3), pp. 316-324 (1994).

18. Thorn M, Ponten F, Bergstrom R, Sparen P, Adami HO, "*Clinical and Histopathologic Predictors of Survival in Patients With Malignant Melanoma: a Population-Based Study in Sweden"*, Journal of the National Cancer Institute, Vol. 86(10), pp. 761-769 (1994).

19. Clark WH Jr, From L, Bernardino EA, Mihm MC, "*The Histogenesis and Biologic Behavior of Primary Human Malignant Melanomas of the Skin" ,* Cancer Research, Vol.29(3), pp. 705-727 (1969).

20. H. Wallace, J. Clark, M.D. Martin, C. Mihm, "*Lentigo maligna and lentigo maligna melanoma",* American Journal of Pathology, Vol. 55(1), pp. 39-67 (1969).

21. V.J. McGovern, "*The Classification of Melanoma and its Relationship with Prognosis",* Pathology, Vol. 2(2), pp. 85-98 (1970).

22. Ascierto P.A. et.al., "*Early Diagnosis of MAlignant MElanoma: Proposal of a working formulation for the Management of Cutaneous Pigmented Lesions for the melanoma Coorperative Group",* International Journal of Oncology, Vol. 22(6), pp. 1209-1215 (2003).

23. R.P. Braun, L.E. French, J.H. Saurat, "*Dermoscopy of Pigmented Lesions: A Valueable Tool in the Diagnosis of melanoma",* Swiss Medical Weekly, Vol. 134(7-8), pp. 83-90 (2004).

24. Fleming, M.G. et al., "*Techniques for a structural analysis of dermatoscopic imagery",* Computerized Medical Imaging and Graphics, Vol. 22(5), pp. 375-389 (1998).

25. Pariser RJ, Pariser DM, "*Primary care physicians' errors in handling cutaneous disorders: A prospective survey" ,* Journal of the American Academy of Dermatology, Vol.17(2, Part 1), pp. 239-245 (1987).

26. Franz Nachbar et.al., "*The ABCD rule of dermatoscopy: High prospective value in the diagnosis of doubtful melanocytic skin lesions",* Journal of the American Academy of Dermatology, Vol.30(4), pp. 551-559 (1994).

27. Cancer Treatment, Staging, & Guideline Syntheses, Available from: http:// http://reference.medscape.com/guide/cancer-guides.

28. Argenziano G, Giorgy VD, "*Interactive Atlas of Dermoscopy. Dermoscopy Tutorial: Vascular Structures",* EDRA Medical Publishing & New Media (2002).

29. Argenziano G. et.al., "*Epiluminescence microscopy for the diagnosis of doubtful melanocytic skin lesions. Comparison of the ABCD rule of dermatoscopy and a new seven-point checklist based on pattern analysis",* Archives of Dermatology, Vol. 134, pp. 1563-1570 (1998).

30. Yuan Xiaojing et. al. "*SVM-based Texture Classification and Application to Early Melanoma Detection",* in Proc. 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, New york, USA, pp. 4775-4778 (2006).

31. Menzies SW, Ingvar C., Crotty KA, McCarthy WH, "*Frequency and morphologic characteristics of invasive melanomas lacking specific surface microscopic features",* Archives of Dermatology, Vol.132, pp. 1178-1182 (1996).

32. D.J. Eedy, "*An atlas of surface microscopy of pigmented skin lesions: dermoscopy" 2nd Edn.* British Journal of Dermatology, Vol.149(5), pp. 1093-1093 (2003).

33. Westerhoff K, McCarthy WH, Menzies SW, "*Increase in the sensitivity for melanoma diagnosis by primary care physicians using skin surface microscopy",* British Journal of Dermatology, Vol.143, pp. 1016-1020 (2000).

34. Soyer HP et. al., "*Three-Point Checklist of Dermoscopy",* Dermatology, Vol. 208(1), pp. 27-31 (2004).

35. Argenziano G et.al. , "*Dermoscopy of Pigmented skin lesions: results of a consensus meeting via the internet",* Journal of American Academy of Dermatology, Vol. 48 pp.679-693 (2003).

36. Crowson A.N, Magro CM, and Mihm MC, "*Prognosticators of melanoma, the melanoma report, and the sentinel lymph node",* Modern Pathology, Vol. 19, pp.71-87 (2006).

37. Wallace H., Clark Jr. et. al., "*A classification of malignant melanoma in man correlated with histogenesis and biological behaviour",* Advances in the Biology of the Skin, pp. 621-647 (1967).

38. Clark Jr WH, From L, Bernardino EA, Mihm MC, "*Histogenesis and biologic behavior of primary human malignant melanomas of the skin",* Cancer Research,Vol. 29, pp. 705-726 (1969).

39. A. Breslow, "*Thickness, cross-sectional areas and depth of invasion in the prognosis of cutaneous melanoma",* Annals of Surgery, Vol.172, pp. 902-908 (1970).

40. Crowson A., et al., "*The pathology of melanoma",* Cutaneous Melanoma, 4th edn. Quality Medical Publishers Inc.: St Louis, pp. 171-206 (2003).

41. Babu, M.N., et al., "*Histo-pathological Image Analysis using OS-FCM and Level Sets", in Proc. of IEEE 39th Applied Imagery Pattern Recognition Workshop (AIPR), pp. 1-8* (2010).

42. Demir Ciqdem, Yener Bulent, "*Automated cancer diagnosis based on histopathological images: a systematic survey"* , in *TR-05-09*, Rensselaer Polytechnic Institute, Department of computer science (2009).

43. Clark Jr WH et.al. , "*Model predicting survival in stage I melanoma based on tumor progression",Journal of* Natural Cancer Inst, Vol.81, pp. 1893-1904 (1989).

44. Clark W.H. , M.C. Mihm, "*Lentigo maligna and lentigo maligna melanoma",* American Journal of Pathology, Vol.55, pp. 39-54 (1969).

45. Reed R.J., "*Acral lentiginous melanoma",* In: New Concepts in Surgical Pathology of the Skin. Wiley, New York, pp. 89-90 (1976).

46. H.K. Koh, E. Michalik, A.J. Sober, "*Lentigo maligna melanoma has no better prognosis than other types of melanoma",* Journal of Clinical Oncology, Vol.2, pp. 994-1001 (1984).

47. R.T. Vollmer, "*Malignant melanoma. A multivariate analysis of prognostic factors"* Pathology Annual, Vol.24, pp. 383-407 (1989).

48. M.A. Weinstock, A.J. Sober, "*The risk of progression of lentigo maligna to lentigo maligna melanoma",* British Journal Dermatology, Vol.116, pp. 303-310 (1987).

49. I.J. Fidler, "*Tumor heterogeneity and the biology of cancer invasion and metastasis",* Cancer Research, Vol.38, pp. 2651-2660 (1978).

50. C.M. Balch, S.J. Soong, J.E. Gershenwald, "*Prognostic factors analysis of 17,600 melanoma patients: validation of the American Joint Committee on Cancer melanoma staging system",* Journal of Clinical Oncology, Vol.19, pp. 3622-3634 (2001).

51. R.L. Barnhill, J.A. Fine, G.C. Roush, "*Predicting fiveyear outcome for patients with cutaneous melanoma in a population-based study",* Cancer, Vol.78, pp. 427-432 (1996).

52. C.M. Balch, T.M. Murad, S.J. Soong, "*A multlifactorial analysis of melanoma: prognostic histopathological features comparing Clark's and Breslow's staging methods",* Annual Surgery, Vol.188, pp. 732-742 (1978).

53. C.M. Balch, S.J. Soong, T.M. Murad, "*A multifactorial analysis of melanoma. II. Prognostic factors in patients with stage I (localized) melanoma",* Surgery, Vol. 86, pp. 343-351 (1979).

54. N. Cascinelli, A. Morabito, R. Bufalino, "*Prognosis of stage I melanoma of the skin",* International Journal of Cancer, Vol. 26, pp. 733-739 (1980).

55. K. Sondergaard, G. Schou, "*Survival with primary cutaneous malignant melanoma, evaluated from 2012 cases. A multivariate regression analysis",* Virchows Archives A Pathological anatomy and Histopathology, Vol.406, pp. 179-195 (1985).

56. D.S. Rigel, R.J. Friedman, A.W. Kopf, "*Factors influencing survival in melanoma"* Dermatologic Clinics, Vol. 9, pp. 631-642 (1991).

57. K.R. Stidham, J.L. Johnson, H.F. Seigler, "*Survival superiority of females with melanoma. A multivariate analysis of 6383 patients exploring the significance of gender in prognostic outcome",* Archives of Surgery, Vol.129, pp. 316-324 (1994).

58. C. Garbe, P. Buttner, J. Bertz, "*Primary cutaneous melanoma. Identification of prognostic groups and estimation of individual prognosis for 5093 patients",* Cancer, Vol.75, pp. 2484-2491 (1995).

59. C.M. Balch, S. Soong, M.I. Ross, "*Long-term results of a multi-institutional randomized trial comparing prognostic factors and surgical results for intermediate thickness melanomas (1.0 to 4.0 mm)" ,* Intergroup Melanoma Surgical Trial. Annals of Surgical Oncology, Vol.7, pp. 87-97 (2000).

60. S. Retsas, K. Henry, M.Q. Mohammed, "*Prognostic factors of cutaneous melanoma and a new staging system proposed by the American Joint Committee on Cancer (AJCC): validation in a cohort of 1284 patients",* European Journal of Cancer, Vol.38, pp. 511-516 (2002).

61. S.G. Ronan, M.C. Han, T.K. Das Gupta, "*Histologic prognostic indicators in cutaneous malignant melanoma",* Seminars in Oncology, Vol.15, pp. 558-565 (1988).

62. R.T.V ollmer, H.F. Seigler, "*A model for pretest probability of lymph node metastasis from cutaneous melanoma",* American Journal of Clinical Pathology, Vol.114, pp. 875-879 (2000).

63. C.L. Day, T.J. Harrist, F. Gorstein, "*Malignant melanoma. Prognostic significance of 'microscopic satellites' in the reticular dermis and subcutaneous fat"*Annals of Surgery, Vol.194, pp. 108-112 (1981).

64. T.J. Harrist, D.S. Rigel, and J.C.L. Day, *"Microscopic satellites' are more highly associated with regional lymph nodes metastases than is primary melanoma thickness",* Cancer, Vol.53, pp. 2183-2187 (1984).

65. P. Leon, J.M. Daly, M. Synnestvedt, "*The prognostic implications of microscopic satellites in patients with clinical stage I melanoma",* Archives of Surgery, Vol.126, pp. 1461-1468 (1991).

66. U.N. Rao, J. Ibrahim, L.E. Flaherty, "*Implications of microscopic satellites of the primary and extracapsular lymph node spread in patients with high-risk melanoma: pathology corollary of Eastern Cooperative Oncology Group Trial E1690",* Journal of Clinical Oncology, Vol. 20, pp.2053-2057 (2002).

67. J.M. Mascaro, M. Molgo, T. Castel, "*Plasma cells within the infiltrate of primary cutaneous malignant melanoma of the skin. A confirmation of its histoprognostic value",* American Journal of Dermatopathology, Vol.9, pp. 497-499 (1987).

68. M. Piepkorn, "*Prognostic factors in cutaneous melanoma",* Pathology of Melanocytic Nevi and Malignant Melanoma, Springer, Barnhill RL, pp. 372-394 (2004).

69. Ammara Masood, Adel Al-Jumaily, "*Computer Aided Diagnostic Support System for Skin Cancer: A Review of Techniques and Algorithms",* International Journal of Biomedical Imaging, pp.1-22 (2013).

70. Ammara Masood, Adel Al Jumaily, Tariq Adnan, "*Development of Automated Diagnostic System for Skin Cancer: Performance Analysis of Neural Network Learning Algorithms for Classification"*, in Artificial Neural Networks and Machine Learning – ICANN, Springer International Publishing, pp. 837-844 (2014).

71. Cancer Council Australia. *Screening and early detection of skin cancer*. 2007; Available from: www.cancer.org.auCancer

http://www.cancercouncil.com.au/editorial.asp?pageid=57.

72. E.G iovannucci, Liu, Y, Rimm, , Hollis, Fuchs, Stampfer, Willett, "*Prospective study of predictors of vitamin D status and cancer incidence and mortality in men"* Journal of the National Cancer Institute, Vol. 98(7), pp. 451-459 (2006).

73. Phillippa Youl et al., "*Diagnosing skin cancer in primary care",* The Medical Journal of Australia, Vol. 187, pp. 215-220 (2007).

74. Nadine Kasparian , McLoone, B.Meiser, "*Skin cancer-related prevention and screening behaviors: a review of the literature",* Journal of Behavioural Medicine,Vol. 32, pp. 406-428 (2009).

75. S. Valiukeviciene, "*Meolanoma Diagnosis and treatment Principles.* Kaunas KMU, pp. 85-89, (2002).

76. H.P ehamberger, A. Steiner, K. Wolff, "*In vivo epiluminescence microscopy of pigmented skin lesions. I. Pattern analysis of pigmented skin lesions",* Journal of the American Academy of Dermatology, Vol.17(4), pp. 571-583 (1987).

77. A. Steiner, H. Pehamberger, K. Wolff, "*Improvement of the diagnostic accuracy in pigmented skin lesions by epiluminescent light microscopy",* Anticancer research, Vol. 7(3 Pt B), pp. 433-434 (1987).

78. Abbasi Nr et al., "*Early diagnosis of cutaneous melanoma: Revisiting the abcd criteria" ,* The Journal of the American Medical Association, Vol. 292(22), pp. 2771-2776 (2004).

79. Robinson, "*Skills training to learn discrimination of abcde criteria by those at risk of developing melanoma",* Archives of Dermatology, Vol.142(4), pp. 447-452 (2006).

80. R.H. Johr, "*Dermoscopy: alternative melanocytic algorithms—the ABCD rule of dermatoscopy, menzies scoring method, and 7-point checklist",* Clinics in dermatology, Vol.20(3), pp. 240-247 (2002).

81. M.F.Healsmith et al., "*An evaluation of the revised seven-point checklist for the early diagnosis of cutaneous malignant melanoma" ,* British Journal of Dermatology, Vol.130(1), pp. 48-50 (1994).

82. V. Dal Pozzo, E. Roscetti, "*The seven features for melanoma: a new dermoscopic algorithm for the diagnosis of malignant melanoma",* European Journal of Dermatology, Vol. 9(4), pp. 303-308 (1999).

83. I.A. Zalaudek et al.,The Dermoscopy Working, Group, "*Three-point checklist of dermoscopy: an open internet study",* British Journal of Dermatology, Vol.154(3), pp. 431-437 (2006).

84. M.E.M. Vestergaard, W. Scott, "*Automated Diagnostic Instruments for Cutaneous Melanoma",* Seminars in Cutaneous Medicine and Surgery,Vol.27(1), pp. 32-36 (2008).

85. Binder M et al., "*Epiluminescence microscopy: A useful tool for the diagnosis of pigmented skin lesions for formally trained dermatologists"* , Archives of Dermatology, Vol.131(3), pp. 286-291 (1995).

86. P.C arli et al., "*Reliability and inter-observer agreement of dermoscopic diagnosis of melanoma and melanocytic naevi",* European Journal of Cancer Prevention, Vol.7(5), pp. 397-402 (1998).

87. I. Stanganelli et al., "*Intraobserver agreement in interpretation of digital epiluminescence microscopy",* Journal of the American Academy of Dermatology, Vol.33(4), pp. 584-589 (1995).

88. N. Cascinelli et al., "*A possible new tool for clinical diagnosis of melanoma: The computer",* Journal of the American Academy of Dermatology, Vol.16(2, Part 1): pp. 361-367 (1987).

89. P.N. Hall, E. Claridge, J.D.M. Smith, "*Computer screening for early detection of melanoma—is there a future?",* British Journal of Dermatology, Vol.132(3), pp. 325-338 (1995).

90. M. Cristofolini et al., "*Diagnosis of cutaneous melanoma: accuracy of a computerized image analysis system (Skin View)",* Skin Research and Technology, Vol. 3(1), pp. 23-27 (1997).

91. S.E. Umbaugh, "*Computer vision in medicine: color metrics and image segmentation methods for skin cancer diagnosis"*, in *Electrical Engineering Department*, University of Missouri/Rolla USA (1990).

92. Ignazio Stanganelli, Luigi Calori, Roberto Gori et. al., "*Computer aided diagnosis of melanocytic tumors,"*AntiCancer Research, Vol.25, pp. 4577-4582 (2005).

93. P.R ubegni et al., "*Automated diagnosis of pigmented skin lesions",* International Journal of Cancer, Vol.101(6), pp. 576-580 (2002).

94. A.J. Sober, J.M. Burstein, "*Computerized digital image analysis: an aid for melanoma diagnosis--preliminary investigations and brief review",* The Journal of dermatology, Vol.21(11), pp. 885-890 (1994).

95. S.M. Rajpara et al., "*Systematic review of dermoscopy and digital dermoscopy/ artificial intelligence for the diagnosis of melanoma",* British Journal of Dermatology, Vol.161(3), pp. 591-604 (2009).

96. Rosado B, M.S., Harbauer A et al., "*Accuracy of computer diagnosis of melanoma: a quantitative meta-analysis",* Archives of Dermatology, Vol.139, pp.361-367 (2003).

97. P.C.Bauer, P. Boi, et al., "*Digital epiluminescence microscopy: usefulness in the differential diagnosis of cutaneous pigmentary lesions. A statistical comparison between visual and computer inspection",*Melanoma Research, Vol.10(4), pp.345-349 (2000).

98. Ilias Maglogiannis, "*Overview of Advanced Computer Vision Systems for Skin Lesions Characterization",* IEEE Transactions on Information Technology in Biomedicine, Vol. 13(5), pp. 721-733 (2009).

99. Friedman Rj et al., "*The diagnostic performance of expert dermoscopists vs a computer-vision system on small-diameter melanoma",*. Archives of Dermatology, Vol. 144(4), pp. 476-482 (2008).

100. A. Blum, I. Zalaudek, G. Argenziano, "*Digital Image Analysis for Diagnosis of Skin Tumors",* Seminars in Cutaneous Medicine and Surgery, Vol.27(1), pp. 11-15 (2008).

101. A.A. Marghoob et al., "*Instruments and new technologies for the in vivo diagnosis of melanoma",* Journal of the American Academy of Dermatology, Vol.49(5), pp. 777-797 (2003).

102. D.S. Rigel, J. Russak, R. Friedman, "*The Evolution of Melanoma Diagnosis: 25 Years Beyond the ABCDs",* CA: A Cancer Journal for Clinicians, Vol. 60(5), pp.301-316 (2010).

103. E.L.P saty, A.C. Halpern, "*Current and emerging technologies in melanoma diagnosis: the state of the art",* Clinics in dermatology, Vol. 27(1), pp.35-45 (2009).

104. T. Schindewolf et al., "*Evaluation of different image acquisition techniques for a computer vision system in the diagnosis of malignant melanoma",* Journal of the American Academy of Dermatology, Vol. 31(1), pp. 33-41 (1994).

105. K.S. Nehal et al., "*Use of and beliefs about baseline photography in the management of patients with pigmented lesions: a survey of dermatology*

residency programmes in the United States", Melanoma Research, Vol.12(2), pp.161-167 (2002).

106. N.E. Feit, S.W. Dusza, A.A. Marghoob, "*Melanomas detected with the aid of total cutaneous photography",* British Journal of Dermatology, Vol.150(4), pp. 706-714 (2004).

107. W. Slue, A.W. Kopf, J.K. Rivers, "*Total-body photographs of dysplastic nevi",* Archives of Dermatology, Vol.124(8), pp.1239-1243 (1988).

108. G. Argenziano, H.P. Soyer, "*Dermoscopy of pigmented skin lesions – a valuable tool for early Diagnosis of melanoma",* The Lancet Oncology, Vol.2(7), pp. 443-449 (2001).

109. R.P. Braun et al., "*Dermoscopy of pigmented skin lesions",* Journal of the American Academy of Dermatology, Vol. 52(1), pp. 109-121 (2005).

110. M. Stante et al., "*Non-invasive analysis of melanoma thickness by means of dermoscopy: a retrospective study",* Melanoma Research, Vol.11(2), pp. 147-152 (2001).

111. D. Massi, V. De Giorgi, H.P. Soyer, "*Histopathologic Correlates of Dermoscopic Criteria",* Dermatologic Clinics, Vol.19(2), pp. 259-268 (2001).

112. F.S tauffer et al., "*The dermatoscope: a potential source of nosocomial infection?"* Melanoma Research, Vol.11(2), pp. 153-156 (2001).

113. S.S eidenari et al., *Computerized evaluation of pigmented skin lesion images recorded by a videomicroscope: comparison between polarizing mode observation and oil/slide mode observation",* Skin Research and Technology, Vol.1(4), pp. 187-191 (1995).

114. M. Carrara et al., "*Multispectral imaging and artificial neural network: mimicking the management decision of the clinician facing pigmented skin lesions",* Physics in Medicine and Biology, Vol. 52(9), pp. 2599-2613 (2007).

115. M. Carrara et al., "*Automated segmentation of pigmented skin lesions in multispectral imaging",* Physics in Medicine and Biology, Vol.50(22), pp.345-357 (2005).

116. M. Elbaum et al., "*Automatic differentiation of melanoma from melanocytic nevi with multispectral digital dermoscopy: a feasibility study",* Journal of the American Academy of Dermatology, Vol.44(2), pp. 207-218 (2001).

117. A. Lorber et al., "*Correlation of image analysis features and visual morphology in melanocytic skin tumours using in vivo confocal laser scanning microscopy,"* Skin Research and Technology, Vol.15(2), pp. 237-241 (2009).

118. A. Gerger et al., "*In vivo confocal laser scanning microscopy in the diagnosis of melanocytic skin tumours",* British Journal of Dermatology, Vol.160(3), pp.475-481 (2009).

119. Hofmann Wellenhof, Ahlgrimm-Siess et al., "*Reflectance confocal microscopy--state-of-art and research overview,"* Seminars in Cutaneous Medicine and Surgery Vol. 28, pp.172-179, (2009).

120. J. Welzel, "*Optical coherence tomography in dermatology: a review"*, Skin Research and Technology, Vol. 7(1), pp. 1-9 (2001).

121. M.C. Pierce et al., "*Advances in Optical Coherence Tomography Imaging for Dermatology",* Journal of Investigative Dermatology, Vol.123(3), pp.458-463 (2004).

122. M. Mogensen et al., "*OCT imaging of skin cancer and other dermatological diseases",* Journal of Biophotonics, Vol. 2(6-7), pp. 442-451 (2009).

123. L. Serrone et al., "*High frequency ultrasound in the preoperative staging of primary melanoma: a statistical analysis",* Melanoma Research, Vol. 2(3), pp. 287-290 (2002).

124. Blum A et al., "*Ultrasound Mapping of Lymph Node and Subcutaneous Metastases in Patients with Cutaneous Melanoma: Results of a Prospective Multicenter Study",* Dermatology, Vol. 212(1), pp. 47-52 (2006).

125. J. Maurer et al., "*Role of High-Resolution Magnetic Resonance Imaging for Differentiating Melanin-Containing Skin Tumors",* Investigative Radiology, Vol. 30(11), pp. 638-643 (1995).

126. I.O no, and F. Kaneko, "*Magnetic resonance imaging for diagnosing skin tumors,"* Clinics in dermatology, Vol.13(4), pp. 393-399 (1995).

127. S. Gammal et al., "*Improved Resolution of Magnetic Resonance Microscopy in Examination of Skin Tumors",* Journal of Investigative Dermatology, Vol.106(6), pp. 1287-1292 (1996).

128. M. Sasikala, N. Kumaravel, "*Comparison of Feature Selection Techniques for Detection of Malignant Tumor in Brain Images",* in proc. of *2005 Annual IEEE India Conference, Dec. 2005, pp.212-215* (2005).

129. S. Nakariyakul, D.P. Casasent, "*Improved forward floating selection algorithm for feature subset selection",* in proc. of International Conference on Wavelet Analysis and Pattern Recognition,August 2008, pp. 739-798 (2008).

130. M.E. Celebi, Y.A. Aslandogan, P.R. Bergstresser, "*Unsupervised border detection of skin lesion images",* in proc. of International Conference onInformation Technology: Coding and Computing, April 2005, pp. 123-128 (2005).

131. S. Sookpotharom, "*Border Detection of Skin Lesion Images Based on Fuzzy C-Means Thresholding" ,* in 3rd International Conference on Genetic and Evolutionary Computing, pp. 777-780 (2009).

132. M. Celebi, Hassan Kingravi, J. Lee. "*Fast and accurate border detection in dermoscopy images using statistical region merging",* in Proc. SPIE Medical Imaging, pp. 1-10 (2007).

133. M. Silveira, J.S. Marques, "*Level set segmentation of dermoscopy images,"*in 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, May 2008, pp. 173-176 (2008).

134. I. Maglogiannis, E. Zafiropoulos, and C. Kyranoudis, "*Intelligent Segmentation and Classification of Pigmented Skin Lesions in Dermatological Images"*, in Advances in Artificial Intelligence, pp. 214-223 (2006).

135. Bulent Erkol et al., "*Automatic lesion boundary detection in dermoscopy images using gradient vector flow snakes",* Skin Research and Technology, Vol.11, pp. 17-26 (2005).

136. M.E. Celebi, H.A. Kingravi, Y.A. Aslandogan, "*Nonlinear vector filtering for impulsive noise removal from color images",* Journal of Electronic Imaging, Vol. 16(3), pp. 1-65 (2007).

137. M. Emre Celebi, Hitoshi Iyatomi, William V. Stoecker, "*Lesion Border Detection in Dermoscopy Images",* Computer Medical Imaging Graphics, Vol.33(2), pp.148-153 (2009).

138. Q. Abbas et al., "*Lesion border detection in dermoscopy images using dynamic programming",* Skin Research and Technology, Vol.17(1), pp. 91-100 (2011).

139. Q.A bbas, I. Fondón, M. Rashid, "*Unsupervised skin lesions border detection via two-dimensional image analysis" ,*Computer Methods and Programs in Biomedicine, Vol.104(3), pp. 1-15 (2011).

140. M. Celenk, "*A color clustering technique for image segmentation",* Computer Vision, Graphics, and Image Processing Journal, Vol.52(2), pp. 145-170 (1990).

141. D.D. Gomez et al., "*Independent Histogram Pursuit for Segmentation of Skin Lesions",* IEEE Transactions on Biomedical Engineering, Vol. 55(1), pp. 157-161 (2008).

142. H.G. Adelmann, "*Butterworth equations for homomorphic filtering of images,"* Computers in Biology and Medicine, Vol. 28(2), pp.169-181 (1998).

143. M. Emre Celebi et al., "*Border detection in dermoscopy images using statistical region merging",* Skin Research and Technology Journal, Vol.14(3), pp.347-353 (2008).

144. M.R. Sweet, "A*daptive and recursive median filtering" , Available from: http://www.easysw.com/~mike/gimp/despeckle.html*.

145. QinzLi et al. "*Dark line detection with line width extraction",* in 15th IEEE International Conference on Image Processing, Octobar 2008, pp. 621-624, (2008).

146. A. Criminisi, P. Perez, and K. Toyama, "*Region filling and object removal by exemplar-based image inpainting",* IEEE Transactions on Image Processing, Vol.13(9), pp.1200-1212 (2004).

147. Lee T, Gallagher R, Coldman A, McLean D. , "*DullRazor: a software approach to hair removal from images",* Computers in Boilogy and medicine, Vol.27(6), pp. 533-543 (1997).

148. P.S chmid, "*Segmentation of digitized dermatoscopic images by two-dimensional color clustering",* IEEE Transactions on Medical Imaging, Vol.18(2), pp. 164-171 (1999).

149. C. Do Hyun, and G. Sapiro, "*Segmenting skin lesions with partial-differential-equations-based image processing algorithms",* IEEE Transactions on Medical Imaging, Vol.19(7), pp. 763-767 (2000).

150. M. d'Amico, M. Ferri, and I. Stanganelli, "*Qualitative asymmetry measure for melanoma detection",* in IEEE International Symposium on Biomedical Imaging: Nano to Macro, pp. 1156-1158 (2004).

151. P. Schmid, "*Lesion detection in dermatoscopic images using anisotropic diffusion and morphological flooding",* in International Conference on Image Processing, pp. 449-453 (1999).

152. P. Schmid-Saugeona, J. Guillodb, and J.-P. Thirana, "*Towards a computer-aided diagnosis system for pigmented skin lesions",* Computerized Medical Imaging and Graphics, Vol.27(1), pp. 65-78 (2003).

153. R.C. Gonzales, *Digital Image Processing*. 2nd ed , New Jersey: Prentice Hall (2002).

154. Wyszecki G, "*Color science: concepts and methods, quantitative data and formulae",* New York: Wiley (1983).

155. Howard Zhou, M.C., Richard Gass et al. "*Feature-preserving Artifact Removal from Dermoscopy Images",* in Proceedings of the SPIE Medical Imaging conference, pp. 1-9 (2008).

156. Paul Wighton, Tim K. Lee, M.S. Atkins, "*Dermoscopic hair disocclusion using inpainting",* in Proceedings of the SPIE Medical Imaging , pp.1-8 (2008).

157. Z. Huiyu et al., "*Anisotropic Mean Shift Based Fuzzy C-Means Segmentation of Dermoscopy Images",* IEEE Journal of Selected Topics in Signal Processing, Vol.3(1), pp. 26-34 (2009).

158. Q. Abbas, M.E. Celebi, and I.F. García, "*Hair removal methods: A comparative study for dermoscopy images",* Biomedical Signal Processing and Control, Vol.6(4), pp.395-404 (2011).

159. W.V. Stoecker et al. , "*Boundary Detection Algorithm Implementation for Medical Images"*, International Journal of Engineering Research & Technology, Vol. 3, pp.285-287 (2014).

160. J.B. Maintz, M.A. Viergever, "*A survey of medical image registration",* Journal on Medical Image Analysis, Vol.2(1), pp. 1-36 (1998).

161. V. Caselles, R. Kimmel, and G. Sapiro, "*Geodesic Active Contours",* International Journal of Computer Vision, Vol.22(1), pp. 61-79 (1997).

162. M. Silveira et al., "*Comparison of Segmentation Methods for Melanoma Diagnosis in Dermoscopy Images",* IEEE Journal of Selected Topics in Signal Processing, Vol.3(1), pp. 35-45 (2009).

163. X.Chenyang, and J.L. Prince, "*Snakes, shapes, and gradient vector flow",* IEEE Transactions on Image Processing, Vol.7(3), pp. 359-369 (1998).

164. Pramod Pagadala, "*Tumor border detection in epiluminescence microscopy images*," Master Thesis in *Department of Electrical Engineering*, University of Missouri-Rolla (1998).

165. Z. Zhang, W.V. Stoecker, and R.H. Moss, "*Border detection on digitized skin tumor images",* IEEE Transactions on Medical Imaging, Vol.19(11), pp.1128-1143 (2000).

166. J. Sikorski, "*Identification of Malignant Melanoma by Wavelet Analysis,*" in Student/Faculty Research Day, CSIS. Pace University (2004).

167. G. Betta et al. "*Automated Application of the 7-point checklist' Diagnosis Method for Skin Lesions: Estimation of Chromatic and Shape Parameters",* in proc. of the

IEEE Instrumentation and Measurement Technology Conference, pp.1818-1822 (2005).

168. K. Ramlakhan, and S. Yi., "*A Mobile Automated Skin Lesion Classification System"* in 23rd IEEE International Conference on Tools with Artificial Intelligence, pp.138-141,(2011).

169. D. Ruiz et al., "*A decision support system for the diagnosis of melanoma: A comparative approach",* Expert Systems with Applications, Vol.38(12), pp.15217-15223 (2011).

170. Valerio De Vita, Gabriella Fabbrocini, Consolatina Liguori, Alfredo Paolillo, Paolo Sommella, "*Statistical Techniques Applied to the Automatic Diagnosis of Dermoscopic Images."*Acta IMEKO, Vol.1(1), pp. 7-18 (2012).

171. S. Angelina, L.P. Suresh, and Krishna Veni, "*Image segmentation based on genetic algorithm for region growth and region merging",* in Proceedings of International Conference on Computing, Electronics and Electrical Technologies, pp.970-974 (2012).

172. Young Won, L. Sang Uk, "*On the color image segmentation algorithm based on the thresholding and the fuzzy c-means techniques",* International Journal on Pattern Recognition, Vol. 23(9), pp. 935-952 (1990).

173. J. Gao, J. Zhang, M.G. Fleming, "*A novel multiresolution color image segmentation technique and its application to dermatoscopic image segmentation"* in proc. International Conference on Image Processing, pp.408-411 (2000).

174. N. Senthilkumaran, R. Rajesh, "*Image Segmentation - A Survey of Soft Computing Approaches",* in International Conference on Advances in Recent Technologies in Communication and Computing, pp. 844-846 (2009).

175. N.R. Pal, and S.K. Pal, "*A review on image segmentation techniques",* Pattern Recognition, Vol. 26(9), pp. 1277-1294 (1993).

176. J. Maeda et al., "*Unsupervised Perceptual Segmentation of Natural Color Images Using Fuzzy-Based Hierarchical Algorithm Image Analysis*," B. Ersbøll and K. Pedersen, Editors., Springer Berlin / Heidelberg, pp. 462-471 (2007).

177. U. Maulik, "*Medical Image Segmentation Using Genetic Algorithms",* IEEE Transactions on Information Technology in Biomedicine, Vol.13(2), pp. 166-173 (2009).

178. Mantas Paulinas, "*A Survey of Genetic Algorithms applications for Image Enhancement and Segmentation,*" Information Technology and Control, Vol. 36, pp. 278-284, (2007).

179. Liu, J. and B. Zuo., "*The Segmentation of Skin Cancer Image Based on Genetic Neural Network*", in Proceedings of 2009 WRI World Congress on Computer Science and Information Engineering, pp 594-599 (2009).

180. G.S. Vennila, L.P. Suresh, and K.L. Shunmuganathan, "*Dermoscopic image segmentation and classification using machine learning algorithms*", in International Conference on Computing, Electronics and Electrical Technologies, pp.1122-1127 (2012).

181. Emre Celebi, Hassan A. Kingravi, Y. Alp Aslandogan, William V, Stoecker, "*Detection of blue-white veil areas in dermoscopy images using machine learning*

techniques", in Proceedings of the SPIE Medical Imaging Conference, pp.1-8 (2006).

182. H. Motoyama et al., "*Feature of malignant melanoma based on color information,*" in *SICE 2004 Annual Conference*, pp.230-233 (2004).

183. L. Xu et al., "*Segmentation of skin cancer images,*"Image and Vision Computing, Vol. 17(1), pp. 65-74 (1999).

184. Andy Chiem, Adel Al-Jumaily, Rami N. Khushaba, "*A Novel Hybrid System for Skin Lesion Detection*", in 3rd International Conference on Intelligent Sensors, Sensor Networks and Information, pp. 567-572 (2007).

185. G. Di Leo et al., "*An improved procedure for the automatic detection of dermoscopic structures in digital ELM images of skin lesions",* in IEEE Conference on Virtual Environments, Human-Computer Interfaces and Measurement Systems, pp. 1-6 (2008).

186. T. Mendonca et al., "*Comparison of Segmentation Methods for Automatic Diagnosis of Dermoscopy Images",* in 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 6572-6575 (2007).

187. M.M. Rahman, P. Bhattacharya, and B.C. Desai., *A multiple expert-based melanoma recognition system for dermoscopic images of pigmented skin lesions*. in Proc. of IEEE International Conference on BioInformatics and BioEngineering, pp. 1-6 (2008).

188. Giuseppe Di Leo et al., "*Towards an automatic diagnosis system for skin lesions: Estimation of blue-whitish veil and regression structures",* in 6th International Multi-Conference on Systems, Signals and Devices, pp. 1-6 (2009).

189. Giuseppe Di Leo et al., "*Automatic Diagnosis of Melanoma: A Software System Based on the 7-Point Check-List",* in 43rd Hawaii International Conference on System Sciences (HICSS),pp. 1-10 (2010).

190. S.E. Umbaugh, R.H. Moss, W.V. Stoecker, "*An automatic color segmentation algorithm with application to identification of skin tumor borders,*"Computerized Medical Imaging and Graphics, Vol.16(3), pp. 227-235 (1992).

191. G.A. Hance et al., "*Unsupervised color image segmentation: with application to skin tumor borders",* IEEE Engineering in Medicine and Biology Magazine, Vol.15(1), pp.104-111 (1996).

192. S.E. Umbaugh et al., "*Automatic color segmentation algorithms-with application to skin tumor feature identification",* IEEE Engineering in Medicine and Biology Magazine, Vol.12(3), pp. 75-82 (1993).

193. M.M. Rahman, B.C. Desai, and P. Bhattacharya, "*Image Retrieval-Based Decision Support System for Dermatoscopic Images",* in 19th IEEE International Symposium on Computer-Based Medical Systems, pp. 285-290 (2006).

194. Maciej Ogorzałek1, Leszek Nowak, Grzegorz Surówka , Ana Alekseenko, "*Modern Techniques for Computer-Aided Melanoma Diagnosis",* Melanoma in the Clinic - Diagnosis, Management and Complications of Malignancy (2011).

195. Dawei Nie, "*Classification of melanoma and Clark nevus skin lesions based on medical image processing techniques",* in 3rd International Conference on Computer Research and Development, pp. 31-34 (2011).

196. C. Grana, et al., "*A new algorithm for border description of polarized light surface microscopic images of pigmented skin lesions,"* IEEE Transactions on Medical Imaging, Vol. 22(8), pp. 959-964 (2003).

197. B. Erkol et al., "*Automatic lesion boundary detection in dermoscopy images using gradient vector flow snakes",* Skin Research and Technology, Vol.11(1), pp. 17-26 (2005).

198. M. Khalad Abu Mahmoud, Adel Al-Jumaily, "*Segmentation of skin cancer images based on gradient vector flow (GVF) snake",* in Proc. of International Conference on Mechatronics and Automation, pp. 216-220 (2011).

199. J.E.Golston, R.H. Moss, and W.V. Stoecker, "*Boundary detection in skin tumor images: An overall approach and a radial search algorithm",* Pattern Recognition, 23(11), pp. 1235-1247 (1990).

200. J.E. Golston et al., "*Automatic detection of irregular borders in melanoma and other skin tumors",* Computerized Medical Imaging and Graphics, Vol.16(3), pp. 199-203 (1992).

201. Andreassi L et al., "*Digital dermoscopy analysis for the differentiation of atypical nevi and early melanoma: A new quantitative semiology"* , Archives of Dermatology, Vol.135(12), pp. 1459-1465 (1999).

202. Marco Burroni et al., "*Melanoma Computer-Aided Diagnosis: Reliability and Feasibility Study",* Clinical Cancer Research, Vol.10, pp. 1881–1886 (2004).

203. Burroni, M. et al., "*Dysplastic naevus vs. in situ melanoma: digital dermoscopy analysis",* British Journal of Dermatology, Vol.152(4), pp. 679-684 (2005).

204. A.G. Manousaki et al., "*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, Vol.45(4), pp. 402-410 (2006).

205. Round, A.J., A.W.G. Duller, and P.J. Fish, "*Colour segmentation for lesion classification",* in Proceedings of the 19th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 582-585 (1997).

206. T. Fikrle, K. Pizinger, "*Digital computer analysis of dermatoscopical images of 260 melanocytic skin lesions; perimeter/area ratio for the differentiation between malignant melanomas and melanocytic nevi",* Journal of the European Academy of Dermatology and Venereology, Vol. 21(1), pp. 48-55 (2007).

207. M. Khalad Abu Mahmoud, Adel Al-Jumaily, and Maen Takruri, "*The automatic identification of melanoma by wavelet and curvelet analysis: Study based on neural network classification"* in 11th International Conference on Hybrid Intelligent Systems (HIS), pp. 680-685 (2011).

208. Emer Celebi et al., "*A methodological approach to the classification of dermoscopy images",* Computerized Medical Imaging and Graphics, Vol.31(6), pp. 362-373 (2007).

209. Karol Przystalski, M.O., Leszek Nowak, Grzegorz Surowka, "*Decision support system for skin cancer diagnosis,"* in 9th Internaltional symposium on operations research and it's applications, pp. 406-413 (2010).

210. A. Khan et al., "*Fuzzy logic techniques for blotch feature evaluation in dermoscopy images",* Computerized Medical Imaging and Graphics, Vol. 33(1), pp. 50-57 (2009).

211. E.Zagrouba & W. Barhoumi, "*An Accelerated System for Melanoma Diagnosis Based on Subset Feature Selection",*Journal of Computing and Information Technology, Vol.13(1), pp. 69-82 (2005).

212. S. Ning et al., "*Automatic Segmentation of Skin Lesion Images using Evolutionary Strategy",* Biomedical Signal Processing and Control, Vol. 3 pp. 220-228 (2008).

213. S.V. Patwardhan, A.P. Dhawan, P.A. Relue, "*Classification of melanoma using tree structured wavelet transforms" ,*Computer Methods and Programs in Biomedicine, Vol. 72(3), pp. 223-239 (2003).

214. G. Surowka, and K. Grzesiak-Kopec, "*Different Learning Paradigms for the Classification of Melanoid Skin Lesions Using Wavelets",* in 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 3136-3139 (2007).

215. *The GLCM Tutorial Home Page*. Available from: http://www.fp.ucalgary.ca/mhallbey/tutorial.htm.

216. T.T.Tanaka et al. "*Pattern classification of nevus with texture analysis",* in *26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society,* (2004).

217. Howard Zhou, James M. Rehg, "*Dermoscopic interest point detector and descriptor*,"in *Proc. of the Sixth IEEE international conference on Symposium on Biomedical Imaging: From Nano to Macro*2009, IEEE Press Boston, Massachusetts, USA. pp. 1318-1321 (2009).

218. L. Chulhee, and D.A. Landgrebe, "*Decision boundary feature extraction for neural networks",* IEEE Transactions on Neural Networks, Vol.8(1), pp. 75-83 (1997).

219. C. Lee, D.A. Landgrebe, "*Decision boundary feature extraction for nonparametric classification" ,* IEEE Transactions on Systems, Man and Cybernetics, Vol. 23(2), pp. 433-444 (1993).

220. C. Lee, and D.A. Landgrebe, "*Feature extraction based on decision boundaries",* IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.15(4), pp. 388-400 (1993).

221. X.Yuan et al., "*Multi-scale feature identification using evolution strategies",* Image and Vision Computing Journal, Vol. 23(6), pp. 555-563 (2005).

222. H. Kittler et al., "*Morphologic changes of pigmented skin lesions: A useful extension of the ABCD rule for dermatoscopy",* Journal of the American Academy of Dermatology, Vol. 40(4),pp. 558-562 (1999).

223. M. Keefe, D.C. Dick, and R.A. Wakeel, "*A study of the value of the seven-point checklist in distinguishing benign pigmented lesions from melanoma",* Clinical and Experimental Dermatology, Vol.15(3), pp. 167-171 (1990).

224. S.W. Menzies, K.A.C., :*An Atlas of Surface Microscopy pf pigmented skin tumors,*" New York: McGraw Hill (1995).

225. Friedman RJ, "*The clinical features of malignant melanoma",* Dermatologic Clinics, Vol.3, pp. 271-283 (1985).

226. Menzies SW, McCarthy WH, "*The morphologic criteria of the pseudopod in surface microscopy",* Archives of Dermatology, Vol. 131, pp. 436-440 (1995).

227. Dolianitis C et al., "*Comparative performance of 4 dermoscopic algorithms by nonexperts for the diagnosis of melanocytic lesion",* Archives of Dermatology, Vol. 141(8), pp. 1008-1014 (2005).

228. G. Betta, G. Fabbrocini, A. Paolillo, P. Sommella, "*Dermoscopic image-analysis system: estimation of atypical pigment network and atypical vascular pattern",* in Proc. of IEEE International Workshop on Medical Measurement and Applications, pp. 63-67 (2006).

229. Gabriella Fabbrocini, Giuseppe Di Leo, et. al, "*Epiluminescence Image Processing for Melanocytic Skin Lesion Diagnosis Based on 7-Point Check-List: A Preliminary Discussion on Three Parameters" ,* The Open Dermatology Journal,Vol. 4, pp. 110-115 (2010).

230. W.V. Stoecker, W.W. Li, and R.H. Moss, "*Automatic detection of asymmetry in skin tumors",* Computerized Medical Imaging and Graphics, Vol.16(3), pp. 191-197 (1992).

231. H. Ganster et al., "*Initial results of automated melanoma recognition",* in *Proceedings of The 9th Scandinavian Conference on Image Analysis*, pp. 1-10 (1995).

232. J. F. Aitken, J.P., D. Battistutta, P. K. O'Rourke, A. C. Green, and N. G. Martin, "*Reliability of computer image analysis of pigmented skin lesions of Australian adolescents",* Journal of cancer, Vol.78(2), pp. 252–257 (1996).

233. Green A, M.N., Pfitzner J, O'Rourke M, Knight N. , "*Computer image analysis in the diagnosis of melanoma",* Journal of American Academy of Dermatology, Vol. 31, pp. 958-964 (1994).

234. Lee, H.C., "*Skin cancer diagnosis using hierarchical neural networks and fuzzy logic",* in *Department of Computer Science*, University of ssouri, Rolla: USA (1994).

235. N.F. Cascinelli, M. Bufalino, R. Zurrida, S. Galimberti, V. Mascheroni, L. Bartoli, C. Clemente, "*Results obtained by using a computerized image analysis system designed as an aid to diagnosis of cutaneous melanoma",* Melanoma Research, Vol.2(3), pp. 163-170 (1992).

236. P. Schmid-Saugeon, "*Symmetry axis computation for almost-symmetrical and asymmetrical objects: Application to pigmented skin lesions",* Medical Image Analysis, Vol. 4(3), pp. 269-282 (2000).

237. E. Claridge et al., "*Shape analysis for classification of malignant melanoma",* Journal of Biomedical Engineering, Vol.14(3), pp. 229-234 (1992).

238. T.K. Lee, D.I. McLean, and M. Stella Atkins, "*Irregularity index: A new border irregularity measure for cutaneous melanocytic lesions",* Medical Image Analysis, Vol.7(1), pp. 47-64 (2003).

239. Ercal, F., et al., "*Neural network diagnosis of malignant melanoma from color images",* IEEE Transactions on Biomedical Engineering, Vol. 41(9), pp. 837-845 (1994).

240. Z. She, and P.J. Fish, "*Analysis of skin line pattern for lesion classification",* Skin Research and Technology, Vol.9(1), pp. 73-80 (2003).

241. S. Gilmore et al.,"*Lacunarity Analysis: A Promising Method for the Automated Assessment of Melanocytic Naevi and Melanoma",* PLoS ONE, Vol.4(10), pp. e7449 (2009).

242. Claridge et al., "*Developing a predictive model of human skin coloring"*, in *Proc. of SPIE Medical Imaging: Physics of Medical Imaging*, pp. 1-12 (1996).

243. Mariam A.Sheha, Amr Sharawy, "*Automatic Detection of Melanoma Skin Cancer using Texture Analysis"*, International Journal of Computer Applications, Vol. 42(20), pp. 22-26 (2012).

244. Kontinen J, RM. MacKie, "*Texture features in the classification of melanocytic lesions",* in Proc. *Ninth International Conference on Image Analysis and Processing,* Florence, Italy: Berlin: Springer, pp. 453-460 (1997).

245. R.M. Haralick, K.Shanmugan, and I. Dinstein, "*Textural features for image classification",* IEEE Transactions on Systems, Man, and Cybernetics, Vol.3(6), pp. 610–621 (1973).

246. D.A. Clausi , "*An analysis of co-occurrence texture statistics as a function of grey level quantization"*, Canadian Journal of Remote Sensing, Vol. 28(1), pp. 45-62 (2002).

247. B.B. Shrestha, Joseph Kam, Keong Chen et.al, "*Detection of atypical texture features in early malignant melanoma",* Skin Research and Technology, Vol. 16(1), pp. 60-65 (2010).

248. Pellacani G, Cucchiara R, Seidenari S, "*Automated Extraction and Description of Dark Areas in Surface Microscopy Melanocytic Lesion Images",* Dermatology,. Vol. 208(1), pp. 21-26 (2004).

249. S.O. Skrovseth et al. "*A computer aided diagnostic system for malignant melanomas",* in 3rd International Symposium on Applied Sciences in Biomedical and Communication Technologies (ISABEL), pp. 1-5 (2010).

250. W.V. Stoecker et al., "*Detection of asymmetric blotches (asymmetric structureless areas) in dermoscopy images of malignant melanoma using relative color",* Skin Research and Technology, Vol.11(3), pp. 179-184 (2005).

251. Hengameh Mirzaalian, Ghassan Hamarneh, "*Learning Features for Streak Detection in Dermoscopic Color Images using Localized Radial Flux of Principal Intensity Curvature",* IEEE workshop on Mathematical Methods for Biomedical Image Analysis, pp.97-101 (2012).

252. W.V. Stoecker et al., "*Detection of granularity in dermoscopy images of malignant melanoma using color and texture features",* Computerized Medical Imaging and Graphics, Vol. 35(2), pp. 144-147 (2011).

253. M.E Celebi et al., "*Automatic detection of blue-white veil and related structures in dermoscopy images",* Computerized Medical Imaging and Graphics, Vol. 32(8), pp. 670-677 (2008).

254. W. Sauerbrei, "*The Use of Resampling Methods to Simplify Regression Models in Medical Statistics"* , Journal of the Royal Statistical Society: Series C (Applied Statistics), Vol. 48(3), pp. 313-329 (1999).

255. Y. Kodratoff, *Introduction to machine learning,* Morgan Kaufmann (2014).

256. M.H.F. Law, Mario AT Jain, Anil K, "*Simultaneous feature selection and clustering using mixture models",* IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.26(9), pp. 1154-1166 (2004).

257. R.J.Kohavi, George H, "*Wrappers for feature subset selection" ,* Artificial intelligence, Vol.97(1),pp. 273-324 (1997).

258. I.E.Guyon, Andre, "*An introduction to variable and feature selection",* The Journal of Machine Learning Research, Vol.3, pp. 1157-1182 (2003).

259. P. Langley, "*Selection of relevant features in machine learning",* Defense Technical Information Center (1994).

260. A.L.L. Blum, Pat, "*Selection of relevant features and examples in machine learning,"* Artificial intelligence, Vol.97(1), pp. 245-271 (1997).

261. H.D. Stoppiglia, Gérard Dubois, Rémi Oussar, Yacine, "*Ranking a random feature for variable and feature selection",* The Journal of Machine Learning Research, Vol.3, pp. 1399-1414 (2003).

262. E.C. Acuna, Frida Gonzalez, Marggie, "*A comparison of feature selection procedures for classifiers based on kernel density estimation",* in Proc. of the International Conference on Computer, Communication and Control technologies, pp. 1-5 (2003).

263. Kira, K.R., Larry A. "*The feature selection problem: Traditional methods and a new algorithm",* in Proc. of National Conference on Artificial Intelligence *AAAI,* pp. 129-134 (1992).

264. Huan Liu, Rudy Setiono, "*A probabilistic approach to feature selection-a filter solution",* in *ICML, pp.1-9 (*1996).

265. C.T. Lazar, Jonatan Meganck, Stijn Steenhoff, David Coletta, Alain et.al, "*A survey on filter techniques for feature selection in gene expression microarray analysis",* IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB), Vol. 9(4), pp. 1106-1119 (2012).

266. G.H.K. John, Ron Pfleger, Karl. "*Irrelevant features and the subset selection problem."* in *Proceedings of the Eleventh International Conference on Machine Learning*, pp. 121-129 (1994).

267. Z.K. Xu, Irwin Lyu, Michael Rung-Tsong Jin, Rong, "*Discriminative semi-supervised feature selection via manifold regularization",* IEEE Transactions on Neural Networks, Vol. 21(7), pp. 1033-1047 (2010).

268. P.M.F. Narendra, Keinosuke, "*A branch and bound algorithm for feature subset selection",* IEEE Transactions on Computers, Vol.100(9),pp. 917-922 (1977).

269. D.E. Golberg, "*Genetic algorithms in search, optimization, and machine learning,"* Addion wesley (1989).

270. J. Kennedy, and R. Eberhart, "*Particle swarm optimization",* in Proceedings of IEEE International Conference on Neural Networks, pp. 1942-1948, (1995).

271. Rami, Khushaba, Ahmed Al Ani, Adel Al-Jumaily, "*Feature Subset Selection Using Differential Evolution*," in *Advances in Neuro-Information Processing,* Springer Berlin Heidelberg, pp.103-110 (2013).

272. Esra Sarac and Selma Ayse Ozel, "*An Ant Colony Optimization Based Feature Selection for Web Page Classification",* The Scientific World Journal, pp. 1-16 (2014).

273. P. Pudil, J. Novovicova, and J. Kittler, "*Floating search methods in feature selection",* Pattern Recognition Letters, Vol. 15(11), pp. 1119-1125 (1994).

274. J. Reunanen, "*Overfitting in making comparisons between variable selection methods",* The Journal of Machine Learning Research, Vol. 3, pp. 1371-1382 (2003).

275. P.P. Somol, Pavel Novovicova, Jana Paclık, Pavel, "*Adaptive floating search methods in feature selection",* Pattern Recognition Letters, Vol. 20(11), pp. 1157-1163 (1999).

276. Y.B. Sun, EJ. CF Delp, "*A comparison of feature selection methods for the detection of breast cancers in mammograms: adaptive sequential floating search vs. genetic algorithm"* , in Proc. of 27th Annual International Conference of the Engineering in Medicine and Biology Society, pp. 6532-6535 (2005).

277. P.A. Devijver and J. Kittler, "*Pattern recognition: A statistical approach",* Vol. 761, Prentice-Hall London (1982).

278. S.C. Nakariyakul, David P, "*An improvement on floating search algorithms for feature subset selection",* Pattern Recognition, Vol.42(9), pp. 1932-1940 (2009).

279. S.D. Stearns, "*On selecting features for pattern classifiers",* in Proceedings of the 3rd International Joint Conference on Pattern Recognition, (1976).

280. S.G. Isabelle Guyon, Massoud Nikravesh, Lofti A.Zadeh , *"Feature Extraction: Foundations and Applications",*"ed. 2006, Netherlands: Springer.

281. M. Kudo and J. Sklansky, "*Comparison of Algorithms that Select Features for Pattern Recognition",* Pattern Recognition, Vol. 33(1), pp. 25-41 (2000).

282. S. Lin, Tsung-Yuan Chou, Shuo-Yan Chen, Shih-Chieh, "*A simulated-annealing-based approach for simultaneous parameter optimization and feature selection of back-propagation networks",* Expert Systems with Applications, Vol.34(2), pp. 1491-1499 (2008).

283. Van Laarhoven, Emile H, "*Simulated annealing: theory and applications,*"Vol. 37. Springer Science & Business Media (1987).

284. R. Haupt, Sue Ellen, "*Practical genetic algorithms",* John Wiley & Sons (2004).

285. C. Huang, Chieh-Jen, "*A GA-based feature selection and parameters optimizationfor support vector machines",* Expert Systems with Applications, Vol.31(2), pp. 231-240 (2006).

286. M.L.P. Raymer, William F Goodman, Erik D Kuhn, Leslie Jain, Anil K, "*Dimensionality reduction using genetic algorithms",* IEEE Transactions on Evolutionary Computation, Vol.4(2), pp. 164-171 (2000).

287. M.B. Dorigo, Mauro, "*Ant colony optimization*," in *Encyclopedia of machine learning*, Springer, pp. 36-39 (2010).

288. Mehdi Hosseinzadeh Aghdam, Nasser Ghasem-Aghaee, Mohammad Ehsan Basiri, "*Text feature selection using ant colony optimization",* Expert Systems with Applications, Vol.36(3), pp. 6843-6853 (2009).

289. Hiram A. Firpi, Erik Goodman, "*Swarmed Feature Selection",* in Proceedings of the 33rd Applied Imagery Pattern Recognition Workshop, pp. 112-118 (2004).

290. Bharathi P.T, P. Subashini, "*Differential Evolution and Genetic Algorithm based feature subset selection for recognition of river ice types",* Journal of Theoretical & Applied Information Technology, Vol.67(1), pp. 254-262 (2014).

291. Swagatam Das, Sayan Maity, Bo-Yang Qu, P.N. Suganthan, "*Real-parameter evolutionary multimodal optimization—a survey of the state-of-the-art,*" Swarm and Evolutionary Computation, Vol.1(2), pp. 71-88 (2011).

292. Trung Thanh Nguyen, Shengxiang Yang, Juergen Branke, "*Evolutionary dynamic optimization: A survey of the state of the art",* Swarm and Evolutionary Computation, Vol.6, pp. 1-24 (2012).

293. Goldberg, John H, "*Genetic algorithms and machine learning",* Machine Learning, Vol.3(2), pp. 95-99 (1988).

294. M.D. Vose, "*The simple genetic algorithm: foundations and theory,*" Vol. 12., MIT press (1999).

295. Hamidreza Rashidy Kanan, Karim Faez, "*GA-based optimal selection of PZMI features for face recognition",* Applied Mathematics and Computation, Vol.205(2), pp. 706-715 (2008).

296. Jianjiang Lu, Tianzhong Zhao, Yafei Zhang, "*Feature selection based-on genetic algorithm for image annotation",* Knowledge-Based Systems, Vol.21(8), pp. 887-891 (2008).

297. O. Cordon, S. Damas, J. Santamaría et al., "*Feature-based image registration by means of the CHC evolutionary algorithm",* Image and Vision Computing, Vol.24(5), pp. 525-533 (2006).

298. Larry J. Eshelman, "*The CHC Adaptive Search Algorithm: How to Have Safe Search When Engaging in Nontraditional* Genetic Recombination" , in proc. of the First Workshop on Foundations of Genetic Algorithms, p. 265 (1990).

299. II Seok Oh, Jin-Seon Lee, Byung-Ro Moon, "*Hybrid genetic algorithms for feature selection",* IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 26(11), pp. 1424-1437 (2004).

300. Ahmed Al-Ani, "*Feature subset selection using ant colony optimization,*" International Journal of Computational Intelligence, Vol.2(1), pp. 53-58 (2005).

301. R. Eberhart, James Kennedy, "*A new optimizer using particle swarm theory",* in Proceedings of the sixth international symposium on micro machine and human science, pp. 39-43 (1995).

302. Mojtaba Ahmadieh Khanesar et al, "*A novel binary particle swarm optimization,*" in Mediterranean Conference on Control & Automation, pp. 1-6 (2007).

303. Yan Jiang, Tiesong Hu, Chong Chao Huang, Xianing Wu, "*An improved particle swarm optimization algorithm*", in Applied Mathematics and Computation, pp. 231-239 (2008).

304. Rainer Storn, Kenneth Price, "*Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces",* Journal of global optimization, Vol.11(4), pp. 341-359 (1997).

305. Rainer Storn, Kenneth Price, "*Differential evolution-a simple and efficient adaptive scheme for global optimization over continuous spaces",* Vol. 3. ICSI Berkeley (1995).

306. Dervis Karaboga, Selcuk Okdem, "*A simple and global optimization algorithm for engineering problems: differential evolution algorithm",* Turk Journal of Electrical Engineering, Vol.12(1) pp. 53-60 (2004).

307. Kenneth Price et al., "*Differential evolution: a practical approach to global optimization",* Springer Science & Business Media (2006).

308. S Das, Ponnuthurai N. Suganthan, "*Differential evolution: a survey of the state-of-the-art",* IEEE Transactions on Evolutionary Computation, Vol.15(1), pp. 4-31 (2011).

309. Rami N. Khushaba, Ahmed Al-Ani, Adel Al-Jumaily, "*Differential evolution based feature subset selection*. in *Pattern Recognition",* in proc. of 19th International Confernece on Pattern Recongnition, pp. 1-4 (2008).

310. T. Bhadra et al., "*Differential Evolution Based Optimization of SVM Parameters for Meta Classifier Design",* Procedia Technology, Vol. 4, pp. 50-57 (2012).

311. Kingshuk Chakravarty, Diptesh Das, Aniruddha Sinha, Amit Konar, "*Feature selection by differential evolution algorithm-a case study in personnel identification",* in IEEE Congress on Evolutionary Computation (CEC), pp. 892-899, (2013).

312. Ahmed Al-Ani, Akram Alsukker, Rami N Khushaba, "*Feature subset selection using differential evolution and a wheel based search strategy",* Swarm and Evolutionary Computation, Vol.9, pp. 15-26 (2013).

313. J. Vesterstrøm, Rene Thomsen, "*A comparative study of differential evolution, particle swarm optimization, and evolutionary algorithms on numerical benchmark problems",* in Proc. of IEEE Congress on Evolutionary Computation, pp. 1980-1987 (2004).

314. Swagatam Dass, Amit Konar, Udau K. Chakraborty, "*Two improved differential evolution schemes for faster global search*,"in Proc. of the 7th annual conference on Genetic and evolutionary computation, pp. 991-998 (2005).

315. S.E. Umbaugh, "*Computer imaging : digital image analysis and processing*,"Baco Raton CRC press (2005).

316. S. Aksoy, R.M. Haralick, "*Feature normalization and likelihood-based similarity measures for image retrieval*," Pattern Recognition Letters, Vol.22(5), pp. 563-582 (2001).

317. T. HC, *Testing for normality*, New York: CRC Press (2002).

318. RO Duda, PE Hart, DG Stork, "*Pattern Classification",* 2nd ed, New York: John Wiley Sons (2000).

319. T. Randen, and J.H. Husoy, "*Filtering for texture classification: a comparative study*", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.21(4), pp. 291-310 (1999).

320. T. Cover, and P. Hart, "*Nearest neighbor pattern classification*", IEEE Transactions on Information Theory, Vol.13(1), pp. 21-27 (1967).

321. B.V. Dasarathy, "*Nearest neighbor (NN) norms: NN pattern classification techniques*"(1991).

322. S. Dreiseitl et al., "*A Comparison of Machine Learning Methods for the Diagnosis of Pigmented Skin Lesions*", Journal of Biomedical Informatics, Vol.34(1), pp. 28-36 (2001).

323. E. El-Kwae et al., "*Detection of suspected malignant patterns in three-dimensional magnetic resonance breast images*", Journal of Digital Imaging, Vol.11(2), pp. 83-93 (1998).

324. H. Handels et al., "*Feature selection for optimized skin tumor recognition using genetic algorithms*", Artificial Intelligence in Medicine, Vol.16(3), pp. 283-297 (1999).

325. BD Ripley, "*Pattern recognition and neural networks*," Cambridge: Cambridge University Press (1996).

326. A. Sboner et al., "*A multiple classifier system for early melanoma diagnosis*," Artificial Intelligence in Medicine, Vol.27(1),pp. 29-44 (2003).

327. SL Salzberg, "*C4.5: programs for machine learning*," Los Altos, CA: Morgan Kaufmann (1993).

328. J.Gehrke, "*Classification and Regression Trees*," in Encyclopedia of Data Warehousing and Mining, Second Edition, IGI Global, pp. 192-195 (2009).

329. W.Y. Loh, "*Classification and regression trees*", Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, Vol.1(1), pp. 14-23 (2011).

330. N. Speybroeck, "*Classification and regression trees*", International Journal of Public Health, Vol.57(1), pp. 243-246 (2012).

331. V.K. Xindong Wu, "*CART : Classification and regression trees*," in *The Top Ten Algorithms in Data Mining*, Boca Raton : CRC Press, pp. 179-203 (2009).

332. Z.M. Kokol P, Stiglic MM, I. Malèiae, "*The limitations of decision trees and automatic learning in real world medical decision making*", Journal of Medical Systems, Vol.21(6), pp. 403-415 (1997).

333. D.E.Clark, "*Computational Methods for Probabilistic Decision Trees*", Computers and Biomedical Research, Vol.30(1), pp. 19-33 (1997).

334. Tim Oates, David Jensen, "*Large Datasets Lead to Overly Complex Models: An Explanation and a Solution*", in *Proc. of the 4th International Conference on Knowledge Discovery and Data Mining*, pp. 294-298 (1998).

335. F.E. Harrell et al., "*Regression modelling strategies for improved prognostic prediction*", Statistics in Medicine, Vol.3(2), pp. 143-152 (1984).

336. D.J. Spiegelhalter, "*Probabilistic prediction in patient management and clinical trials*", Statistics in Medicine, Vol.5(5), pp. 421-433 (1986).

337. D.Altman, "*Practical statistics for medical research" ,* London/Newyork: Chapman & Hall (1991).

338. S.S. Cross, R.F. Harrison, and R.L. Kennedy, "*Introduction to neural networks,"* The Lancet, Vol. 346(8982), pp. 1075-1079 (1995).

339. C.M. Bishop, "*Neural networks for Pattern Recognition,"*Oxford U.K: Clarendon (1995).

340. Skapura et al, "*Neural Networks, Algorithms, Applications and Programming Techniques,"*Addison-Wesley Publishing Company (2002).

341. R.M. Neal, "*Bayesian learning for neural networks*,"New York: Springer (1996).

342. Christopher M. Bishop, "*Neural Networks for pattern recognition"*, London: Oxford Univ. Press (1995).

343. L. Fausett, "*Fundamentals of neural network: architectures, algorithms and applications*,"Cliffs, Prentice Hall (1994).

344. J. Abdul Jaleel, Sibi Salim, Aswin R.B., "Artificial Neural Network based Detection of Skin Cancer", International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, pp. 200-205 (2012).

345. Hagan MT, D.H., Beal MH, *Neutral network design*2003: PWS, Beston.

346. Saeed Ayat, Hojjat A. Farahani, Mehdi Aghamohamadi, Mahmood Alian, Somayeh Aghamohamadi, "*A comparison of artificial neural networks learning algorithms in predicting tendency for suicide",* Neural Computing & Applications,. Vol. 23, pp. 1381-1386 (2013).

347. D. Marquardt, "*An algorithm for Least-Squares Estimation of Nonlinear Parameters,"* SIAM Journal on Applied Mathematics, Vol.11(2), pp. 431-441 (1963).

348. M.F. Moller, "*A scaled conjugate gradient algorithm for fast supervised learning,"* Neural Networks, Vol. 6, pp. 525-533 (1993).

349. VN Vapnik, "*Statistical learning theory*,"New york: Wiley (1998).

350. VN Vapnik, "*The nature of statistical learning theory",* 2nd ed, New York: Springer (2000).

351. Nello Cristianini, John Shawe Taylor, "*An introduction to support vector machines and other kernel-based learning methods*," Cambridge: Cambridge University Press (2000).

352. Chuan-Min Zhai, Ji-Xiang Du., "*Applying extreme learning machine to plant species identification" ,* in International Conference on Information and Automation, pp. 879-884 (2008).

353. A.Bharathi, A.M. Natarajan, "*Cancer Classification using Modified Extreme Learning Machine based on ANOVA Features",* European Journal of Scientific Research, Vol.58(2), pp. 156-165 (2011).

354. Nathalie Japkowicz, "*Learning from Imbalanced Data Sets: A Comparison of Various Strategies",* in AAAI Technical Report (2000).

355. Ronaldo C. Prati et al., "*Class Imbalances versus Class Overlapping: An Analysis of a Learning System Behavior,*" MICAI 2004: Advances in Artificial Intelligence, pp. 312-321 (2004).

356. M. Schumacher, Hollander N, Sauerbrei W, "*Resampling and cross-validation techniques: a tool to reduce bias caused by model building?"*, Statistics in Medicine, Vol. 16(24), pp. 2813-2827 (1997).

357. R.W.Johnson, "*An Introduction to the Bootstrap",* Teaching Statistics, Vol.23(2), pp. 49-54 (2001).

358. B. Efron, "*Estimating the Error Rate of a Prediction Rule: Improvement on Cross-Validation",* Journal of the American Statistical Association, Vol.78(382), pp. 316-331 (1983).

359. Patwardhan, Shuangshuang Dhawan, P. Atam, "*Multi-spectral image analysis and classification of melanoma using fuzzy membership based partitions,*" Computerized Medical Imaging and Graphics, Vol. 29(4), pp. 287-296 (2005).

360. C. Garbe et al., "*Primary cutaneous melanoma. Identification of prognostic groups and estimation of individual prognosis for 5093 patients",* Cancer, Vol. 75(10), pp. 2484-2491 (1995).

361. P. Rubegni et al., "*Digital Dermoscopy Analysis and Artificial Neural Network for the Differentiation of Clinically Atypical Pigmented Skin Lesions: A Retrospective Study,*"Journal of Investigative Dermatology, Vol.119(2), pp. 471-474 (2002).

362. R. Marchesini et al., "*In vivo Spectrophotometric Evaluation of Neoplastic and Non-plastic Skin Pigmented Lesions. II Discriminant Analysis between Nevus and Melanoma,*" Photochemistry and Photobiology, Vol.55(4), pp. 515-522 (1992).

363. I. Maglogiannis, S. Pavlopoulos, and D. Koutsouris, "*An integrated computer supported acquisition, handling, and characterization system for pigmented skin lesions in dermatological images",* IEEE Transactions on Information Technology in Biomedicine, Vol.9(1), pp. 86-98 (2005).

364. T. Schindewolf et al., "*Classification of melanocytic lesions with color and texture analysis using digital image processing",* Analytical and quantitative cytology and histology / the International Academy of Cytology [and] American Society of Cytology, Vol.15(1), pp. 1-11 (1993).

365. R.J. Stanley et al., "*A fuzzy-based histogram analysis technique for skin lesion discrimination in dermatology clinical images",* Computerized Medical Imaging and Graphics, Vol.27(5), pp. 387-396 (2003).

366. J. Chen et al., "*Colour analysis of skin lesion regions for melanoma discrimination in clinical images",* Skin Research and Technology, Vol.9(2), pp. 94-104 (2003).

367. R.J. Stanley, W.V. Stoecker, and R.H. Moss, "*A relative color approach to color discrimination for malignant melanoma detection in dermoscopy images",* Skin Research and Technology, Vol.13(1), pp. 62-72 (2007).

368. R. Joe Stanley et al., "*A basis function feature-based approach for skin lesion discrimination in dermatology dermoscopy images"* , Skin Research and Technology, Vol.14(4), pp. 425-435 (2008).

369. M. Binder et al., "*Application of an artificial neural network in epiluminescence microscopy pattern analysis of pigmented skin lesions: a pilot study",* British Journal of Dermatology, Vol.130(4), pp. 460-465 (1994).

370. R. Husemann et al., "*Computerized diagnosis of skin cancer using neural networks",* Melanoma Research, pp. 1052-1063 (1996).

371. M.K. Binder, H. Seeber, A. Steiner, A. Pehamberger, H. Wolff, K., "*Epiluminescence microscopy-based classification of pigmented skin lesions using computerized image analysis and an artificial neural network",* Melanoma Research, Vol. 8(3), pp. 261-266 (1998).

372. H. Handels et al., "*Computer-Supported Diagnosis of Melanoma in Profilometry,*" Methods of Information in Medicine, Vol. 38, pp. 43-49 (1999).

373. Gilmore S et al., "*A support vector machine for decision support in melanoma recognition,*"Experimental Dermatology, Vol.19(9), pp. 830-835 (2010).

374. Menzies SW et al., "*Automated instrumentation for the diagnosis of invasive melanoma: image analysis of oil epiluminescence microscopy,*"Skin Cancer and UV-Radiation, Berlin, Springer-Verlag, pp. 1064-1070 (1997).

375. S. Seidenari, G. Pellacani, and P. Pepe, "*Digital videomicroscopy improves diagnostic accuracy for melanoma",* Journal of the American Academy of Dermatology, Vol. 39(2), pp. 175-181 (1998).

376. S. Seidenari, G. Pellacani, and A. Giannetti, "*Digital videomicroscopy and image analysis with automatic classification for detection of thin melanomas",* Melanoma Research, Vol.9(2), pp. 163-172 (1999).

377. H. Ganster et al., "*Automated melanoma recognition",* IEEE Transactions on Medical Imaging, Vol. 20(3), pp. 233-239 (2001).

378. Y. Vergouwe et al., "*Validity of prognostic models: when is a model clinically useful?,*" Seminars in urologic oncology, Vol.20(2), pp. 96-107 (2002).

379. Altman, Patrick, "*What do we mean by validating a prognostic model?*"Statistics in Medicine, Vol.19(4), pp. 453-473 (2000).

380. Ammara Masood, Adel Al Jumaily, "*Fuzzy C mean Thresholding based Level Set for Automated Segmentation of Skin Lesions,*" Journal of Signal and Information Processing, Vol.4(3B), pp. 66-71 (2013).

381. Ammara Masood, Adel Al Jumaily, Yashar Maali, "*Level Set Initialization Based on Modified Fuzzy C Means Thresholding for Automated Segmentation of Skin Lesions*,"in *Neural Information Processing*, pp. 341-351 (2013).

382. Ammara Masood, Adel Al Jumaily, Azadeh Noori Hoshyar, Omama Masood, "*Automated segmentation of skin lesions: Modified Fuzzy C mean thresholding based level set method"*, in Proc. of 16th International Multi Topic Conference (INMIC),pp. 201-206, (2013).

383. Ammara Masood, Adel Al Jumaily, "*Integrating soft and hard threshold selection algorithms for accurate segmentation of skin lesion*," in Proc. of Middle East Conference on Biomedical Engineering , pp. 83-86 (2014)

384. S. Raut et al., "*Image Segmentation - A State-Of-Art Survey for Prediction"* in Proc. International Conference on Advanced Computer Control, pp. 420-424 (2009).

385. S.N. Deepa et. al., "*A survey on artificial intelligence approaches for medical image classification",* Indian Journal of Science and Technology, Vol.4(11), pp. 1583-1595 (2011).

386. P. K. Sahoo, S. Soltani, A. K.C.Wong , "*A survey of thresholding techniques*", Computer Vision, Graphics, and Image Processing, Vol. 41(2), pp. 233-260(1988).

387. Joan S. Weszka, "*A survey of threshold selection techniques*", Computer Graphics and Image Processing, Vol. 7(2), pp. 259-265(1978).

388. K. Korotkov, Garcia Rafael, "*Computerized analysis of pigmented skin lesions: A review*" , Artificial Intelligence in Medicine, Vol.56(2) , pp. 69-90 (2012).

389. C.A. Glasbey, "*An Analysis of Histogram-Based Thresholding Algorithms*" , CVGIP: Graphical Models and Image Processing, Vol. 55(6) , pp. 532-537(1993).

390. Isard, "*Active Contours*", Springer Verlag (1998).

391. S.Osher, R. Fedkiw, "*Level Set Methods and Dynamic Implicit Surfaces*", New York: Springer-Verlag (2002).

392. Chunming Li et al., "*Level set evolution without re-initialization: a new variational formulation*", in Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 430-436 (2005).

393. Petro Perona, Jitendra Malik "*Scale-space and edge detection using anisotropic diffusion*",IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.12(7), pp. 629-639 (1990).

394. Nascimento JC1, Marques JS. , "*Adaptive snakes using the EM algorithm*", IEEE Transactions on Image Processing,Vol. 14(11), pp. 1678-1686 (2005).

395. Bing Nan Lia, Chee Kong Chuib, Stephen Change, S.H. Ong, "*Integrating spatial fuzzy clustering with level set methods for automated medical image segmentation*", Computers in Biology and Medicine,Vol. 41(1), pp. 1-10(2011).

396. Bing Nan Li et al., "I*ntegrating FCM and Level Sets for Liver Tumor Segmentation*", in 13th International Conference on Biomedical Engineering, pp. 202-205 (2009).

397. Dongju Liu, Jian Yu, "*Otsu Method and K-means*", in Ninth International Conference on Hybrid Intelligent Systems, pp. 344-349 (2009).

398. Do Hyun Chung, G. Sapiro, "*Segmenting Skin Lesions with Partial-Differential-Equations-Based Image Processing Algorithms*" , IEEE Transactions on Medical Imaging,Vol. 19(7) ,pp. 763-767 (2000).

399. Chivukula A. Murthy, Sankar K. Pal, "*Fuzzy thresholding: mathematical framework, bound functions and weighted moving average technique*", Pattern Recognition Letters, Vol.11(3) , pp. 197-206 (1990).

400. C.V. Jawahar, P.K. Biswas, A.K. Ray, "*Analysis of fuzzy thresholding schemes*", Pattern Recognition, Vol.33(8), pp. 1339-1349 (2000).

401. O. Demirkaya, "*Image Processing with Matlab: application in medicine and biology*", Boca Raton : CRC Press, (2009).

402. Nobuyuki Otsu, "*A Threshold Selection Method from Gray-Level Histograms*", IEEE Transactions on Systems, Man and Cybernetics,Vol. 9(1) ,pp. 62-66 (1979).

403. Jun Zhang, and Hu Jinglu, "*Image Segmentation Based on 2D Otsu Method with Histogram Analysis*", in International Conference on Computer Science and Software Engineering, pp. 105-108 (2008).

404. Hong Yan, "*Unified formulation of a class of image thresholding techniques*", Pattern Recognition, Vol. 29(12), pp. 2025-2032 (1996).

405. H.D. Cheng , Yen-Hung Chen, "*Fuzzy partition of two-dimensional histogram and its application to thresholding*", Pattern Recognition, Vol. 32(5),pp. 825-843(1999).

406. O. J. Tobias , R. Seara , "*Image segmentation by histogram thresholding using fuzzy sets*" , IEEE Transactions Image Processing, on,Vol. 11(12) ,pp. 1457-1465 (2002).

407. Liang-Kai Huang, Mao-Jiun J. Wang, "*Image thresholding by minimizing the measures of fuzziness*", Pattern Recognition, Vol. 28(1), pp. 41-51 (1995).

408. J. Kittler, J. Illingworth, "*Minimum error thresholding*", Pattern Recognition,Vol. 19(1), pp. 41-47 (1986).

409. J.N.Kapur, P.K. Sahoo, A.K.C. Wong, "*A new method for gray-level picture thresholding using the entropy of the histogram*", Computer Vision, Graphics, and Image Processing, Vol. 29(3), pp. 273-285 (1985).

410. Santiago Aja-Fernández, Gonzalo Vegas-Sánchez-Ferrero ; Miguel A. Martín Fernández, "Soft thresholding for medical image segmentation", in Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 4752-4755 (2010).

411. T.F. Chan , L.A. Vese, "*Active contours without edges*", IEEE Transactions on Image Processing, Vol. 10(2), pp. 266-277 (2001).

412. M. Muthu Rama Krishnana, Chandan Chakrabortya, Ranjan Rashmi Paulb, Ajoy K. Ray , "*Hybrid segmentation, characterization and classification of basal cell nuclei from histopathological images of normal oral mucosa and oral submucous fibrosis*" , Expert Systems with Applications, Vol. 39(1), pp. 1062-1077(2012).

413. Yener, "*Automated cancer diagnosis based on histopathological images: a systematic survey*" Rensselaer Polytechnic Institute, Department of Computer Science (2009).

414. AD Belsare, MM Mushrif, "*Histopathological image analysis using image processing techniques: An overview*", Signal & Image Processing,Vol. 3(4), pp. 23-36 (2012).

415. M. Nareash Babu, Vamsi K. Madasu ; M. Hanmandlu ; S. Vasikarla, "*Histo-pathological image analysis using OS-FCM and level sets*", in IEEE 39th Applied Imagery Pattern Recognition Workshop, pp. 1-8 (2010).

416. Bauer J1, Metzler G, Rassner G, Garbe C, Blum A., "*Dermatoscopy turns histopathologist's attention to the suspicious area in melanocytic lesions*", Archives of Dermatology, Vol.137(10), pp. 1338-1340, (2001).

417. Verhaegen PD, van Zuijlen PP, Pennings NM, van Marle J, Niessen FB, van der Horst CM, Middelkoop E. , "*Differences in collagen architecture between keloid, hypertrophic scar, normotrophic scar, and normal skin: An objective histopathological analysis*" ,Wound Repair and Regeneration, Vol. 17(5) , pp. 649-656 (2009).

418. M. El-Domyati, S. Attia, F. Saleh, D. Brown, D. E. Birk, F. Gasparro, H. Ahmad,J. Uitto, "*Intrinsic aging vs. photoaging: a comparative histopathological, immunohistochemical, and ultrastructural study of skin*" , Experimental Dermatology,. Vol. 11(5), pp. 398-405(2002).

419. S. Ben Chaabane et al., "Color image segmentation using automatic thresholding and the fuzzy C-means techniques" , in Proc. of 14th IEEE Mediterranean Electrotechnical Conference, pp. 857-861 (2008).

420. P. Zhenkui, Z. Yanli, and L. Zhen, "*Image segmentation based on Differential Evolution algorithm*" in Proc. of International Conference on Image Analysis and Signal Processing, pp. 48-51 (2009).

421. J.M. Geusebroek, A.W.M. Smeulders, and J. van de Weijer, "*Fast anisotropic Gauss filtering*", IEEE Transactions on Image Processing, Vol. 12(8), pp. 938-943 (2003).

422. J.-F.H. Yang, Shu-Sheng Chung, Pau-Choo, "*Color image segmentation using fuzzy C-means and eigenspace projections*", Signal Processing, Vol. 82(3), pp. 461-472 (2002).

423. T.F. Chan, B.Y. Sandberg, and L.A. Vese, "*Active Contours without Edges for Vector-Valued Images*", Journal of Visual Communication and Image Representation, Vol. 11(2), pp. 130-141 (2000).

424. V.M. Grau, AUJ Alcaniz, M Kikinis, Ron Warfield, Simon K, "*Improved watershed transform for medical image segmentation using prior information*", IEEE Transactions on Medical Imaging, Vol. 23(4), pp. 447-458 (2004).

425. H.J. Fatakdawala, Xu Basavanhally, A. Bhanot, G. Ganesan, S. Feldman, M. Tomaszewski, J. E. Madabhushi, "*Expectation Maximization-Driven Geodesic Active Contour With Overlap Resolution (EMaGACOR): Application to Lymphocyte Segmentation on Breast Cancer Histopathology*", IEEE Transactions on Biomedical Engineering, Vol. 57(7), pp. 1676-1689 (2010).

426. Ammara Masood, Adel Al-Jumaily, "*Semi Advised SVM with Adaptive Differential Evolution Based Feature Selection for Skin Cancer Diagnosis*", Journal of Computer Science & Communications, Vol. 3(11), pp. 184-190 (2015).

427. Ammara Masood, Adel Al-Jumaily, Khairul Anam, "T*exture Analysis Based Automated Decision Support System for Classification of Skin Cancer Using SA-SVM*", in Neural Information Processing, Springer International Publishing, pp. 101-109 (2014).

428. Ammara Masood, Adel Al-Jumaily, Khairul Anam, "*Adaptive Differential Evolution Based Feature Selection and Parameter Optimization for Advised SVM*

Classifier", in Neural Information Processing, Springer International Publishing, pp. 401–410, (2015).

429. Ammara Masood, Adel Al-Jumaily, Khairul Anam, "*Differential evolution based advised SVM for histopathalogical image analysis for skin cancer detection*", in Proc. of 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 781-784 (2015).

430. Ammara Masood, Adel Al-Jumaily, Yee Mon Aung, "*Scaled Conjugate Gradient based Decision Support System for Automated Diagnosis of Skin Cancer*", in 11th IASTED International Conference on Biomedical Engineering, Zurich, Switzerland, paper Id.818-020 (2014).

431. A.F. Yang L, Lonnestad T, Grottum P, "*Methods to estimate areas and perimeters of blob-like objects: a comparison*", in Proceedings of the IAPR workshop on machine vision applications, pp. 27227-6 (1994).

432. W.K. Pratt, "*Digital Image Processing*", PIKS Scientific Inside, John Wiley Sons (2007).

433. T. Gevers, A. Smeulders, "Color based object recognition Image Analysis and Processing", Springer Berlin / Heidelberg, pp. 319-326 (1997).

434. A.W.M.S. Theo Gevers, "*Color-based object recognition*", Pattern Recognition, pp. 453-464 (1999).

435. M.M. Galloway, "*Texture analysis using gray level run lengths*", Computer Graphics and Image Processing, pp. 172-179 (1975).

436. A.S. Chu, C. M. Greenleaf, "*Use of gray value distribution of run lengths for texture analysis*", Pattern Recognition Letters, Vol. 11(6): pp. 415-419 (1990).

437. B.V.H. Dasarathy, Edwin B, "*Image characterizations based on joint gray level—run length distributions*", Pattern Recognition Letters, Vol. 12(8), pp. 497-502 (1991).

438. R.K M. Amadasun, "*Textural Features Corresponding to Textural Properties*", IEEE Transactions on System, Man Cybernetics, Vol. 19(5), pp. 1264-1274 (1989).

439. C.W.S. Anderson, E. A. Shamsunder, "*Multivariate autoregressive models for classification of spontaneous electroencephalographic signals during mental tasks*", IEEE Transactions on Biomedical Engineering, Vol. 45(3), pp. 277-286 (1998).

440. J.G. Proakis, "*Digital Signal Processing Principles, Algorithms, and Applications*", New Jersey, Prentice-Hall (1996).

441. S.G. Mallat, "*A theory for multiresolution signal decomposition: the wavelet representation*", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 11(7), pp. 674-693 (1989).

442. C.S.G. Burrus, Ramesh A Guo, Haitao, "*Introduction to wavelets and wavelet transforms*", a primer, USA, Prentice Hall (1997).

443. P.S. Jackman, Da-Wen Allen, Paul, "*Comparison of various wavelet texture features to predict beef palatability*", Meat science, Vol. 83(1), pp. 82-87 (2009).

444. E. Avci, "*Comparison of wavelet families for texture classification by using wavelet packet entropy adaptive network based fuzzy inference system*", Applied Soft Computing, Vol. 8(1), pp. 225-231 (2008).

445. I. Daubechies, "T*en lectures on wavelets*", Vol. 61 (1992).

446. B.S.M. Manjunath, Wei-Ying, "*Texture features for browsing and retrieval of image data*", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 18(8), pp. 837-842 (1996).

447. M.M.R.C. Krishnan, Chandran Ray, Ajoy Kumar, "*Wavelet based texture classification of oral histopathological sections*", International Journal of Microscopy, Science, Technology, Applications and Education, Vol. 2(4), pp. 897-906 (2010).

448. R.R.M. Coifman, Yves Wickerhauser, "*Wavelet analysis and signal processing*", in In Wavelets and their Applications, Citeseer (1992).

449. R.N.K. Khushaba, S. Lal, S. Dissanayake, "*Driver Drowsiness Classification Using Fuzzy Wavelet-Packet-Based Feature-Extraction Algorithm*", IEEE Transactions on Biomedical Engineering, Vol. 58(1), pp. 121-131 (2011).

450. Rami Khushaba, Ahmed Al-Ani, Adel Al-Jumaily, "*Feature subset selection using differential evolution and a statistical repair mechanism*", Expert Systems with Applications, Vol. 38(9), pp. 11515-11526 (2011).

451. Sk. Minhazul Islam ,Swagatam Das , Saurav Ghosh ,Subhrajit Roy , Ponnuthurai Nagaratnam Suganthan , "*An adaptive differential evolution algorithm with novel mutation and crossover strategies for global numerical optimization*" , IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, Vol.42(2), pp. 482-500 (2012).

452. P. Kumar, P. Millie, "*A Self Adaptive Differential Evolution Algorithm for Global Optimization*", in Proc. of Swarm, Evolutionary, and Memetic Computing, pp. 103-110 (2010).

453. Carlos A. Coello, "A comparative study of differential evolution variants for global optimization", in Proc. of the 8th annual conference on Genetic and evolutionary computation, pp. 485-492 (2006).

454. Ferrante Neri, Ville Tirronen, "Recent advances in differential evolution: a survey and experimental analysis", Artificial Intelligence Review, Vol.33(1-2), pp. 61-106 (2010).

455. Satoshi Yamaguchi, "An automatic control parameter tuning method for differential evolution", Electrical Engineering in Japan, Vol.174(3), pp. 25-33 (2011).

456. A.K. Qin, Ponnuthurai N. Suganthan, "Self-adaptive differential evolution algorithm for numerical optimization" in Proc. of the IEEE Congress on Evolutionary Computation, pp. 1785-1791 (2005).

457. Sambarta Dasgupta, S.Das, Arijit Biswas, Ajith Abraham, "*On stability and convergence of the population-dynamics in differential evolution*" , AI Communications,Vol. 22(1), pp. 1-20 (2009).

458. G. Jeyakumar, C. Shanmugavelayutham, "*Analyzing the Explorative Power of Differential Evolution Variants on Different Classes of Problems*", in Swarm, Evolutionary, and Memetic Computing, pp. 95-102 (2010).

459. Xin Yao, Yong Liu , Guangming Lin, "*Evolutionary programming made faster*" , IEEE Transactions on Evolutionary Computation,Vol. 3(2) , pp. 82-102 (1999).

460. Matthieu Weber, Ville Tirronen, Ferrante Neri, "*Scale factor inheritance mechanism in distributed differential evolution*", Soft Computing,Vol. 14(11), pp. 1187-1207 (2010).

461. Bharathi P ,P. Subashini, "*Optimal Feature Subset Selection Using Differential Evolution and Extreme Learning Machine*", International Journal of Science and Research, Vol. 3(7), pp. 1898-1905 (2012).

462. Rami N. Khushaba, Akram AlSukker ,Ahmed Al-Ani, Adel Al-Jumaily , Albert Y. Zomaya , "*A novel swarm based feature selection algorithm in multifunction myoelectric control*" , Journal of Intelligent and Fuzzy Systems, Vol. 20(4-5), pp. 175-185 (2009).

463. Ammara Masood, Adel Al-Jumaily, "*SA-SVM based automated diagnostic system for skin cancer*", in Sixth International Conference on Graphic and Image Processing (ICGIP 2014).

464. Ammara Masood, Adel Al Jumaily, Khairul Anam, "*Self-Supervised Learning Model for Skin Cancer Diagnosis",* in 7th Annual International IEEE EMBS Conference on Neural Engineering , pp. 1012-1015 (2014).

465. Fabien Lauer, Gérard Bloch, " *Incorporating prior knowledge in support vector machines for classification: A review*" , Neurocomputing, Vol. 71(7-9), pp. 1578-1594 (2008).

466. Yiqiang Zhana, Dinggang Shen, "*An adaptive error penalization method for training an efficient and generalized SVM*" , Pattern Recognition, Vol. 39(3), pp. 342-350 (2006).

467. P. Williams, Li S. ,Feng, Jianfeng ,Wu Si, "*A geometrical method to improve performance of the support vector machine*" , IEEE Transactions on Neural Networks,Vol. 18(3), pp. 942-947 (2007).

468. Hyrum S. Anderson, Maya R. Gupta, "*Expected kernel for missing features in support vector machines*", in IEEE Statistical Signal Processing Workshop (SSP), pp. 285-288 (2011).

469. Tasadduq Imam, Kai Ming Ting, Joarder Kamruzzaman, "z-SVM: an SVM for improved classification of imbalanced data" , in AI 2006: Advances in Artificial Intelligence, pp. 264-273 (2006).

470. Kunlun Li, Xuerong Luo, Ming Jin, "*Semi-supervised Learning for SVM-KNN*", Journal of Computers, Vol.5 , pp. 671-678 (2010).

471. Yashar Maali,Adel Al-Jumaily, "*Self-advising support vector machine*", Knowledge-Based Systems, Vol.52(0), pp. 214-222 (2013).

472. Zhi Hua Zhou, De-Chuan Zhan, Qiang Yang, "Semi-supervised learning with very few labeled training examples", in Proc. of the National Conference on Artificial Intelligence, pp. 675-680 (2007).

473. Kamal Nigam, Andrew Kachites Mccallum, Sebastian Thrun, Tom Mitchell, "*Text Classification from Labeled and Unlabeled Documents using EM*" ,Mach. Learn", Vol. 39(2-3),pp. 103-134 (2000).

474. Ronan Collobert, Fabian Sinz, Jason Weston, Léon Bottou, "*Large Scale Transductive SVMs*" , Journal of Machine Learning Research,Vol. 7, pp. 1687-1712 (2006).

475. Avrim Blum, Shuchi Chawla, "*Learning from Labeled and Unlabeled Data using Graph Mincuts*", in International Conference on Machine Learning, (2001).

476. A.B.T. Mitchell, "*Combining labeled and unlabeled data with co-training*," in Proc. of the 11th Annual Conference on Computational learning theory, pp. 92-100 (1998).

477. S. Sengupata, "*Semi-Supervised Classification using Support Vector Machine*", (2010).

478. O Chapelle, A Zien, "S*emi-supervised classification by low density separation*" in International Workshop on Artificial Intelligence (2005).

479. Pascal Vincent, Hugo Larochelle, Isabelle Lajoie, Yoshua Bengio, Pierre-Antoine Manzagol , "*Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion*" , Journal of Machine Learning Research,Vol. 11, pp. 3371-3408 (2010).

480. Shusen Zhou, Qingcai Chen , Xiaolong Wang, "*Discriminative Deep Belief Networks for image classification",* in 17th IEEE International Conference on Image Processing, pp. 1561-1564 (2010).

481. Yoshua Bengio , "*Learning deep architectures for AI*" , Foundations and Trends in Machine Learning, Vol. 2(1),pp. 1-71(2009).

482. Y. Tang, "*Deep learning using linear support vector machines,*" in Workshop on Challenges in Representation Learning, ICML (2013).

483. Weston, J.R., Frédéric Mobahi, Hossein Collobert, Ronan, "*Deep Learning via Semi-supervised Embedding*", in Neural Networks: Tricks of the Trade, pp. 639-655 (2012).

484. Azizi Abdullah, Remco C. Veltkamp, Marco A., "*Wiering An Ensemble of Deep Support Vector Machines for Image Categorization*", in International Conference of Soft Computing and Pattern Recognition, pp. 301-306 (2009).

485. Hyun-Chul Kim, Shaoning Pang, Hong-Mo Je, Daijin Kim, Sung-Yang Bang, "*Support Vector Machine Ensemble with Bagging*", in Pattern Recognition with Support Vector Machines, pp. 397-408 (2002).

486. Jianzhao Qin, Yuanqing Li, and Wei Sun, "*A semisupervised support vector machines algorithm for BCI systems*" Computational Intelligence and Neuroscience, pp. 1-9 (2007).

487. Ruslan Salakhutdinov, "*On the Quantitative Analysis of Deep Belief Networks*", in Proc. of International Conference on Machine Learning, pp. 872-879 (2008).

488. G. E. Hinton, R. R. Salakhutdinov, "*Reducing the dimensionality of data with neural networks*", Science, Vol. 313(5786) , pp. 504-507 (2006).