

Cátia Sofia Pinho da Silva

Feature Extraction for Automatic Classification of Cutaneous Lesions from Dermoscopy Images

Dissertation

Supervisor:
Prof. Dr André Marçal

Co-supervisor:
Prof. Dr Teresa Mendonça

Master degree in Biomedical Engineering

Universidade do Porto, Faculdade de Engenharia

October 9, 2012

Dedicated to

my parents

Acknowledgments

I would like to thank the persons and entities that in one way or another contributed to the development of this thesis and whose contribution was invaluable. So, I thank:

To Professor Doctor André Marçal, for the availability and orientation.

To Professor Doctor Teresa Mendonça, in particular to the ADDI group, by the availability of data used in this thesis.

To Professor Doctor Aurélio Campilho, of whom I had the privilege of being a student in two courses of image analysis and computer aided diagnosis, thus learning the basic concepts for the realization of this thesis.

To my colleagues at the Masters and INESC Porto, and to the faculty, which contributed to my intellectual evolution and openness to new scientific areas.

To Professor Doctor Vladimiro Miranda, my supervisor at INESC Porto, by the confidence, availability, willingness, motivation and advice.

To Miguel, for companionship, patience and help in the revision of the thesis.

To Catarina, for the friendship.

Abstract

Feature Extraction for Automatic Classification of Cutaneous Lesions from Dermoscopy Images

by

Cátia Sofia Pinho da Silva

Dermoscopy (dermatoscopy or epiluminescence microscopy) is a non-invasive diagnostic technique for the *in vivo* observation of pigmented skin lesions used in dermatology. There is currently a great interest in the prospects of automatic image analysis methods for dermoscopy, both to provide quantitative information about a lesion, which can be of relevance for the clinician, and as a stand alone early warning tool. The standard approach in automatic dermoscopic image analysis has usually three stages: (i) segmentation, (ii) feature extraction and selection, (iii) lesion classification. This thesis evaluates the potential of an alternative approach based on the Menzies method - presence of 1 or more of 6 color classes, indicating that the lesion should be considered a potential melanoma. This method does not require stages (i) and (ii) - lesion segmentation and feature extraction.

The Jeffries-Matusita and Transformed Divergence separability distances were used to evaluate the color class separability for an experimental evaluation with 28 dermoscopic images. The identification of color classes in dermoscopic images is a subjective task, which poses great challenges for an automatic implementation. The purpose of this thesis is to evaluate the potential discrimination between the various Menzies color classes in dermoscopic RGB images.

The distribution of spectral values from each of the 6 color classes is quite different when the same color is taken in different images. This original behavior of the color classes provides only a little more than half of the pairs separable in the RGB feature space. Considering the skin as an additional class, an image intensity

calibration was applied to the dataset which increased slightly the rate of separable color class pairs. A non-linear cluster transformation allowed almost the total separation of each color class in the feature space.

Several neural networks in competition were used as a classifier, which lead to loss of arbitrariness and perfect knowledge of each cluster surface. The identification of color classes was validated by, at least, one dermatologist. The discrimination between the various Menzies color classes in dermoscopic RGB images performed 93% of sensibility, 62% of specificity and 74% of accuracy averaged measures. This result indicates that it is possible to discriminate the different colors inside a dermoscopic lesion, mimicking the human diagnosis.

Keywords: Dermoscopy, Menzies method, color, separability measures, image calibration, neural networks classifier.

Resumo

Extração de características e Classificação Automática de Lesões Pigmentadas da Pele

por

Cátia Sofia Pinho da Silva

A dermoscopia (dermatoscopia ou microscopia epiluminescence) é uma técnica de diagóstico para a observação *in vivo* de lesões pigmentadas da pele utilizada por dermatologistas. Atualmente, existe um grande interesse nas perspetivas que os métodos automáticos de análise de imagem proporcionam para a dermoscopia, fornecendo informação quantitativa da lesão ao médico dermatologista e servindo também como uma ferramenta relevante e independente de diagnóstico precoce. A abordagem típica da análise de imagem automática envolve, usualmente, três estados: (i) segmentação, (ii) extração e seleção de características, (iii) classificação da lesão. Esta tese avalia a capacidade de uma abordagem baseada no método de Menzies - presença de 1 ou mais classes de cor num total de 6, indicando assim se a lesão deve ser considerada como um potencial melanoma. Esta metodologia não requer os passos (i) e (ii) - segmentação da lesão e extração de características.

As medidas de distância de Jeffries-Matusita e Divergência Transformada foram utilizadas de forma a avaliar a separabilidade das classes de cor em 28 images dermoscópicas. A identificação de classes de cor nestas imagens, é uma tarefa subjetiva e coloca desafios de grau elevado para uma implementação automática. O objectivo desta tese é, assim, avaliar a potencialidade discriminatória das várias classes de cor de Menzies em imagens dermoscópicas no espaço RGB.

A gama espetral de valores da distribuição de cada uma das 6 classes de cor difere consideralvelmente quando a cor é proveniente de imagens diferentes. Este comportamento original das classes de cor fornece, apenas, um pouco mais de metade de pares de classes separáveis no espaço de características RGB. Considerando a

pele como uma classe adicional, uma calibração de intensidades da imagem aplicada ao conjunto de dados resultou num ligeiro aumento do número de classes de cor separáveis. Uma transformação não-linear dos *clusters*, garantiu a separabilidade de quase a totalidade dos pares de classes no espaço de características.

A aplicaçã o de métodos baseados em redes neuronais na etapa de classificação, diminui a taxa de arbitrariedade (dispensa limiares de aceitação) e permitiu o conhecimento da superfície de cada cluster. A identificação de classes de cor foi validada por, pelo menos, um dermatologista. A discriminação entre as várias classes de cor de Menzies em imagens dermoscópicas no espaço de características RGB conduziu a um desempenho médio de 93% de sensibilidade, 62% de especificidade e 74% de precisão do classificador. Este resultado indica que é possível a discriminação de diferentes cores numa lesão dermoscópica, reproduzindo um diagnóstico humano fidedigno.

Palavras-chave: Dermoscopia, método de Menzies, cor, medidas de separabilidade, calibração, redes neuronais.

Contents

Acknowledgments iii

Abstract v

Resumo vii

Content xii

List of figures xvii

List of tables xx

Abbreviations xxi

1. Introduction 1
- 1.1 General Context 1
- 1.2 History 2
- 1.3 The Cutaneous Melanoma 5
  - 1.3.1 Medical Basis 5
  - 1.3.2 Relevance of Early Diagnosis 7
- 1.4 Dermoscopy 8
- 1.5 Motivation and Objectives 9
- 1.6 Main Contributions of the Thesis 9
- 1.7 Outline 9

2. Computer Aided-Diagnosis: an Overview 11
- 2.1 Pattern Recognition Principles and Fundamentals 11
- 2.2 Steps of Digital Image Processing 12

2.3 Medical Image Formation 14
2.3.1 Imaging Modalities 14
2.3.2 Digital Image Formation 15
2.4 Image Enhancement 15
2.5 Digital Image Analysis - CADx system 17
2.5.1 Segmentation 17
2.5.2 Feature Choice 18
2.5.3 Classification 19
2.5.4 Validation 21

3. State-of-the-Art 25
3.1 Diagnostic Algorithms - Human Dermoscopy 25
3.1.1 Pattern Analysis 26
3.1.2 ABCD rule 26
3.1.3 Menzies Method 27
3.1.4 7-Point Checklist 27
3.1.5 Other diagnostic algorithms 28
3.2 Diagnostic Accuracies 28
3.3 Dermoscopy Image Analysis - Computer Dermoscopy 29
3.3.1 Image Analysis Approaches 29
3.3.2 Segmentation of Skin Lesions 30
3.3.3 Classification Methods 32
3.4 Evaluation Issues 35
3.4.1 Clinical Case Databases 35
3.4.2 Evaluation Methods 36
3.5 Commercial Systems 38
3.5.1 System Design Principles 38
3.5.2 Image Capture Devices 38
3.5.3 Dermocopy Computer Systems 39
3.6 Concluding Remarks 39

4. Proposed Method 41
4.1 Description 41
4.2 Dataset 44
4.3 Color Regions 48
4.4 Calibration 48

4.4.1 Skin Intensity Calibration 48
4.4.2 Calibration based on Skin-Lesion Rate 49
4.4.3 Calibration with the Magnitude of the Reference Mean Vector Constant 50
4.5 Color Analysis 51
4.6 Separability Measures 55
4.6.1 Divergence 55
4.6.2 Bhattacharyya Distance 58
4.6.3 Jeffries-Matusita 58
4.6.4 Transformed Divergence 60
4.6.5 Kolmogorov-Smirnov Normality Test 60
4.7 Inter- and Intra-Class Tests and Grouping Data 61
4.8 Non-Linear Technique for Increasing Clusters Separability 62
4.9 Classification 64
4.10 Post-Processing 68
4.11 Validation 69

5. Results and Results Discussion 71
5.1 Experimental Data Description 71
5.1.1 Intra-Class Tests 72
5.2 Pre-Processing and Feature Selection 77
5.2.1 Single Image Training 77
5.2.2 Multiple Image Training 81
5.2.3 With Calibration 83
5.2.4 Including the Skin 86
5.2.5 Discussion 87
5.3 Train and Test Sets Description 88
5.4 Increasing Clusters Separability 98
5.5 Classification and Post-Processing - Step 1 100
5.6 Classification and Post-Processing - Step 2 104
5.7 Validation 108

6. Concluding Remarks 111

Bibliography 114

A. Dataset of Dermoscopy Images 123

B. Intra-Classes Tests 125

C. Color Analysis 137

D. Resulted Images of Image Intensity Calibration 139

E. Resulted Images of Calibration based on Skin-Lesion Rate 141

F. Resulted Images of Calibration with the Magnitude of the Reference Mean Vector Constant 143

G. Classified Images 145

List of Figures

1.1 *George Ebers* German Egyptologist and novelist, discovered the Egyptian medical *Papyrus Ebers*. 3
1.2 Scottish surgeon John Hunter (1728-1793) and Rudolf Virchow (1821-1902) "the father of modern pathology". 4
1.3 Louis Leakey (1903-1972). 5
1.4 Human skin anatomy and histology. 6
1.5 Example of a malignant melanoma. 6
1.6 Skin anatomy - the melanocytics. 7

2.1 Overview of the most common modules of an image processing system. 13
2.2 Dermatoscope. 15
2.3 Geometric distortion and brightness variation. 16
2.4 Medical image classification task. 17
2.5 Patterns - image regions. 18
2.6 Classifier illustration. 19
2.7 Clustering illustration. 21
2.8 Validation - the performance metrics for classification problems. . . . 22
2.9 Block diagram for the design of a classification system. 23

3.1 Flow chart for the differential diagnosis of pigmented skin lesions is performed with a 2-step approach. 26
3.2 Thresholding as segmentation technique in feature domain. 31
3.3 Principle of a 3-layer ANN. 33
3.4 Bayesian classifier. 34
3.5 kNN classification method. 35
3.6 Support Vector Machine example. 35

4.1 Classification algorithm of dermoscopic PSL: flowchart of the proposed method. 43
4.2 Identification of color regions in PSL images. 47
4.3 Procedure of the skin-lesion rate calibration. 50
4.4 Representative scheme of the technique of calibration with magnitude of reference constant. 51
4.5 Schematic showing how pixels of a RGB color image are formed from the corresponding pixels of the three component images. 52
4.6 Schematic of the RGB color cube showing and secondary colors of light at the vertices. 52
4.7 Hue and saturation in the HSI color model. 53
4.8 Two dimensional multispectral space showing a hypothetical degree of separation. 56
4.9 Divergence. 57
4.10 A problem with divergence. 58
4.11 Jeffries-Matusita measure. 59
4.12 Non-linear cluster transformation. 64
4.13 Proposed approach for the classification task. 65
4.14 Biological neuron scheme. 66
4.15 Operational techniques in a generic neuron. 66
4.16 Illustration of filter applies in a pseudo-region of classified image. . . 68

5.1 Boxplot for subset of images that have the BG (Blue Gray) color class in the Red channel of RGB color space. 74
5.2 Boxplot for subset of images that have the BG (Blue Gray) color class in the Green channel of RGB color space. 74
5.3 Boxplot for subset of images that have the BG (Blue Gray) color class in the Blue channel of RGB color space. 75
5.4 Boxplot of all (28) test images for skin class - evaluation of the Red color channel. 75
5.5 Boxplot of all (28) test images for skin class - evaluation of the Green color channel. 76
5.6 Boxplot of all (28) test images for skin class - evaluation of the Blue color channel. 76
5.7 Scatterplots for the single image training test without calibration. . . 78
5.8 Scatterplots for the multiple image training test without calibration. . 82

5.9 Scatterplots for the MIT test with skin intensity calibration (method I). 84
5.10 Scatterplots for the training set without calibration. 93
5.11 Scatterplots for the training set with calibration method I. 94
5.12 Scatterplots for the test set without calibration. 95
5.13 Scatterplots for the test set with calibration method I. 96
5.14 Scatterplots for the training data with skin intensity calibration (method I) and a non-linear technique which increase the clusters separability distances. 99
5.15 Hyperbolic tangent sigmoid transfer function. 101
5.16 Structure of the neural network classifier. 102
5.17 Results of the neural networks in competition classification considering class groups for image IMD06. 103
5.18 Results of morphological operators and filtering in the step 1 classified images. 104
5.19 Scatterplots for the DB, BG and Bk color classes after 12 Polikar transformation. 105
5.20 Results of the second classification and post-processing for image IMD06. 107

A.1 Dermoscopy Images Dataset - Melanocytic melanomas. 123
A.2 Dermoscopy Images Dataset - Melanocytic nevus. 124

B.1 Boxplot for subset of images that have the Wh (White) color class in the Red color channel of RGB feature space. 126
B.2 Boxplot for subset of images that have the Wh (White) color class in the Green color channel of RGB feature space. 126
B.3 Boxplot for subset of images that have the Wh (White) color class in the Blue color channel of RGB feature space. 127
B.4 Boxplot for subset of images that have the LB (Light Brown) color class in the Red color channel of the RGB feature space. 127
B.5 Boxplot for subset of images that have the LB (Light Brown) color class in the Green color channel of the RGB feature space. 128
B.6 Boxplot for subset of images that have the LB (Light Brown) color class in the Blue color channel of the RGB feature space. 129
B.7 Boxplot for subset of images that have the DB (Dark Brown) color class in the Red color channel of the RGB feature space. 130

B.8 Boxplot for subset of images that have the DB (Dark Brown) color class in the Green color channel of the RGB feature space. 131
B.9 Boxplot for subset of images that have the DB (Dark Brown) color class in the Blue color channel of the RGB feature space. 132
B.10 Boxplot for subset of images that have the BG (Blue Gray) color class in the Red color channel of the RGB feature space. 132
B.11 Boxplot for subset of images that have the BG (Blue Gray) color class in the Green color channel of the RGB feature space. 133
B.12 Boxplot for subset of images that have the BG (Blue Gray) color class in the Blue color channel of the RGB feature space. 133
B.13 Boxplot for subset of images that have the Bk (Black) color class in the Red color channel of the RGB feature space. 134
B.14 Boxplot for subset of images that have the Bk (Black) color class in the Green color channel of the RGB feature space. 134
B.15 Boxplot for subset of images that have the Bk (Black) color class in the Blue color channel of the RGB feature space. 135

D.1 Dermoscopy Images Dataset resulting from Calibration method I - Melanocytic melanomas. 139
D.2 Dermoscopy Images Dataset resulting from Calibration method I - Melanocytic nevus. 140

E.1 Dermoscopy Images Dataset resulting from Calibration method II - Melanocytic melanomas. 141
E.2 Dermoscopy Images Dataset resulting from Calibration method II - Melanocytic nevus. 142

F.1 Dermoscopy Images Dataset resulting from Calibration method III - Melanocytic melanomas. 143
F.2 Dermoscopy Images Dataset resulting from Calibration method III - Melanocytic nevus. 144

G.1 Results for neural networks in competition classification considering 5 class groups. 146
G.2 Dilation - step 1. 147
G.3 Mode filter - step 1. 148

G.4 Final results obtained with the neural networks in competition in order to divide class group 4 into the three possible color classes (DB, BG and Bk). 149
G.5 Filtering - Step 2. 150
G.6 Black corners removal. 151

List of Tables

3.1 Evaluation of diagnostic algorithms. 26
3.2 Classifier comparison on 466 PSLs (125 CM, 341 benign PSL). 33
3.3 State-of-the-art methodologies summary. 37
3.4 Commercial dermoscopic systems with image analysis component. . . 39

4.1 Presence of color in the dataset. 45
4.2 Colors indication and color regions identification. 46
4.3 Number of image color samples for each color class in the 28 dataset images. 48

5.1 Original data behavior for single image training and multiple image training groups without calibration. 79
5.2 Jeffries-Matusita and Transformed Divergence distance values, for the single image training. 80
5.3 Jeffries-Matusita and Transformed Divergence distance values, for the multiple image training. 81
5.4 Data behavior for multiple image training group with image calibration method I. 85
5.5 Jeffries-Matusita and Transformed Divergence distance values, for the multiple image training with skin intensity calibration (method I). . . 86
5.6 Summary of JM results for the various scenarios tested (number of class pairs). 87
5.7 Summary of JM results for the various scenarios tested (number of class pairs) including the skin class. 87
5.8 Training and Testing sets. 89
5.9 Training data without and with calibration. 90

5.10 Test data without and with calibration. 91
5.11 Jeffries-Matusita and Transformed Divergence distance values, for the training and test sets, with and without calibration method I. 97
5.12 Jeffries-Matusita and Transformed Divergence distance values, for the dataset resulting of five Polikar's transformations 100
5.13 Jeffries-Matusita and Transformed Divergence distance values, for the grouping dataset resulting of the 5th Polikar transformation. 100
5.14 Parameters of the neural network classifier. 101
5.15 Classification results - step 1: classification in 5 rearranged groups. . 102
5.16 Polikar transformations for increase separability distances in the three color classes that compose group label 4. 106
5.17 Validation results. 108

C.1 Jeffries-Matusita and Transformed Divergence distance values, for the multiple image training using the HSI color space. 137
C.2 Jeffries-Matusita and Transformed Divergence distance values, for the multiple image training using the Opponent color space. 138
C.3 Jeffries-Matusita and Transformed Divergence distance values, for the multiple image training using the XYI color space. 138

Abbreviations

ABCD	Asymmetry, Border, Color, Differential structures
ABCDE	Asymmetry, Border, Color, Differential structures, Evolution
ADDI	Automatic computer-based Diagnosis system for Dermoscopy Images
ALM	Acral Lentiginous Melanoma
AMN	Atypical Melanocytic Nevi
ANN	Artificial Neural Network
AUC	Area Under the Curve
BG	Blue Gray
Bk	Black
CAD	Computer-Aided Diagnosis
CADe	Computer-Aided Detection
CADx	Computer-Aided Diagnosis
CART	Classification And Regression Tree
CASH	Color, Architecture, Symmetry, Homogeneity
CCD	Charge-Coupled Device
CI	Confidence Interval
CM	Cutaneous Melanoma
CNMD	Consensus Net Meeting on Dermoscopy
DANAOS	Diagnostic and neural analysis of skin cancer
DB	Dark Brown
DPV	Dermatoscopic Point Value
DICOM	Digital Imaging and Communications in Medicine
ELM	Epiluminescence Microscopy
FCUP	Faculdade de Ciências da Universidade do Porto
FN	False Negative
FP	False Positive
FPF	False Positive Fraction
HPH	Hospital Pedro Hispano
HSI	Hue, Saturation, and Intensity color space
HSV	Hue, Saturation, and Value color space
ITL	Information Theoretic Learning
JM	Jeffries-Matusita

kNN	k-Nearest Neighbor
KS	Kolmogorov-Smirnov
LB	Light Brown
LDA	Linear Discriminant Analysis
LMM	Lentigo Maligna Melanoma
Logreg	Logistic regression
MSE	Mean Square Error
NCT	Nonlinear Cluster Transformation
NM	Nodular Melanoma
NN	Neural Network
NNT	Number Needed to Treat
OPP	Opponent color space
P	Precision
PACS	Picture Archiving and Communication Systems
PCA	Pricipal Component Analysis
PSL	Pigmented Skin Lesion
Re	Red
RGB	Red, Green, and Blue color channels
ROC	Receiver Operating Characteristic
SE	Sensivity
SIM	Scaling Index Method
SP	Specificity
SSM	Superficial Spreading Melanoma
SVM	Support Vector Machines
tansig	hyperbolic tangent sigmoid
TM	Thin Melanoma
TN	True Negative
TP	True Positive
TPF	True Positive Fraction
Wh	White
WHO	World Health Organization

Chapter 1

Introduction

Skin cancer is the most incident cancer in the 20th century with 8 new cases per 100,000 population per year in Portugal [46]. The worst case scenario can lead to death if the cancer is not diagnosed early.

This chapter introduces the skin cancer, more particularly, the cutaneous melanoma, indicating some history of the current biological knowledge about it. Dermoscopy, a non-invasive technique for observation of pigmented skin lesions and its importance for an early diagnosis is also presented. The objectives, motivation and main contributions of this work also appears in the last sections of the chapter, as well as the outline of the thesis.

1.1 General Context

Human cutaneous melanoma, a deadly skin cancer that develops through the malignant transformation of melanocytes, can be a highly aggressive neoplasm, characterized by a rapidly growing incidence and an elevated mortality rate [38]. Although early melanomas are locally controlled with surgical excision, up to 20% of patients will develop metastatic tumours due to its high capability to invade and rapidly metastasize to distinct organs [38]. Melanoma and non-melanoma skin cancers currently constitute one of the most common malignancies in the portuguese population, and the worldwide incidence and mortality rates are continuously increasing. In particular, melanoma incidence has increased more than any other cancer, and currently has reached 8 new cases per 100,000 population per year in Portugal [46]. Because advanced skin cancer remain incurable, early detection and surgical excision is currently the only approach to reduce mortality. An action must be taken in

order to develop new methods for early diagnosis leading to the decrease of mortality rate. This work presents a new approach that allows the diagnosis of a pigmented skin lesion as benign or potential malignant, aiding the early diagnosis.

1.2 History

Dermatology had a very small beginning, and its origin is found in the early times of Egypt, Persia, India, and among the Hebrews. But the human cancer is probably as old as the human race. It is obvious that cancer did not suddenly start appearing after modernization or industrial revolution. The oldest known description of human cancer is found in seven Egyptian papyri between 3000-1500 BC. Two of them, known as the *Edwin Smith* and *George Ebers* papyri, contain details of conditions that are consistent with modern descriptions of cancer (Figure 1.1) [63].

Hippocrates, the great Greek physician (460-370 BC), who is considered the father of medicine is though to be the first person to clearly recognize difference between benign and malignant tumours. His writings include description of cancers involving various body sites [63].

Later in the course of history, Constantinople became the intellectual headquarters of medicine. The ancient teachings of Hippocrates and Galen continued to influence the physicians in Constantinople, Cairo, Alexandria, and Athens. During this period the cause of cancer was explained as the result of an excess of black bile [63].

During the Renaissance, beginning in the 15th century, physicians acquired greater knowledge of the anatomy and physiology of the human body. This started an era, that has seen the advancement of surgery and development of rational therapies based on clinical observations. The theory that cancer is caused by excess of black bile continued to prevail through the 16th century. At this time cancer was considered incurable, however a variety of temporary measures were available including creams and pastes containing arsenic. Autopsies, performed by Harvey in the 17th century, gave an insight into the circulation system. By about the same period Gaspare Aselli discovered the lymphatic system and this led to the end of the old theory of black bile as the cause of cancer. The new theory suggested that abnormalities in the lymph and lymphatic system as the primary cause of cancer. A French physician, Claude Gendron suggested that cancer arises locally as a hard,

[2] `http://en.wikipedia.org/wiki/Georg_Ebers`



(a) (b)

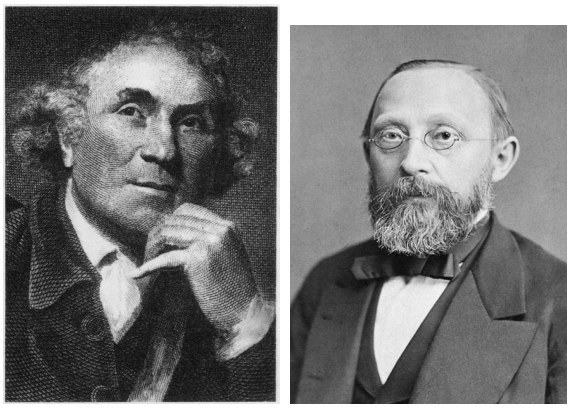
Fig. 1.1: (a) *George Ebers*: German Egyptologist and novelist, discovered the Egyptian medical (b) *Papyrus Ebers*[2].

growing mass, untreatable with drugs, and must be removed with all its "filaments" [63].

Jean Astruc and chemist Bernard, two 18th century physicians, conducted research to confirm or disprove their current theories related to the origin of cancer. These efforts were the very first steps of experimental oncology. The art and science of seeking better diagnosis, treatments and understanding of the causes of cancer evolved from many who followed their path. In 1761, Giovanni Morgagni of Padua was the first to do an autopsy to look for the pathological findings in a patient after death. This and the efforts of many great physicians who followed them laid the foundation for scientific oncology. John Hunter (1728-1793) (Figure 1.2(a)), a famous Scottish surgeon suggested that some cancers might be cured by surgery and described methods by which we can distinguish the surgically removable tumours. He suggested that, if a tumour has not encroached to the nearby tissue and was still moveable, "there is no impropriety in removing it" [63].

The discovery of anesthesia in 1844 by Wells allowed surgery to flourish and the classic cancer operations such as radical mastectomy were developed. The discovery of microscope by Leeuwenhoek in the late 17th century added momentum to the quest for the cause of cancer. By late 19th century, with the development of better microscopes to study cancer tissues scientists gained more knowledge about the cancer process. These studies showed that cancer cells are markedly different in appearance compared to the surrounding normal cells from which they originated.

[4] `http://en.wikipedia.org/wiki/Rudolf_Virchow`



(a) (b)

Fig. 1.2: (a) John Hunter (1728-1793): was a Scottish surgeon regarded as one of the most distinguished scientists and surgeons of his day [44]; (b) Rudolf Virchow[4](1821-1902): referred to as "the father of modern pathology".

Rudolf Virchow (Figure 1.2(b)), who is often called the founder of cellular pathology, provided the scientific basis for the modern pathologic study of cancer and correlated the clinical course of illness with microscopic findings. This approach led to the development of modern cancer surgery. Tissues removed by the surgeons were examined under the microscope to make a precise diagnosis of cancer. By looking at the cut margin of surgery the pathologist were able to tell the surgeon if the cooperative procedure had completely removed the tumour [52].

The early 20th century saw great progress in our understanding of microscopic structure and functioning of the living cells. A virus causing cancer in chickens was identified in 1911. Existence of many chemical and physical carcinogens were conclusively identified during later part of the 20th century. Later part of the 20th century showed tremendous improvement in our understanding of the cellular mechanisms related to cell growth and division. Many factors that suppress and activate the cell growth and division were identified [52].

The oldest available specimen of a human cancer was found in the remains of skull of a female who lived during the Bronze Age (1900-1600 BC). The mummified skeletal remains of Peruvian Incas, dating back 2400 years ago, contained abnormalities suggestive of involvement with malignant melanoma. Cancer was also found in fossilized bones recovered from ancient Egypt. Louis Leakey (Figure 1.3) found the oldest possible hominid malignant tumour in 1932 from the remains of a body,

[5] `http://www.blm.gov/ca/st/en/fo/barstow/calico.html`

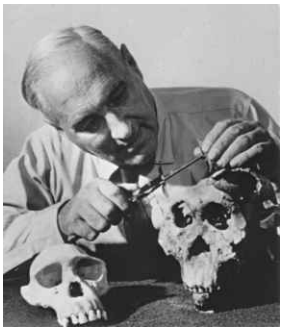


Fig. 1.3: Louis Leakey[5](1903-1972).

which could be either that of Homo erectus or an Australopithecus. This tumour had features suggestive of a Burkitts lymphoma [63].

1.3 The Cutaneous Melanoma

Currently, skin cancer is one of the most incident diseases in Portugal and all over the world, and it is associated to several implications at the physical, psychological and social level. Usually it leads to a diminished quality of life, being one of the main health problems of the XXI century [46].

Among the majority of cancers, skin cancer is a type of cancer with greater growth in recent years. This type of cancer annually represents about one third of all cancers detected in Portugal, affecting one in seven individuals throughout life. It is estimated that the malignant melanoma (skin cancer), accounts for only 8% of skin cancers detected in Portugal, however being responsible for more than 70% of deaths, a figure that has tripled since 1980. According to the Institute of Oncology, the incidence rate in our country is estimated to be 8 cases per 100,000 inhabitants/year, which gives about 800 new cases per year [46].

Skin cancer develops in the upper layer of the skin, the epidermis (Figure 1.4). The most common types are the basal cell carcinoma, the squamous cell carcinoma, and the Cutaneous Melanoma (CM). Of these three types, CM is the most aggressive one.

1.3.1 Medical Basis

The CM is a malignant type of Pigmented Skin Lesion (PSL) or melanocytic skin lesion. A PSL typically is a *dark spot*, *mole* or *nervus* on the skin (Figure 1.5) originating from an aggregation of the skin color pigment melanin. Special cells, the

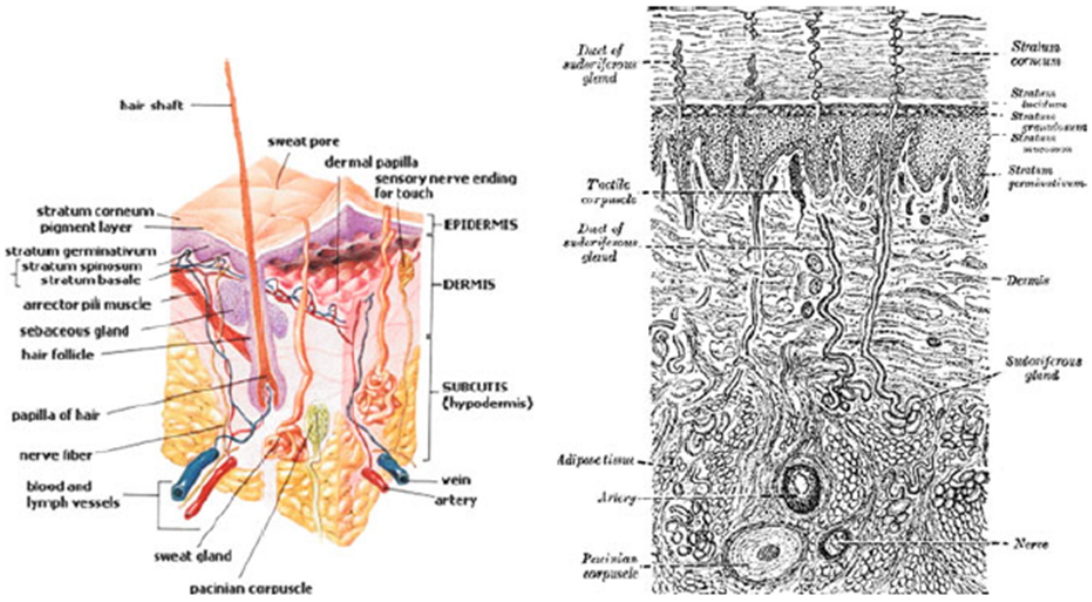


Fig. 1.4: The *human skin*. Anatomy (left) and histology (right) [23].

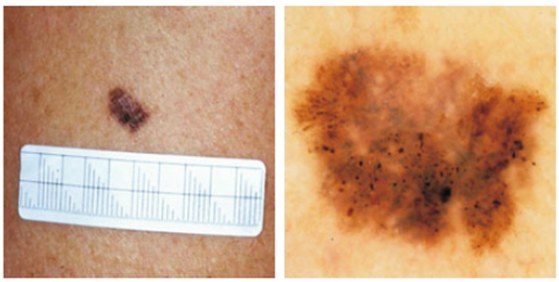


Fig. 1.5: Example of a *malignant melanoma*. Macroscopic ("clinical") view with the naked eye (left), and microscopic ("dermoscopic") view (right) [23].

melanocytes (Figure 1.3.1), produce melanin in small granules called melanosomes. The melanin is then transported to cells in the outer skin keratocytes, where they appear as "color" of the skin [23].

The CM, also referred to as malignant melanoma or simply melanoma, is a cancerous tumour that arises in the melanocytes. It accounts for about 3% of all skin cancers [3]. The most common CM types are [75]:

- The Superficial Spreading Melanoma (SSM), usually spreads along the epidermis and then grows deeper to the dermis (Figure 1.4). It is curable when it is removed before it invades the dermis.

- The Acral Lentiginous Melanoma (ALM) is the most common melanoma in dark-skinned races [21]. It is frequently mistaken for a bruise or nail streak in

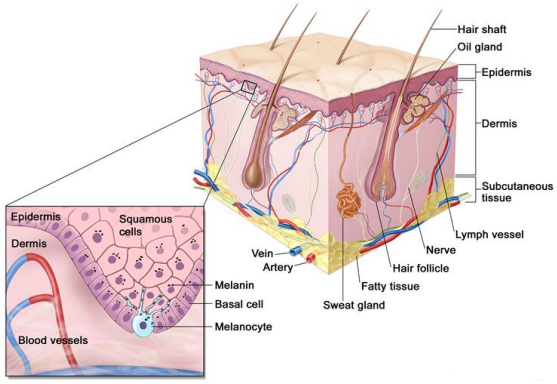


Fig. 1.6: Skin anatomy showing the epidermis, dermis and subcutaneous tissue. The melanocytes are in the basal layer cells in the deepest part of the epidermis [23].

its early stages and therefore often diagnosed in a later stage when it is very aggressive.

- The Lentigo Maligna Melanoma (LMM) develops in a sun-induced freckle and typically occurs on sun-damaged skin in the elderly and may in its early stages be mistaken for a benign age spot or sun spot.

- The Nodular Melanoma (NM) is a very aggressive CM type that lends to grow downwards into deeper skin tissue rather than along the surface of the skin.

Cancer statistics prove increasing incidence of CM in many countries with white population over the last four decades [30]. With an estimated increase of 3-7%, it is the most rapidly increasing cancer in white populations, suggesting a doubling of rates every ten to twenty years. Incidence rates in central Europe are about 5% with a North-South gradient due to the different skin types between North and South populations: in the North skin is more pale and in the South it is more "tanned". Risk factors include: the number of PSLs on the body; the presence of atypical PSLs; sun-exposure and sunburns; and hereditary factors ([29],[3]).

1.3.2 Relevance of Early Diagnosis

Unlike other types of cancer, there is a good chance for 100% cure in cases of CM. Detected in an early, non-invasive stage, a CM can be surgically removed (excised), with an excellent prognosis for the patient. CM screening programs are therefore promoted [48]. Nowadays, in about 90% of cases a CM is diagnosed at the primary

tumour stage. However, the rate of diagnostic excisions, i.e. the number of PSLs excised per CM, also referred to as Number Needed to Treat (NNT), is very high, with estimated averages of 10-90%, depending on the patients age group and the experience of the examiner ([26],[33]). CM diagnoses is very demanding and requires a lot of experience. Given the dangerous nature of CM, many suspicious of PSLs are excised. However, as they are not malignant, it would not be necessary to remove them - sensitivity is put above specificity. PSL excisions are performed with certain safety margins. Current recommendations are 0.5, 1, and 2 cm for *in situ* CM, CM with Breslow thickness $\leq$ 2 mm and CM with thickness $>$ 2 mm, respectively [29]. As surgical interventions, excisions implicate inherent risks and costs. Therefore, the average NNT should be reduced by suitable measures such as enhanced training and improvements on diagnostic imaging techniques and algorithms. Furhermore, according Kittler et al. [42], the accuracy of clinical diagnosis of CM to naked eye is only 60%. The performance of diagnostic PSLs increases when the diagnosis is made by a group of experienced specialists in consensus and decreases as the prevalence of melanomas increases.

1.4 Dermoscopy

Dermoscopy (dermatoscopy, epiluminescence microscopy, incident light microscopy, skin surface microscopy) is a non-invasive diagnostic technique for the *in vivo* observation of PSLs, allowing a better visualization of surface and subsurface structures (from the epidermis to the papillary dermis - Figure 1.6). This diagnostic tool permits the visualization of morphological structures of the epidermis and the dermo-epidermal junction until the papillary dermis, thus opening a new dimension in the analysis of the clinical morphologic features of PSLs [9]. Several studies [77] have shown that this method may improve diagnostic sensitivity by 20-30% compared with clinical diagnosis by the naked eye. For the success of clinical applicability of dermoscopy, experienced physicians are needed. In order to reduce the learning-curve of non-expert clinicians and to mitigate problems inherent in the reliability and reproducibility of the diagnostic criteria used in pattern analysis, several indicative methods based on diagnostic algorithms have been introduced in the last few years. The ABCD rule, the 7-point checklist, and the Menzie's method are the most relevant ones ([7],[70],[77]) - more details in Chapter 3.

As such, the diagnostic of early malignant melanoma contributes essentially to a decrease in the rate of mortality from this disease.

1.5 Motivation and Objectives

As digital image analysis is concerned (see Chapter 2), this work focuses on the diagnosis (i.e., malignancy assessment) of PSLs, not on their detection. Both tasks, detection and diagnosis, can be supported by digital image analysis. Related systems are commonly referred to as Computer-Aided Detection (CADe), and Computer-Aided Diagnosis (CADx) [23], respectively.

Because an early diagnosis of PSLs is critical for patient survival, the development of CADx system is the main goal of this thesis. This system is focused on the development of an algorithm for a color-based automatic classification of a dermoscopic image - more details in Chapter 4.

1.6 Main Contributions of the Thesis

The main contributions of this thesis to the state-of-the-art are:

- Evaluation of the potential application of the Menzies method on an image processing based diagnosis
- Development of a potential new diagnostic algorithm of PSLs.

The work developed for this thesis resulted in two conference papers:

1. ICIAR: Silva, Cátia S. P.; Marçal, André R. S.; Pereira, Marta A.; Mendonça, Teresa; Rozeira, Jorge: Separability Analysis of color classes on dermoscopic images. Lecture Notes on Computer Science, Part II, pp. 268-77, ICIAR 2012.
2. CompImage: Marçal, André R. S.; Silva, Cátia S. P.; Mendonça, Teresa; Pereira, Marta A.; Rozeira, Jorge: Evaluation of the Menzies method potential for automatic dermoscopic image analysis. Computational Modeling of Objects Presentd in Images (CompImage): Fundamentals, Methods and Applications, 3rd edition, Taylor & Francis. 2012 (in press).

1.7 Outline

The thesis is divided in six chapters. The first chapter addresses the history and appearance of the first cancers, the concept of cutaneous melanoma, the medical basis and the importance of early diagnosis. In the end, the main objectives and

their motivations are presented, as well as a summary of the main contributions of the thesis.

In the second chapter, digital image processing will be described, addressing the work that will be presented and introducing the relevant concepts and notation.

The third chapter contains the state-of-the-art of CM diagnosis, with particular emphasis on the methodology followed by all existing methods.

The fourth chapter introduces the theory of the proposed method for CM diagnosis, following its stages of development, all of them relevant to the results obtained.

The fifth chapter presents the main results of the proposed method for pigmented skin lesions diagnosis.

The sixth and last chapter presents conclusions about the work and suggestions for future work.

Chapter 2

Computer Aided-Diagnosis: an Overview

This chapter describes the computer-aided diagnosis application of pattern recognition in a brief and concise overview. This theme holds the theoretical basis in which this work is inserted: biomedical image processing, in particular, dermoscopy image processing. An introduction to the methods of biomedical image processing is also made. After some fundamental preliminary remarks to the terminology used, medical imaging modalities (epiluminescence microscopy) are introduced (Section 2.3). Sections 2.4 and 2.5 deal with low-level image processing and visualization, respectively, as far as necessary to understand the following chapters. Subsequently, the core steps of image analysis, namely feature choice, segmentation, classification and quantitative measurements (or validation) are presented in separate sections.

2.1 Pattern Recognition Principles and Fundamentals

Pattern recognition is the scientific discipline whose goal is the classification of objects into a number of categories or *classes*. Depending on the application, these objects can be images or signal waveforms or any type of measurements that need to be classified. These objects are called *patterns*. Pattern recognition has a long history, but before the 1960s it was mostly the output of theoretical research in the area of statistics. As with many other subjects, the advent of computers increased the demand for practical applications of pattern recognition, which in turn set new demands for further theoretical developments. As our society evolves from the industrial to its postindustrial phase, automation in industrial production and the need for information handling and retrieval are becoming increasingly important.

This trend has pushed pattern recognition to the high edge of today's engineering applications and research. Pattern recognition is an integral part of most *machine intelligence* systems built for decision making [80].

Computer-aided diagnosis is an important application of pattern recognition, aiming at assisting doctors in making diagnostic decisions. The final diagnosis is, of course, made by the doctor. Computer-aided diagnosis has been applied and is of interest for a variety of medical data, such as dermoscopy images. The need for a computer-aided diagnosis stems from the act that medical data are often not easily interpretable, and the interpretation can depend very much on the skill of the physician (Section 1.3.2). Although dermoscopy is the best method for detecting skin cancer, the false negative rate for the clinical diagnosis for malignant melanoma is reported to be of the order of 20 to 50% ([20], [55]). Current estimates indicate that between 15 and 40 benign lesions are excised for biopsy [20] in order to confirm one melanoma case. This may be caused by poor image quality, eye fatigue of the physician, the subtle nature of the findings or the skill of the examining physician. The percentage of correct classifications improves at a second reading by another physician. Thus, one can aim to develop a pattern recognition system in order to assist physicians with a "second" opinion. Increasing confidence in the diagnosis based on dermoscopy images would, in turn, decrease the number of patients with suspected, but not malignant, skin cancer who have to undergo surgical skin biopsy, with its associated complications.

By the increasing use of direct digital imaging systems for medical diagnostics, ***digital image processing*** becomes more and more important in health care. Digital images are composed of individual pixels (this acronym is formed from the words "picture" and "element"), to which discrete brightness or color values are assigned. They can be efficiently processed, objectively evaluated, and made available at many places at the same time by means of appropriate communication networks and protocols, such as Picture Archiving and Communication Systems (PACS) and the Digital Imaging and Communications in Medicine (DICOM) protocol, respectively. Based on digital imaging techniques, the entire spectrum of digital image processing is now applicable in medicine [23].

2.2 Steps of Digital Image Processing

The commonly used term "biomedical image processing" means the provision of digital image processing for biomedical sciences. In general, digital image processing

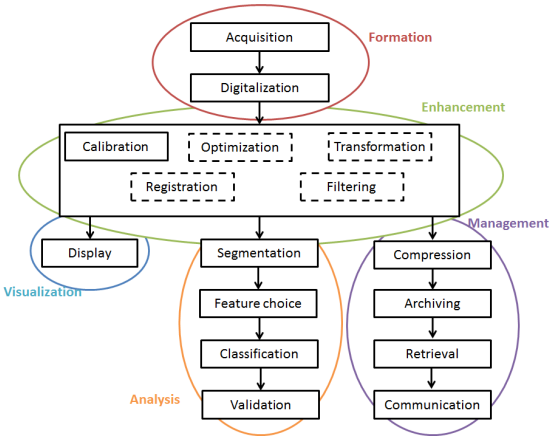


Fig. 2.1: Overview of the most common modules of an image processing system.

covers four major areas ([23], [24], [80]) - Figure 2.1:

1. *Image formation* includes all the steps from capturing the image to forming a digital image matrix.

2. *Image visualization* refers to all types of manipulation of this matrix, resulting in an optimized output of the image.

3. *Image analysis*, also called as Computer-Aided Diagnosis (CADx), includes all the steps of processing, which are used for quantitative measurements as well as abstract interpretations of biomedical images. These steps require a priori knowledge on the nature and content of the images, which must be integrated into the algorithms on a high level of abstraction. Thus, the process of image analysis is very specific, and the algorithms developed are usually not directly transferable into other application domains.

4. *Image management* sums up all techniques that provide the efficient storage, communication, transmission, archiving, and access (retrieval) of image data. Thus, the methods of telemedicine are also a part of the image management.

In contrast to image analysis, which is often also referred to as high-level image processing, low-level processing denotes manual or automatic techniques, which can be realized without a priori knowledge on the specific content of images. This type of algorithms has similar effects regardless of the content of the images. Therefore, low-level processing methods are usually available with programs for image enhancement [23].

2.3 Medical Image Formation

Since the discovery of X-rays by Wilhelm Conrad Röntgen in 1895, medical images have become a major component of diagnostics, treatment planning and procedures, and follow-up studies. Furthermore, medical images are used for education, documentation, and research describing morphology as well as physical and biological functions in 1D, 2D, 3D, and even 4D image data. Today, a large variety of imaging modalities have been established, which are based on transmission, reflection or refraction of light, radiation, temperature, sound, or electron spin.

2.3.1 Imaging Modalities

From the various of the medical imaging modalities, optical microscopy will be the focus for this work, in particular epiluminescence microscopy.

Epiluminescence Microscopy

Microscopy is the technical field of using microscopes to view samples and objects that cannot be seen with the unaided eye (objects that are not within the resolution range of the normal eye). Optical microscopy involve the diffraction, reflection, or refraction of electromagnetic radiation/electron beams interacting with the specimen, and the subsequent collection of this scattered radiation or another signal in order to create an image [1]. The resulting image can be detected directly by the eye, imaged on a photographic plate or captured digitally. The single lens with its attachments, or the system of lenses and imaging equipment, along with the appropriate lighting equipment, sample stage and support, makes up the basic light microscope. The most recent development is the digital microscope, which uses a CCD camera to focus on the exhibit of interest. The image is shown on a computer screen, so eye-pieces are unnecessary - an example is the Epiluminescence microscopy.

Epiluminescence microscopy (ELM) is a non-invasive technique that, by employing the optical phenomenon of oil immersion, makes subsurface structures of the skin accessible for *in vivo* examination and thus provides additional criteria for the clinical diagnosis of pigmented skin lesions. At present, almost all studies about the value and clinical importance of ELM are based on data derived from ELM experts (i.e., dermatologists specifically trained in this technique) [10].

The technique consists in placing mineral oil, alcohol or even water on the skin lesion that is subsequently inspected using a hand-held lens, a hand-held scope (also

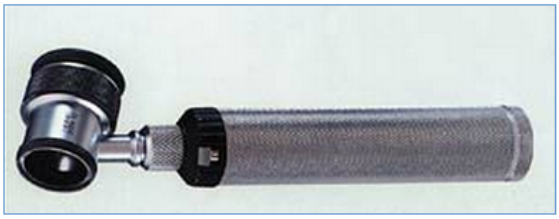


Fig. 2.2: Dermatoscope: instrument widely used by dermatologists, in European countries, for pigmented skin lesions observation [22].

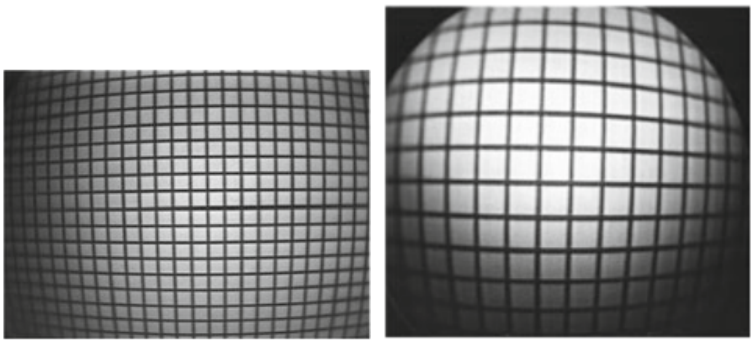
called dermatoscope, Figure 2.2), a stereomicroscope, a camera, or a digital imaging system. The magnifications of these various instruments range from 6x to 40x and even up to 100x. The widely used dermatoscope has a 10-fold magnification permitting a sufficient assessment of PSLs in daily routine. The fluid placed on the lesion eliminates surface reflection and renders the cornified layer translucent, thus allowing a better visualization of pigmented structures within the epidermis, the dermoepidermal junction and the superficial dermis (Figure 1.3.1). Moreover, size and shape of vessels of the superficial vascular plexus can be easily appreciated by this procedure. Nowadays dermoscopy is widely used, especially in European countries, by all physicians managing PSLs, and particularly by dermatologists. Most of them use the traditional dermatoscope (Figure 2.2), but a growing number are already equipped with digital imaging systems [22].

2.3.2 Digital Image Formation

Digital image processing implies a discrete nature of the images. Digitalization of the range value, also known as quantization. It is necessary to determine the maximal number of gray scales for every images. Usually, 8 bit and 24 bit are chosen for gray scale and full color images (RGB color space), respectively, allowing 256 different values in each band/channel. Epiluminescence microscopy of PSL or dermoscopy usually delivers full color images, in RGB color space, with 24 bit per image.

2.4 Image Enhancement

Low-level methods of imaging processing, i.e., procedures and algorithms that are performed without a priori knowledge about the specific content of an image, are mostly applied in pre- or post-processing of medical images (Figure 2.1), in particular, for dermoscopy images. As a special pre-processing method for dermoscopy



(a) (b)

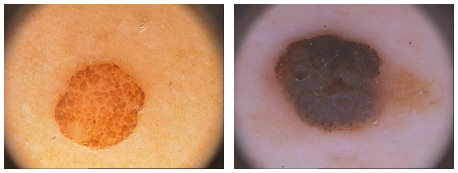
Fig. 2.3: Geometric distortion and brightness variation [23]. By dermoscopic examinations, barrel distortions are often generated, which must be corrected before the image can be analyzed quantitatively. In addiction, the boundary areas in the image appear darker and blurred.

images, techniques for calibration are briefly introduced.

If the physician intents to take quantitative measurements from an image, a careful calibration for the dermoscopy imaging is required. Both, geometry (spatial domain) and brightness or color intensity (value domain) must be adapted to dermoscopy. Calibration is device-specific but disregards the biological content captured, and thus, it is partially low-level processing methods. While reading a dermoscopy image, calibration is made by the dermatologist. However, it must be explicitly implemented for computerized image analysis and measurements.

The positioning of the imaging device must not affect any measurements. In dermoscopic analysis, the devices used originate barrel distortions (Section 2.3.1) (Figure 2.3). Even in simple planar dermoscopy, the objects, which are far away from the image plane, appear larger than those which are located close to the imaging device. This must be kept in mind whenever geometric measurements in digital dermoscopy are taken and displayed to the physicians: point distances in digital images can be converted into length measurements only if a fixed scale is assumed, which is often not fulfilled.

In the same way, the absolute assignment of the pixel values to physical measurements is usually problematic. However, different illumination of the same scene may still alter the captured colors.



(a) (b)

Fig. 2.4: Medical image classification task: examples of an image that have regions corresponding to (a) class A and (b) class B.

2.5 Digital Image Analysis - CADx system

Digital image analysis, in the context of digital image processing, is also known as a CADx system. The CADx system include all the steps of processing which are used for quantitative measurements. The main steps are: (i) segmentation, (ii) feature choice, (iii) classification, and (iv) valuation ([23], [24]); keeping in mind that the feature choice step, comprises two more stages: feature selection and feature extraction.

2.5.1 Segmentation

Segmentation generally means dividing an image into connected and non-overlapping regions. Under this definition, the production of regions is emphasized as the pre-stage of classification. Other definitions accentuate the various diagnostically or therapeutically relevant image areas and, thus, focus the most common application of medical imaging, namely, the discrimination between healthy anatomical structures (healthy skin) and pathological tissue (lesion). By definition, the result of segmentation always refers to identification of the individual patterns (regions inside the image), so, the regions have to be segmented.

Consider the dermoscopy image diagnosis task. Figure 2.4 shows two images, each having a distinct region inside it. The two regions are also themselves visually different. We could say that the region of Figure 2.4(a) results from a benign lesion, class A, and that of Figure 2.4(b) from a malignant one (cancer), class B.

The result of segmentation is exemplified in Figure 2.5. The main goal of this step is to separate the lesion from the surrounding skin.



(a) (b)

Fig. 2.5: Patterns - image regions, corresponding to (a) class A and (b) class B.

2.5.2 Feature Choice

In Figure 2.1, feature choice is defined as the second stage of intelligent (high-level) image analysis. It is followed by classification. Although not explicitly illustrated, this step is composed by two stages: *feature selection* and *feature extraction*. The number of features is a key factor that determines the size of the hypothesis space containing all hypotheses that can be learned from data [34]. A hypothesis is a pattern or function that predicts classes based on given data. The more features, the larger the hypothesis space and thereby the harder the classification [47]. **Feature selection** is essentially a task to remove irrelevant and/or redundant features. It can be achieved in two ways: one is to rank features according to some criterion and select the top k features, and the other is to select a minimum subset of features without learning performance deterioration. In other words, subset selection algorithms can automatically determine the number of selected features, while feature ranking algorithms need to rely on some given threshold to select features [47]. **Features extraction** can be considered as a mapping from the original space to a lower dimensional feature space. The mapping can be carried out with respect to different criteria. The most common features extraction methods are the principal component analysis (PCA) and the linear discriminant analysis (LDA) [47].

In other words, the goal of feature choice is to identify the measurable quantities that make the different patterns *distinct* from each other. Figure 2.6 shows a plot of the mean value of the intensity in each region of interest versus the corresponding standard deviation around this mean. Each point corresponds to a different image from the available database. It turns out that class A patterns tend to spread in a different area from class B patterns. The straight line seems to be a good candidate for separating the two classes. Let us now assume that we are given a new image with a region to be classified. Then we measure the mean intensity and standard

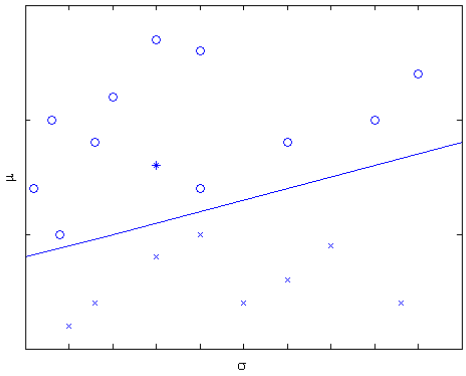


Fig. 2.6: Plot of the mean value versus the standard deviation for a number of different images originating from class A (o) and class B (x). In this case, a straight line separates the two classes.

deviation of that region and we plot the corresponding point. This is shown by the asterisk ($*$) in Figure 2.6. Then it is sensible to assume that the unknown pattern is *more likely* to belong to class A than class B.

2.5.3 Classification

According to the general processing chain (Figure 2.1), the task of the classification step is to assign all connected regions, which are obtained from the *segmentation*, to particularly specified classes of objects, using the features chosen in *feature choice* step.

The classification task has outlined the rationale behind a large class of pattern recognition problems. In the more general case l features x_i, $i = 1, 2, ..., l$, are used, and they from the *feature vector*

$$x = [x_1, x_2, ..., x_l]^T \tag{2.1}$$

where T denotes transposition. Each of the feature vectors identifies *uniquely* a single pattern (object). However, the measurements resulting from different patterns exhibit a random variation. This is due partly to the measurement noise of the measuring devices and partly to the distinct characteristics of each pattern. In dermoscopy images large variations are expected because of the scattering of the differences in the physiology among individuals. This is the reason for the scattering of the points in assigned to each class, shown in Figure 2.6.

The straight line in Figure 2.6 is known as the *decision line*, and it constitutes

the *classifier* whose role is to divide the feature space into regions that correspond to either class A or class B. If a feature vector x, corresponding to an unknown pattern, falls in the class A region, it is classified as class A, otherwise as class B. This does not necessarily mean that the decision is correct. If it is not correct, a *misclassification* has occurred. In order to draw the straight line in Figure 2.6, the labels (class A or B) are known for each point of the region. The patterns (feature vectors) whose true class is known and which are used for the design of the classifier are known as *training patterns* (*training feature vectors*).

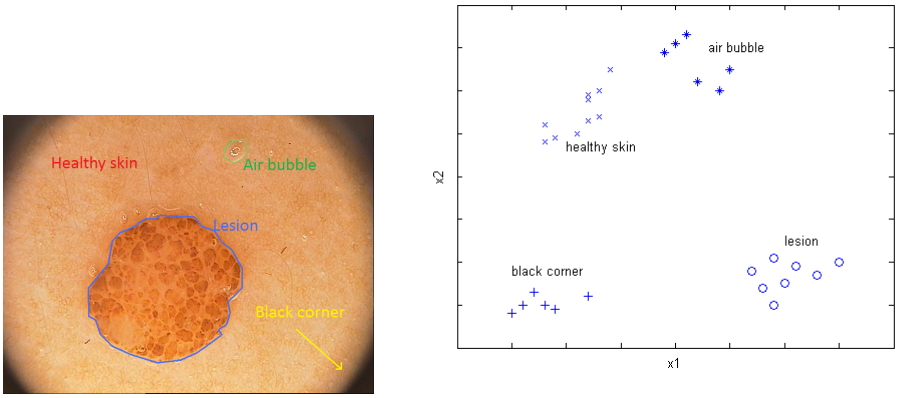
Supervised, Unsupervised and Semi-Supervised Learning

Although a number of basic problems at the heart of the design of a CADx system has been exposed, there are still a few things to be said.

In the example of Figure 2.4, it was assumed that a set of training data were available and the classifier was designed by exploiting this *a priori* known information. This is known as *supervised pattern recognition* or in the more general context of machine learning as *supervised learning*. However, this is not always the case. Actually, there is another type of pattern recognition tasks for which training data of known class labels are not available. In this type of problem a set of feature vectors x is given and the goal is to unravel the underlying *similarities* and *to cluster* (group) "similar" vectors together. This is known as *unsupervised pattern recognition* or *unsupervised learning* or *clustering* [80].

In *multispectral dermoscopic imaging*, the recording of light of PSLs is measured by an optical microscopy (Section 2.3.1), for instance, a dermatoscope (Figure 2.2). The color images formed by the dermatoscope have the information of each basic color band (red, green and blue channel - Section 2.3.2). Thus, each band measures different properties of the same PSL. In this way, images of skin lesions corresponding to the spacial distribution of the reflected light in each band can be created. The task now is to exploit this information in order to identify the various structures types, that is, lesion, healthy skin, black corners, and air bubbles (Figure 2.7). To this end, one feature vector x for each cell from the "sensed" skin's surface is formed. The elements x_i, $i = 1, 2, ..., l$, of the vector are the corresponding image pixel intensities in the various spectral bands. (In practice, the number of spectral bands can vary).

A *clustering* algorithm can be employed to reveal the groups in which feature vectors are clustered in the l-dimensional feature space. Points that correspond to the same structure, such as healthy skin, are expected to cluster together. Figure



(a) (b)

Fig. 2.7: (a) An illustration of various types of structures of dermoscopy images and (b) clustering of the respective features for multispectral imaging using two bands.

2.7 demonstrates the procedure.

A major issue in unsupervised learning is that of defining the "similarity" between two feature vectors and choosing an appropriate measure for it. Another issue of importance is choosing an algorithmic scheme that will cluster (group) the vectors on the basis of the adopted similarity measure. In general, different algorithmic schemes may lead to different results, which the expert has to interpret.

Semi-supervised learning for designing a classification system shares the same goals as the supervised case, however now, the designer has at its disposal a set of patterns of unknown class origin, in addition to the training patterns, whose true class is known. Usually, we refer to the former as *unlabeled* and the latter as *labelled* data [80].

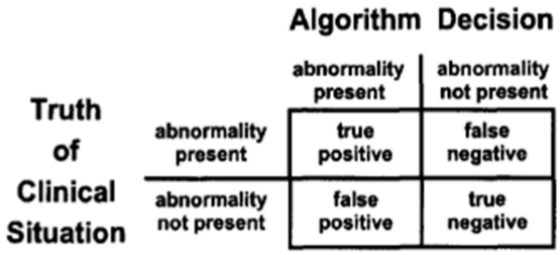
2.5.4 Validation

In the validation step the goal is to confirm that the classes in the previous steps are correct. Validation is strongly related to the more general concept of performance evaluation, see, for example, [84]. Validation usually starts with a visual evaluation of the results, that is, classification results are compared with the original images to see whether the segmentation seems to have been done correctly. However, visual evaluation is not a well-defined way of validating image analysis results. It is always subjective. A *ground truth* evaluation should therefore be made when possible, that

is, images for which the correct features are known should be used in order to make an assessment of how well the image analysis method can match these results.

The set of ground truth images (**validation dataset**) should be representative of the new images that will be encountered later when the image analysis method that is being validated will be used. It is also good idea to perform some worst-case testing, that is, to test what happens when the image contains some rare or unusual features (**test dataset**).

In the case presented, an alternative to the ground truth validation is to let one or more experts repeat classifications of the image data. If the variation between these "human classifications" is as large as it is between the automated classification (Figure 2.6), we know that the accuracy of the automated method is comparable to the "human classification".



(a)

Fig. 2.8: Performance metrics for classification problems [76].

The performance metrics for classification problems involve the relative frequency of correct and incorrect decisions. The terms true positive (TP), true negative (TN), false positive (FP), and false negative (FN) are commonly used in the context of malignancy diagnosis of PSLs (Figure 2.8). Positive/negative refers to the decision made by the algorithm; true/false refers how the decision agrees with the actual clinical state [76].

Several additional performance metrics are derived from the true/false positives/negatives metrics.

- **Accuracy:** it a measure of the global performance of the algorithm in what concerns to correct decisions.

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{2.2}$$

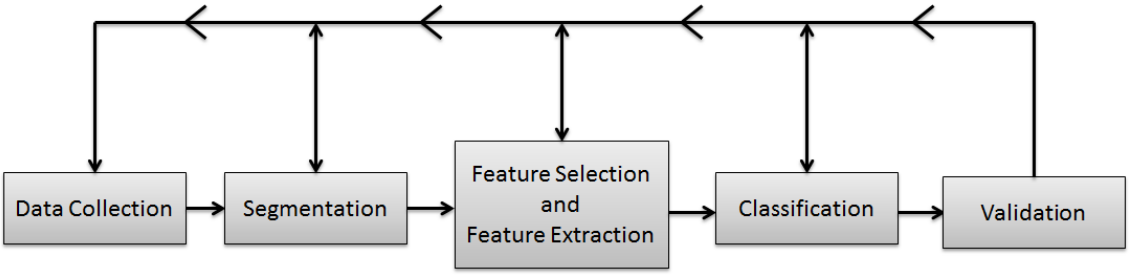


Fig. 2.9: The basic stages involved in the design of a classification system.

- **Sensivity (SE):** it measures how often the algorithm reports that an abnormality exists in the instances where it actually exists. Is also called **true positive fraction** (TPF).

$$SE = \frac{TP}{TP + FN} \tag{2.3}$$

- **Specificity (SP):** it measures how frequently the algorithm correctly reports normal when no abnormality exists. **False positive fraction** (FPF) is the same as 1-specificity.

$$SP = \frac{TN}{TN + FP} \tag{2.4}$$

Figure 2.9 shows the various steps of digital image analysis. As is apparent from the feedback arrows, these steps are interrelated (but not dependent!) and, depending on the results, one may go back to improve the overall performance. Furthermore, there are some methods that combine steps, for example, the feature selection and the classification step, in a common optimization task. Data collection step already contains the stages of formation and enhancement presented in Figure 2.1.

Chapter 3

State-of-the-Art

Diagnosis of the Cutaneous Melanoma (CM) has been a very active research field over the past decades, both in terms of diagnostic algorithms to be applied by human examiners and in terms of skin imaging techniques, which include the support by dermoscopy computer-aided diagnosis (CADx) systems. However, diagnose CM as benign or malign Pigmented Skin Lesions (PSL) still is a challenging task.

This chapter presents the state-of-the-art for dermoscopy image analysis based on CADx system for human dermoscopy and computer dermoscopy.

3.1 Diagnostic Algorithms - Human Dermoscopy

The differential diagnosis for PSL is challenging even for specialists. For this reason, a 2-step approach (Figure 3.1) was suggested by a virtual consensus meeting of 40 actively participating experienced clinicians known as the Consensus Net Meeting on Dermoscopy (CNMD). In the first step, a classification in melanocytic and non-melanocytic lesions is performed, following a 6-step algorithm. In the second step, melanocytic lesions are classified in benign and malignant. It is primarily the second step that is currently addressed by dermoscopy CADx.

Current diagnostic algorithms in dermoscopy evaluated by the CNMD [16] are summarized in Table 3.1.

The following sections present the current diagnostic algorithms that the second step involves.

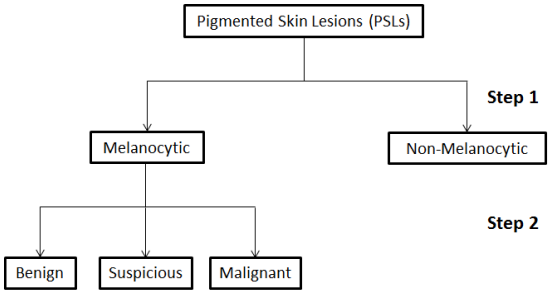


Fig. 3.1: *Flow chart* for the differential diagnosis of pigmented skin lesions is performed with a 2-step approach.

Tab. 3.1: *Evaluation of diagnostic algorithms.* The algorithms were evaluated by the CNMD regarding the inter- and intra-observer agreement [16].

Algorithm	**Intra-observer** κ (95% CI)	**Inter-observer** κ (range)
First-step	0.63 (0.62-0.63)	1.00 (0.73-1.00)
Pattern analysis	0.55 (0.54-0.56)	0.85 (0.12-1.00)
ABCD rule	0.48 (0.47-0.48)	0.72 (0.11-1.00)
Menzies method	0.52 (0.51-0.52)	0.75 (0.21-1.00)
7-point checklist	0.47 (0.46-0.47)	0.72 (0.29-1.00)

3.1.1 Pattern Analysis

Pehamberger et al. introduce the *pattern analysis* as diagnostic algorithm of PSL in 1987 [58]. This method is based on pattern analysis, as the name indicates. Pattern analysis ([15],[57]) performed by physicians allows distinction between benign and malignant growth features. Typical patterns of some common melanocytic lesions are: dots, globules, streaks, blue-white veil, blotch, network, network borders.

Pattern analysis algorithm is very often used because it increases the number of melanocytic lesions correctly diagnosed by 10 to 30% [6]. However, this method presents some problems especially in the reliability of the results, since these are qualitative assessments, which naturally vary slightly and significantly from doctor to doctor. As such, three additional diagnostic methods have been introduced to increase the diagnosis sensitivity: ABCD rule, 7-point checklist, and Menzies method.

3.1.2 ABCD rule

The ABCD rule of dermoscopy was the second algorithm developed, after the pattern analysis, in an attempt to simplify and improve the process. This diagnostic

algorithm was introduced by Stolz et al. in 1994 [79]. After a rigorous multivariate analysis of 31 dermoscopic features, only 4 features were considered important cofactors to the melanocytic lesions diagnosis.

The Asymmetry, Border, Color, and Differential structures features of the PSL constitutes the ABCD rule ([54], [78]). The ABCD criteria are semi-quantitative: it is a mathematical approach that assigns a score to each of the criteria identified in the PSL:

1. *Asymmetry*, 0-2 points: 0 = symmetry in two axis; 1 = symmetry in one axis; 2 = no symmetry at all;
2. *Border*, 0-8 points: number of half-quadrants with diffuse border;
3. *Color*, 1-6 points: number of colors appearing in the lesion (white, red, light-brown, dark-brown, blue-gray, black); and
4. *Differential structures*, 1-5 points: number of differential structures in the lesion (pigment network, dots, globules, streaks, structureless areas)

with particular weight factors and sums up to a total score ranging from 1.0 to 8.9. Depending on this Dermatoscopic Point Value (DPV), $DPV = 1.3A+0.1B+0.5C+0.5D$, a lesion is considered benign ($DPV = 1.00-4.75$), suspicious ($DPV = 4.75-5.45$), or malignant ($DPV = 5.45 - 8.90$). Later, a fifth criterion called evolution was added to describe morphologic changes in the lesion over time (ABCD$\underline{\text{E}}$ rule).

3.1.3 Menzies Method

Menzies method for melanocytic lesions diagnosis was introduced in 1996 by Menzies et al. [53]. For a diagnosis of melanoma according to the Menzies method both of the two negative features (lesion shows only a single color; patterns are point or axial asymmetric) must *not* be found, and at least one of 9 positive features (blue-white veil; multiple brown dots; pseudopods; radial streaming; scar-like depigmentation; peripheral black dots-globules; multiple colors; multiple blue/gray dots; broadened network) must be found.

3.1.4 7-Point Checklist

In 1998, Argenziano et al. [7], proposed the 7-point checklist in an attempt to find a less complex method of melanocytic lesions diagnosis. This method distinguishes

3 major criteria, each with a score of 2 points (atypical pigment network; blue-white veil; atypical vascular pattern) and four minor criteria, each with a score of 1 point (irregular streaks; irregular pigmentation; irregular dots/globules; regression structures). Diagnosis of CM requires a minimum total score of 3, out of a maximum of 10 points.

3.1.5 Other diagnostic algorithms

Two more diagnostic algorithms which were not evaluated by the CNMD are the *3-point score* and the Color, Architecture, Symmetry, and Homogeneity (*CASH*) system [14]. Both are based in a total score that permits to diagnose the melnocytic lesion as benign, malignant or suspicious.

3.2 Diagnostic Accuracies

Compared to purely visual examination, dermoscopy is clearly more accurate [8] - Section 1.4. According to a meta-analysis by Vestergaad et al. [82] from 2008, the relative diagnostic odds ratio for CM in this comparison is 15.6 (95% CI 2.9-83.7, $p = 0.016$) for 9 eligible studies published from 1987. Similarly, CADx systems accuracies on dermoscopic images outperform CADx systems accuracies on clinical images (log odds ratios 4.2 vs. 3.4, $p = 0.08$) [67].

Annessi et al. [4] compared the performance of three diagnostic algorithms - pattern analysis, ABCD rule, 7-point checklist - on doubtful PSLs in 195 patients: 198 consecutive atypical melanocytic lesions, of these 102 Atypical Melanocytic Nevi (AMN) (Clark's melanocytic nevus, benign), 96 Thin Melanoma (TM) (malignant, 24 *in situ* melanomas and 72 SSMs with an average tumor thickness of 0.3 mm). Two dermoscopy-experienced dermatologists classified the lesions as benign and malignant using the three algorithms. Surgical excision followed. The diagnostic accuracy, sensitivity and specificity (see definitions in Section 2.5.4) were: 70.8%, 85.4%, 79.4% for the pattern analysis, 67.8%, 84.4%, 74.5% for the ABCD rule, and 57.7%, 78.1%, 64.7% for the 7-point checklist, respectively.

There is evidence that dermoscopy CADx systems can reach the performance of CM experts and help improve early CM diagnosis [29]. Rajpara et al. [64] reviewed 765 articles and performed a meta-analysis on 30 eligible studies comparing human dermoscopy to computer dermoscopy. They found the pooled sensitivity for computer dermoscopy was slightly higher than for human dermoscopy (91% vs. 88%,

$p = 0.076$), while pooled specificity was significantly better for human dermoscopy (86% vs. 79%, $p = 0.001$). In the meta-analysis of Rosado et al. [67] from 2003 on 10 eligible studies with a total of 6579 cases CADx systems diagnosis was statistically not different from human diagnosis (lod odds ratios 3.36 vs. 3.51, $p = 0.80$).

3.3 Dermoscopy Image Analysis - Computer Dermoscopy

The analysis of digital dermoscopy images can follow different approaches, either driven by the medical expert perspective, or driven by a machine intelligence perspective. Generally, these approaches share basic steps in the processing pipeline, including: segmentation, feature choice, and classification.

3.3.1 Image Analysis Approaches

Two major approaches can be distinguished in dermoscopy digital image analysis for melanoma diagnosis [23]

1. *Machine learning approach* or *black box approach*. Typically, there is an extraction of a large number of low-level features which allow the discrimination between melanoma and benign PSL. Depending on each analysis strategy, the number of low-level features used greatly reduces and a classifier is constructed based on this new number of features. In general, it is not transparent to the human user what such features really measure and how the classifier comes to its conclusion.

2. *Mimic approach*. Typically, a small number of high-level diagnostic features used by medical experts for visual evaluation of the images are modeled mathematically. Semi-quantitative diagnostic algorithms such as the ABCD rule are then automated using the computed features instead of the human scores. Mainly, the user can easily understand the meaning of the features and the way the proposed diagnosis is created by the classification algorithm.

In practice, both approaches seldom appear in a pure form. Rather, following a mimic approach, a system developer will slightly extend the number of features (e.g., by introducing variants) and adapt the classification algorithm, if this improves the performance of the system and the accessibility of the users. On the other hand,

applying the black box approach, the number and type of features as well as the choice of classifiers can be steered to gain transparency.

An example for the black box approach is Blum et al. [12], where the researchers did not follow a preformed strategy (e.g., ABCD rule). Instead, the authors applied a large number of algorithms of vision algebra to the pictures. In particular, the authors used 64 analytical parameters to build their classifier.

An example for the mimic approach is Horsch et al. [37]. This research group used high-level features in line with the ABCD rule of dermoscopy.

3.3.2 Segmentation of Skin Lesions

The first step in the analysis of a dermoscopic PSL image is segmentation, i.e. the automatic or semiautomatic separation of the lesion from the surrounding skin [40].

A semi-automatic segmentation can be done manually by experienced physicians. In [27] a graphical user interface for computer-aided manual segmentation of dermoscopic images is presented. This tool allows building up a reliable ground truth database of manually segmented images. Besides the manual segmentation of the lesion, this tool also allows the segmentation of other specific regions of interest (such as color regions), which are essential for the development of newly computer-aided diagnosis systems.

Several algorithms and techniques for automatic segmentation of images have been developed, without a general solution to this problem. However, this step is applied in two domains: the feature domain and the image domain. Within the feature domain, clustering techniques and thresholding are applied. In the image domain, the result of image segmentation is a set of segments that collectively cover the entire image, or a set of contours extracted from the image (edge detection). Each of the pixels in a region are similar with respect to some feature or computed property, such as color, intensity, or texture. Adjacent regions are significantly different with respect to the same feature(s) [71].

For example, the histogram analysis or thresholding of the dermoscopic image provides the distribution of image pixels, i.e., the number of pixels in the feature domain for each light intensity. Consider the intensity histogram showed in Figure 3.2(c) that corresponds to the grayscale image, $f(x, y)$ (Figure 3.2(b)). As we can see in Figure 3.2(c), the histograms of dermoscopy images are frequently bimodal: one peak corresponds to the area of the lesion and another peak corresponding to the surrounding skin. Thus, the valley between the two major peaks, is considered

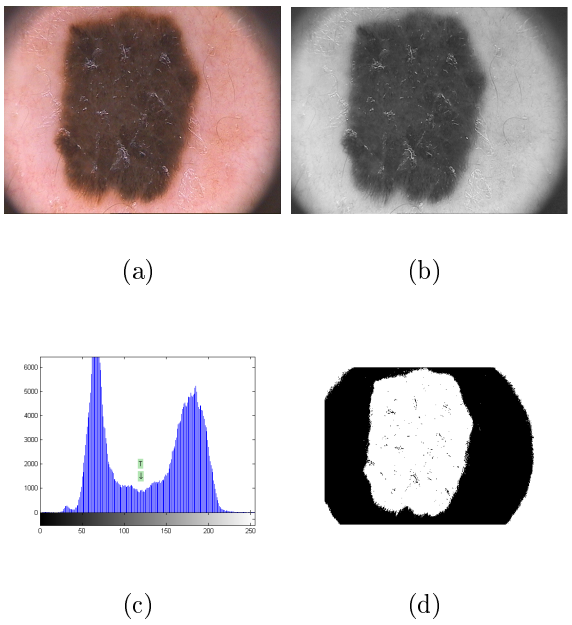


Fig. 3.2: Thresholding as segmentation technique in feature domain: (a) original image; (b) gray-scale image; (c) histogram of image (b); (d) binary image for threshold T=120.

a good approximation to the threshold separation between these two zones.

The thresholding applied to the image is given by:

$$f_T(x,y) = \begin{cases} 1 & if f(x,y) \leq T \\ 0 & otherwise \end{cases}, \tag{3.1}$$

where T is the threshold that separate the two regions (lesion and surrounding skin - Figure 3.2(c)). Pixels labeled 1 correspond to the lesion, whereas pixels labeled 0 correspond to surrounding skin (for instance, see Figure 3.2(d)). When T is constant, this approach is called *global thresholding* [32].

Otsu [56] proposed an image segmentation method which consists in choosing the parameter T such that the between-classes variations is maximized, and the intra-classes variation is minimized. The parameter T is chosen automatically without operator intervention.

Another segmentation procedure, employing a hybrid method that combines clustering of the color space and hierachical region growing, is used in the Artificial Neural Network (ANN) classifier of the DANAOS expert system [35]. A validation and comparison of automatic segmentation with manual segmentation by the

physician by drawing the lesion outline into the image can be found in [40].

Ganster et al. [28] also proposed a hybrid method for the segmentation step that combines thresholding and clustering of the color space.

In Silveira et al. [74] is presented the evaluation of six methods for segmentation of skin lesions in dermoscopic images.

In summary, the common segmentation methods are: thresholding (global or local/adaptative thresholding), edge detection (discontinuities), region growing, texture thresholding, and multi-spectral techniques.

3.3.3 Classification Methods

A type of classification method frequently used in a dermoscopic CADx system is the Artificial Neural Networks (ANN) (Figure 3.3). The ANN is trained by a set of features computed from images of *nevi* with know malignancy. Applying the trained ANN to a feature vector of an unknown PSL will result in a certain malignancy score (in the simplest case with only two possibilities, *benign* and *malignant*). Unless other methods are used in parallel for computation and visualization of diagnostic criteria, the way in which the ANN come to its conclusion is normally not transparent to the user. This has often been criticized as one of the major drawbacks of ANN. An example of a system using an ANN classifier is the DANAOS expert system ([35], [51]). For this system it has been shown that the area under the Receiver Operating Characteristic (ROC) curve (AUC) increases rapidly with an increase of training dataset size from 50 images ($AUC = 0.67$) to 2.000 images ($AUC = 0.86$).

Classification methods appropriate for the mimic approach are e.g. logistic regression (Logreg) and Classification And Regression Tree (CART). In [31], CART analysis was performed in order to differentiate between melanocytic skin lesions and surrounding skin. Instance-based learning (1-NN) was tested for differentiating between benign and malignant tumour elements. For diagnostic assessment, only the percentage of elements suggestive for malignancy in each lesion was used. A total of 369 standardised dermatoscopic images (93 melanomas, 276 benign *nevi*) were evaluated. When instance-based learning was applied to an independent test set, a threshold value of 27.4% of elements suggestive for malignancy recognised 35 out of 35 melanomas and 100 out of 101 *nevi* (sensitivity 100% and specificity 99%).

In [23], both methods (Logreg and CART) have been applied to a set of 466 PSLs (125 melanomas, 341 benign *nevi*), using the 5 mimic features A_H (field color asymmetry), A_C (pixel color asymmetry), B_SIM (border diffusiveness), C_H (color

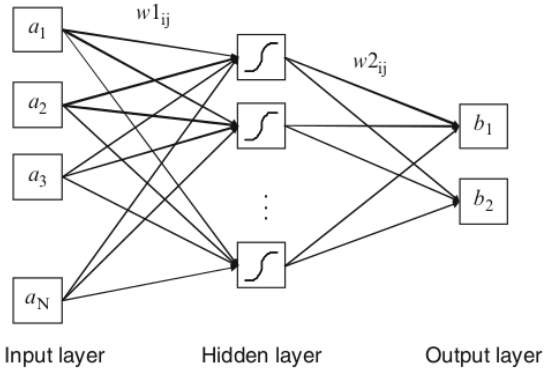


Fig. 3.3: Principle of a 3-layer ANN.

Tab. 3.2: Classifier comparison on 466 PSLs (125 CM, 341 benign PSL).

Classifier	Validation	ρ	SN	SP
Logreg	10 x 10-fold cross-validation	0.895	0.862	0.911
CART	10-fold cross-validation	0.89	0.76	0.94
ANN (linear layer)	10 x 100/100 train/test	0.896	0.896	0.896

variety) and S_L (lesion size).

For comparison, in [23], ANN has also been trained and tested with a cross-validation scheme on the same set of 466 PSLs. Figure 3.3 shows the principle layout of the employed 3-layer ANN, a so-called linear-layer perceptron. The weight factors $w1_{ij}$ and $w2_{ij}$ (see Figure 3.3) reflect the knowledge about how to distinguish CMs from benign PSLs. These are the factors to be trained by a suitable learning strategy such as error back propagation. The input layer receives the feature values and the classification result is obtained at the output layer: $b_1 = malignant$ and $b_2 = benign$. Assuming five input nodes ("neurons") for the five features, four hidden layer nodes, and two output layer nodes, this network has 5x4+4x2=28 weight factors that had to be learned from the training set in each of the cross-validation tests.

Classification results achieved with Logreg, CART and ANN classifiers are compared in Table 3.2. The cross-validated total correct-classification rate ρ is around 0.89 for all three types of classifiers. Logreg and CART show, at the operating point that minimizes the distance between the ROC curve and the ideal classifier point ($sensitivity = specificity = 1$), specificities superior to sensitivities due to the fact that the dataset contained less CMs (125) than benign PSLs (341) and therefore correct assessment of benign lesions is trained better than of malignant lesions. In contrast to this, the ANN compensates different class sizes ($sensitivity = specificity$).

Another classifiers are often used as: Bayesian classifiers ([69]), k-Nearest Neigh-

bor (kNN) ([28], [17], [69]) and Support Vector Machines (SVM) ([19]).

- **Bayesian classifier.** It is based on Bayes decision theory (Figure 3.4), a fundamental statistical approach: quantifying the tradeoffs between various classification decisions using probability and the costs that accompany such decisions. It makes the assumption that the decision problem is posed in probabilistic terms, and that all of the relevant probability values are known (*a priori* probability) [24].

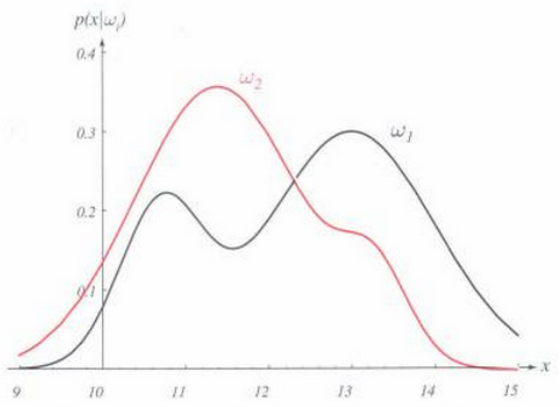


Fig. 3.4: Hypothetical class-conditional probability density functions show the probability density measuring a particular feature value x given the pattern is in category w_i [24].

- **kNN.** It is a method for classifying objects based on closest training examples in the feature space. kNN is a type of instance-based learning where the function is only approximated locally and all computation is deferred until classification. The kNN algorithm is amongst the simplest of all machine learning algorithms: an object is classified by a majority vote of its neighbors, with the object being assigned to the class most common amongst its k nearest neighbors (k is a positive integer, odd and typically small). If $k = 1$ (Figure 3.5), then the object is simply assigned to the class of its nearest neighbor.

- **SVM.** Given an implicit embedding ϕ and training data (x_i, y_i) from 2 classes such that $y_i == {}^{+}_{-}1$, a SVM finds a hyperplane $w^T\phi(x) + b = 0$ that best separates the two classes (see Figure 3.6). The learnt hyperplane is optimal in the sense that minimizes the margin while minimising some measure of loss on the training data [61].

[2] http://blog.csdn.net/alaclp/article/details/7457290

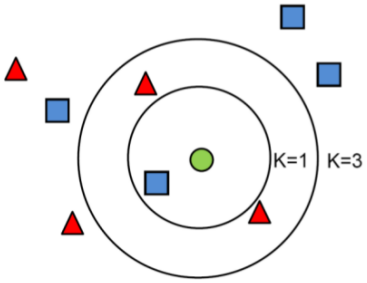


Fig. 3.5: kNN classification method. The object to be classified (green circle) will be classified as belonging to the class of blue squares if K = 1, and classified as belonging to the class of red triangles case k = 3.

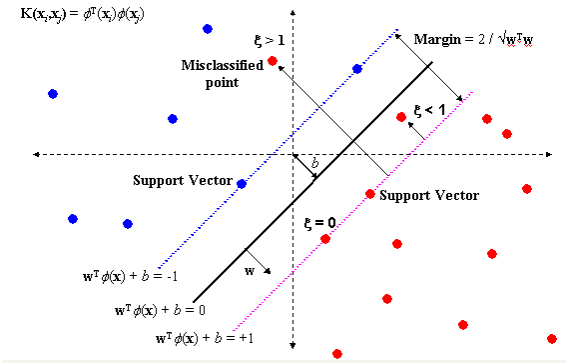


Fig. 3.6: The SVM learns a hyperplane which bests separates the two classes. Red dots have label $y_i = +1$ while blue dots have a label $y_i = -1$[2].

3.4 Evaluation Issues

High-quality, state-of-the-art evaluation of CADx systems is a challenging task. Attempts to give reliable estimates of the performance of such systems are hampered by different types of biases caused by limitations in resources and study designs. This may, to a certain extent, explain why the methodological quality of studies is generally not good ([64],[65],[67]).

3.4.1 Clinical Case Databases

Particularly critical for evaluation is the composition of clinical case databases used for evaluation. In order to guarantee that a case database reflects realistic clinical conditions, the choice of cases must meet certain requirements [67]:

- Random or consecutive selection of lesions (to avoid selection bias).
- Clear definition of inclusion and exclusion criteria.

- Inclusion of all lesions clinically diagnosed as PSL.
- Inclusion of clearly benign lesions that were not excised (to avoid verification bias); diagnostic gold standard in these cases is short-term follow-up.

In order to gain sufficient accuracy in statistical terms, the size of the case database has to be large enough. Depending on the classification method under evaluation and the narrowness of the Confidence Interval (CI) aimed at point estimates such as sensitivity and specificity, datasets will have to be of size of some hundreds to some thousands of cases with biopsy-proven and clearly benign PSLs.

3.4.2 Evaluation Methods

In order to avoid methodological weaknesses, requirements have to be added to the list of database-related requirements of Section 3.4.1 ([67], extended):

- Clear definition of the study setting.
- Reporting of instrument calibration.
- Intra- and inter-instrumental repeatability.
- Classification on independent test set.
- Comparison of computer diagnosis with human diagnosis.
- Comparison of human diagnosis without and with CADx support.

Some publications indicate that there is awareness for the problem, especially in the field of lung and breast cancer CADe and CADx (see e.g. [25], [36]). But the implementation of good practice addressing the above listed requirements is still a great challenge also in these fields.

Table 3.3 summarizes some recent methods used for dermoscopy image analysis and the respective statistical measures (SE and SP - Section 2.5.4) of the performance of the classification step.

Tab. 3.3: State-of-the-art methodologies summary.

Author	Year	Segmentation Method	Feature Extraction	Classifier	SE	SP	Comments
Ganster et al. [28]	2001	Tresholding + Clustering of color space	122 features	kNN	73%	89%	Features based on ABCD rule
Rubegni et al. [68]	2002	Thresholding	48 features	ANN	94%	94%	Features grouped in 4 categories (geometry, color, texture, and islands of colors)
Burroni et al. [17]	2004	Laplacian filter + Zero-crossing algorithm	49 features	kNN + Linear classifier	98% + 95%	79% + 78%	Features grouped in 4 categories (geometry, color, texture, and islands of colors)
Blum et al. [12]	2004	–(a)	64 features	Linear regression	82%	87%	–(b)
Celebi et al. [19]	2007	Region Growing	437 features	SVM	93%	92%	11 features of shape, 354 features of color and 72 features of texture
Celebi et al. [18]	2008	Manual (with supervision)	Blue-whitish veil	CART	69%	90%	Features obtained by the 7-point checklist algorithm
Iyatomi et al. [39]	2008	Clustering + Region Growing	428 features	ANN	86%	86%	Features obtained by the ABCD rule
Ruiz et al. [69]	2011	Adaptative Thresholding	4 features	kNN + Bayesian classifier + ANN	77% + 70% + 78%	95% + 78% + 98%	Features obtained by the ABCD rule

(a) not mencioned; (b) without relevant comments

SE = sensibility, SP = specificity

3.5 Commercial Systems

Currently there is a wide offer of dermoscopy devices and systems. Beyond common system design principles, the products present some essential differences in terms of image acquisition devices, system functionalities, and the approach and degree of CADx support.

3.5.1 System Design Principles

The design of dermoscopy systems for the medical practice is commonly characterized by the goal of supporting the entire workflow of diagnostic decision making and decision management. Hereby, digital image analysis is one aspect only. Required functions of a dermoscopy system include: electronic patient record; image acquisition and management; report creation and storage; image analysis and decision support; support of continued medical training; tele-consultation (optional).

Important features of a dermoscopy system built according to these requirements include: patient-, case- and lesion-based documentation; follow-up support by means of (qualitative and quantitative) comparison of the same lesion at different examination dates; automatic computation of scores derived from diagnostic algorithms (Section 3.1) such as ABCD rule or the 7-point checklist; comparison of pathohistologic report with the corresponding dermoscopic image, so that the physician can validate his diagnostic skills; support of diagnostic skills training by offering similar cases from a reference database; explanations and visualizations of diagnostic scores and the features employed for a skin lesion under examination.

3.5.2 Image Capture Devices

Common image capture devices use calibrated Charge-Coupled Device (CCD) video camera technology or digital still image camera technology for dermoscopic image acquisition ([62], [83]). The devices acquire either one image in the visible light spectrum (mono-spectral imaging), or a set of images in different wavelength bands (multi-spectral imaging). The latter is motivated by the fact that light of different wavelengths penetrates the skin to different depths.

Tab. 3.4: Commercial dermoscopic systems with image analysis component ([62]-[50]).

System	Company	Spectrum	Image analysis
DB-Mips	Biomips Engineering	mono	ANN, similarity
DermAssist	Romedix	mono	$clin - ABCD^a$
DermoGenius ultra	BIOCAM	mono	$ABCD^b$, $ref - DB^c$
Fotofinder dermoscope	FotoFinder Systems	mono	ref-DB
MelaFind	Electro-Optical-Sciences	multi	ref-DB
MicroDerm	Visiomed AG	mono	ANN, ABCD
MoleMax II/3	Derma Medical Systems	mono	ABCD, 7-point
MoleView	Astron Clinica	multi	$SIAscopyscore^d$, 7-point
Solarscan	Polartechnics	mono	ref-DB
VideoCap 100	DS Medica	mono	ref-DB

(a) clin-ABCD = clinical ABCD rule with D for diameter; (b) ABCD = dermatoscopic ABCD rule with D for differential structures; (c) ref-DB = reference image database; (d) SIAscopyscore = proprietary scoring scheme

3.5.3 Dermocopy Computer Systems

Several commercial dermoscopy systems with integrated image analysis components are on the market (Table 3.4)([50]-[62]). Most of these systems support both microscopic (typically 5- to 15-fold magnification) and macroscopic imaging (e.g., EasyScan TS Pico, FotoFinder dermoscope). In addiction, some systems comprise room cameras capable of capturing whole body images (mole mapping) in a standardized way to aid the detection of new moles (e.g., MoleMax II, Molemax 3, DB-Mips). The systems are therefore equipped with one up to three digital video and still image cameras of specific types.

3.6 Concluding Remarks

In the published literature, the accuracy reached by current dermoscopy CADx systems for CM diagnosis is comparable to the accuracy achieved by human specialists. But this evidence is still weak. Large randomized controlled trials evaluating dermoscopy CADx systems under routine conditions are lacking. Many dermatologists therefore continue to be sceptical about the routine use of such systems [65]. Obviously, there is a need for large, high-quality clinical case databases to support system validation. Future trials should not only evaluate the accuracy of the computer system and compare this accuracy to that of human experts, but should also compare the diagnostic performance of humans alone with the performance of humans using

a dermoscopy system as second opinion.

New imaging techniques, such as multi-spectral imaging, could be a next step in the development of more accurate systems for the classification of PSLs. Intensified efforts towards an extension of the computer support to non-melanocytic skin lesions are another important step forward. To develop and validate sophisticated algorithms and systems supporting such new imaging techniques for an extended range of skin lesions and bring them into routine use, embedded in the physician's work flow as one of several components of computer support - electronic patient record, whole body imaging (CADe), tele-dermatology, and others - will demand a lot of research and innovative system development. But certainly there is the potential to further improve early recognition of CM and other threatening diseases of the human skin.

Chapter 4

Proposed Method

This chapter presents all the theoretical basis of the proposed method for the analysis of dermoscopic images.

Firstly, a description of the proposed method is presented. In Section 4.2 and 4.3 the dataset used and the way how color regions were acquired are presented. The following section, 4.4, introduces and describes the image calibration techniques applied, and Section 4.5 explains some color models used in the development of the proposed method. Section 4.6 refers the separability measures or separability distances between classes which are very useful measures to assess in first-hand the performance of the classifier. Some classes can have separability problems, in that case intra- and inter-classes tests can be performed, as addressed in a qualitatively way in Section 4.7. This section also refers to grouping classes in order to rearrange data for the classification task. A non-linear technique for increasing clusters separability is also implemented in the proposed algorithm, as presented in Section 4.8. The proposed classification of artificial neural networks in competition in a transformed space where the classes are well separable from each other is described in Section 4.9. The post-processing and validation are also referred in the last sections of this chapter.

4.1 Description

The standard methods for dermoscopy image analysis, or CADx system, generally involve 4 main steps (Section 2.5): (i) segmentation, (ii) features choice, (iii) classification, and (iv) validation (classifier evaluation). All these steps contribute to the performance of the CADx system.

The method proposed is not a standard dermoscopy image analysis, because it does not require the segmentation step and the sub-step of feature extraction of a CADx system. As already mentioned in Section 2.5.4, although the standard approach involves this 4 steps, they are not dependent.

The proposed method takes the following steps: **(i) feature selection**, **(ii) classification**, and **(iii) validation**.

Menzies et al. [53] proposed a method for diagnosis of PSLs (Section 3.1.3) and they state, as one criteria, that a lesion must be considered malignant or suspicious (Figure 3.1) if it contains more than one color. Lamarão and Cabral [45], reinforced the idea that the presence of color in a PSL is a strong indicator of malignancy of the lesion. According to that, the only feature that will be collected, in order to decide whether a PSL is malignant or not, will be color and so there is no need for the feature extraction step.

The dermoscopic colors that can be present in a PSL are: white (Wh), red (Re), light-brown (LB), dark-brown (DB), blue gray (BG), and black (Bk) [53].

In Figure 4.1 the flowchart of the proposed method is presented. After the image is acquired by a dermoscope, the physician classifies the PSL as melanocytic or non-melanocytic (Figure 3.1). If the lesion is classified as melanocytic, the corresponding image will be analysed. The procedure consists in the selection of color features using different color models and some data transformation, resulting in the input data (well separable clusters in the feature space) of the classifier. The classified images are subjected to post-processing, as filtering and "black corners" removal. Once the final classified images are obtained, Equation 4.1 is applied to decide if the melanocytic PSL should be classified as benign or suspicious (in the worst case is malignant). At the final step, an evaluation of the machine vs. human classification is performed, in order to decide if the classifier can mimic the human procedure of identifying dermoscopic colors.

$$PSL = \begin{cases} benign & \#colors = 1 \\ malignant \vee suspicious & \#colors \geq 2 \end{cases} \tag{4.1}$$

This procedure is fixed for the training data, which serve to teach the classifier the surface of each class (with local and global structures of the surface). The training data uses images from the dataset available (see Section 4.2), where color regions were segmented by physicians. Thus, the classifier will learn the correct dermoscopic color data. When the color classes are already stored (Figure 4.1), some image

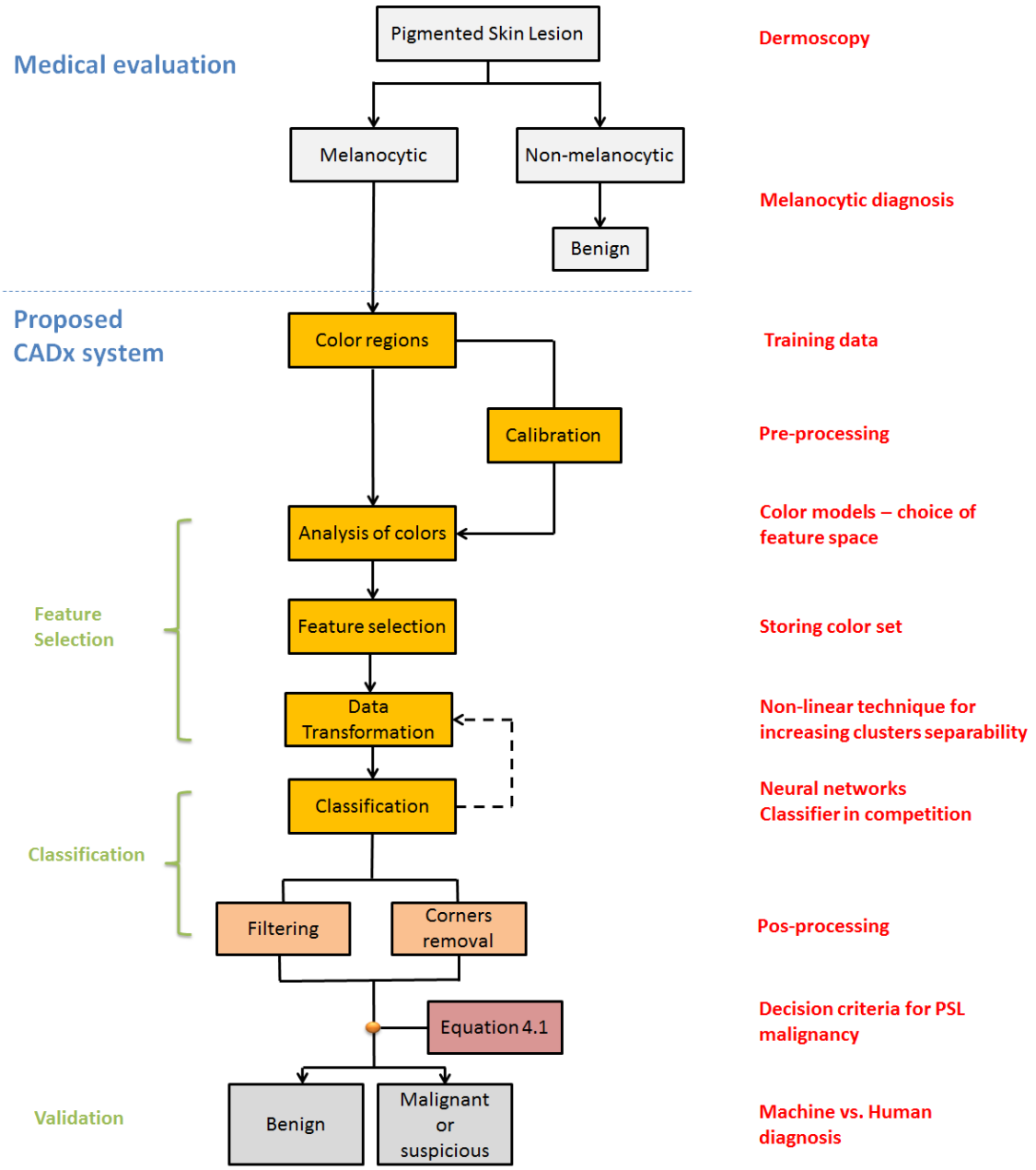


Fig. 4.1: Classification algorithm of dermoscopic PSL: flowchart of the proposed method.

calibration techniques (Section 4.4) can be applied to the data in order to reduce spectral variability of each class. Thereafter, the color features are chosen using different color models (Section 4.5) and measuring separability distances (Section 4.6) to assess the features choice. Some data transformation can also be applied to selected features in order to increase the separability distance between classes (Section 4.8). According to Figure 4.1, the following step is the lesion classification.

The transformed data in the feature space chosen (ideally, with high separability distances between clusters) is the input data of the classifier. In the proposed method an artificial neural network (ANN) was chosen as classifier with supervised learning. NNs in competition (Section 4.9) lead to loss of arbitrariness and perfect knowledge of each cluster (Wh, Re, LB, DB, BG, Bk) surface. Since the data is completely separable in the feature space, the choice of the classifier is highly correct. However, some isolated points can occur in the classified images due to poor quality of images or air bubbles or even hair (Figure 2.7). To solve that problem some image post-processing will be applied, such as filtering (to "normalize" labels) and black corners removal (because they lead to false labelling and are not needed to diagnose images), see Section 4.10. Besides the training set, the classification task uses the test set in order to test the reproducibility capacity of the NNs for new points (different from those used in the training set). Finally, the validation step (Section 4.11) is applied in order to decide if machine learning can mimic human diagnosis in the dermoscopic color identification. But for this, physicians already classified all images as benign, malignant or suspicious. Because of that, and according to the color presence indicated by the machine, it is possible to perform the validation step.

To collect training dermoscopic color data, Marta Pereira, M.D., physician of Hospital Pedro Hispano[1] dermatology service, used a segmentation technique presented in [27] to segment the color regions.

The working tool used in the following sections was MATLAB: signal processing and neural networks toolboxes.

4.2 Dataset

The dataset of images available for this work comes from the database of Hospital Pedro Hispano (HPH) dermatology service[2], directed by Jorge Rozeira, M.D.. The dataset is a working data for the research group Automatic computer-based

[1] http://www.ulsm.min-saude.pt/servicecontent.aspx?menuid=203
[2] http://www.ulsm.min-saude.pt/servicecontent.aspx?menuid=203

Tab. 4.1: Presence of color in the dataset.

No. of colors	Wh	Re	LB	DB	BG	Bk	Images (IMDxx)
			x				05, 08, 10, 14, 23, 24, 25
1				x			02, 07, 15, 16, 21
						x	18
			x	x			01, 03, 04, 06, 09, 11, 22
2				x	x		12, 19, 26
		x	x				27
			x			x	13
	x			x	x		17
3	x				x	x	20
			x	x	x		28

Wh = white, Re = red, LB = light brown, DB = dark brown, BG = blue gray, Bk = black

Diagnosis system for Dermoscopy Images[3] (ADDI) of Faculdade de Ciências da Universidade do Porto[4] (FCUP), and it contains 206 dermoscopy images (52 non-melanocytics + 154 melanocytics in which 135 are nevus and 19 are melanomas).

A subset of 28 images (IMD01-IMD28) was selected for medical evaluation, since only those (28 melanocytics: 23 *nevus* + 5 melanomas) were obtained with, at least, a safe medical validation. These images will be used to present the results of the proposed method and can be seen in Appendix A. Table 4.2 shows the presence of colors in each of these 28 images.

The medical specialist used the software tool described in [27] that allows manual segmentation of the different color classes present in each image. The segmented regions are stored as logical masks. As an example, Figure 4.2 shows the original image IMD17 (Figure 4.2(a)) (which according Table 4.2 has the color Wh, DB and BG) and the respective logical masks (Figure 4.2(b),(c),(d)), for Wh, DB and BG color classes respectively). The opinion of two medical specialists contributed to the construction of Table 4.2. Both specialists indicate the presence of colors, but only one segmented or identified the corresponding areas (e.g. Figure 4.2) for each color (Table 4.2).

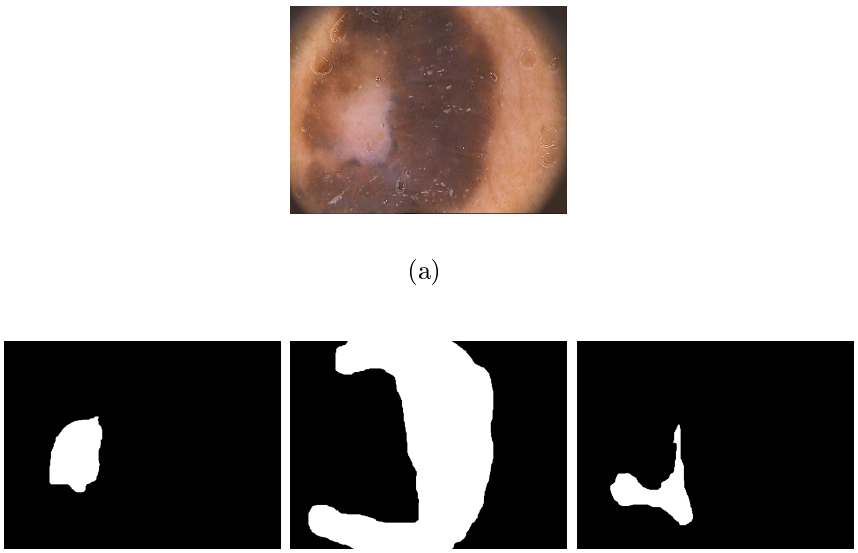
The observation of dermoscopic colors is a subjective process and depends on the experience of the examiner. As we can see in Table 4.2, the same physician indicated different colors for some images on different occasions. It is difficult even for specialists to clearly identify the presence of dermoscopic colors. There are many possible reasons for this problem: poor quality of images, ambiguity between some

[3] http://www2.fc.up.pt/addi/
[4] http://www.fc.up.pt/

Tab. 4.2: Colors indication and color regions identification. The indication of colors was made by two physicians, A and B, with clearly differences of color perception. The same physician A segmented the color regions (second observation) and indicated and identified different colors than in the first evaluation.

Images (IMDxx)	Indication		Identification
	Physician A	**Physician B**	**Physician A**
01	LB, DB	LB, DB	LB, DB
02	DB	DB	DB
03	LB, DB	Wh, LB, DB	LB, DB
04	LB	LB, DB	LB, DB
05	Wh, LB, DB	Wh, LB, DB	LB
06	LB, DB	LB, DB	LB, DB
07	DB, Bk	LB, DB	DB
08	LB	LB, DB	LB
09	LB, DB	LB, DB	LB, DB
10	LB	LB, DB	LB
11	LB, DB	Wh, LB, DB	LB, DB
12	DB, BG	Wh, LB, DB	DB, BG
13	LB, Bk	Wh, LB, DB	LB, Bk
14	Re, LB	LB, DB	LB
15	DB, Bk	LB, DB	DB
16	DB, Bk	Wh, LB, DB	DB
17	DB, BG	Wh, LB, DB	Wh, DB, BG
18	Bk	LB, DB	Bk
19	Wh, BG, Bk	Wh, LB, DB	DB, BG
20	DB, BG, Bk	Wh, LB, DB	Wh, BG, Bk
21	DB, Bk	Wh, DB	DB
22	Wh, Re, DB, BG	Wh, LB, DB	LB, DB
23	Wh, LB	Wh, LB, DB	LB
24	Wh, LB	Wh, LB, DB	LB
25	Re, LB, DB	LB, DB	LB
26	DB, BG, Bk	Wh, DB, BG	DB, BG
27	Re, LB	LB, DB	Re, LB
28	DB, BG, Bk	Wh, LB, DB, Bk	LB, DB, BG

Wh = white, Re = Red, LB = light brown, DB = dark brown, BG = blue gray, Bk = back



(a)

(b) (c) (d)

Fig. 4.2: Identification of color regions in PSL images of the training data. (a) original dermoscopic image (IMD17); (b) Medical identification of Wh, DB and BG color classes; Logical mask for: (c) Wh, (d) DB, and (e) BG color classes, respectively.

color classes, eye fatigue, etc. But that is the reality and a CADx system have to be strongly accurate to pass through those "human limitations".

Despite the differences in medical opinions the evaluations made by the physician A were considered for the identification of color regions task (Figure 4.2). In Table 4.2 the existing colors are noted. Based on that, the dataset can be divided into three cases: images with just one, two or three colors. In Table 4.2 the number of image color samples for each class in a total of 28 images that constitutes the dataset are summarized.

Unfortunately, our dataset does not contain any image of patients with black skin. However, according to [21], dermoscopy in nonwhite populations remains the same as in white populations because darker skin does not hinder the identification of unique dermoscopic features.

Tab. 4.3: Number of image color samples for each color class in the 28 dataset images.

Color class	No. of sample images
Wh	2
Re	1
LB	17
DB	17
BG	6
Bk	3
Total	28

Wh = white, Re = red, LB = light brown, DB = dark brown, BG = blue gray, Bk = black

4.3 Color Regions

For the training set it the best representation of dermoscopic color population respecting the medical evaluation was selected. The *best representation* involve not choosing air bubbles or hair in the image but only pixels that contain a "pure color lesion". Although the selection was very carefully made, the resulting samples may contain noise. This is an inherent problem to all applications, and should be taken into account.

4.4 Calibration

The technique of PSLs image acquisition, known as epiluminescence microscopy or dermoscopy (Section 2.3), have the disadvantages associated with the use of a dermatoscope (Figure 2.2). As such, the images of the dataset have been subjected to processing calibration. Visually, all images of the dataset (Appendix A) have different illumination conditions. There are many image calibration techniques that can be performed.

In order to reduce the effect of variability in illumination and background skin brightness, the following sections present image calibration techniques that will be applied to the dermoscopic dataset (Appendix A).

4.4.1 Skin Intensity Calibration

Image intensity calibration (**method I**) calibrates the intensity values of images based on a reference. The main goal of this technique is to put all images at the

same intensity level, so that there is consistency in the range of values. Therefore, it is necessary to have a *reference image* to calibrate all other images to the that intensity level. The assumption is that the color of the healthy skin is the same for all patients.

For application of intensity calibration in the dataset, it has to be assumed that the skin color is the same for all patients.

Consider an image, I_{ref}, in the RGB feature space. From this image a sample of healthy skin, M_{skin}, is extracted, for all 3 color channels (Red, Green, Blue), so M_{skin} is a sample vector. The intensity of an RGB image is defined as [32]

$$I = \frac{R + G + B}{3}, \tag{4.2}$$

so, the mean intensity is given for

$$< I >= \frac{< R > + < G > + < B >}{3} \tag{4.3}$$

where $< \bullet >$ represents the mean of corresponding color channel.

The objective is to calibrate all other images having as reference $< I_{ref} >$, i.e., we want to find [49]

$$\alpha :< I_{ref} >= \alpha < I_{obs} > \tag{4.4}$$

with $< I_{obs} >$ representing the mean of a tridimensional vector with the mean of each of the 3 color channels and $\alpha = [\alpha_{red}, \alpha_{green}, \alpha_{blue}]$, so that the calibrated images are those that satisfy:

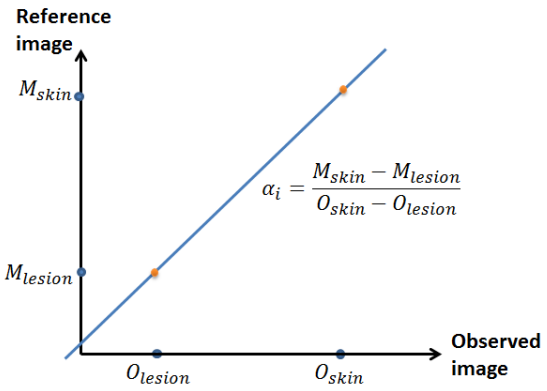
$$I'_{obs} = [\alpha_{red} I_{obs,red}, \alpha_{green} I_{obs,green}, \alpha_{blue} I_{obs,blue}] \tag{4.5}$$

The resulting images, I'_{obs}, correspond to a result of an internal operation and do not serve as a visual aid for the physician.

4.4.2 Calibration based on Skin-Lesion Rate

The **method II** for calibration also performs image calibration but, in this case, maintaining the skin-lesion rate constant for all images. For that, a reference image, I_{ref}, is used.The assumption is that the correlation between the color of healthy skin and the color of the lesion is the same for all images.

Consider the same case, with reference image, I_{ref}, in the RGB feature space.



(a)

Fig. 4.3: Procedure of the skin-lesion rate calibration.

Both the mean vector of healthy skin, M_{skin}, and mean vector of the lesion, M_{lesion}, are extracted, for all 3 color channels. For each image the mean vector for the healthy skin, O_{ref}, and other mean vector for the lesion, O_{lesion}, of all 3 color channels are also stored. The calibrated images are such that [49]:

$$\begin{cases} I'_{red} = \alpha_{i,1} R \\ I'_{green} = \alpha_{i,2} G \\ I'_{blue} = \alpha_{i,3} B \end{cases} \tag{4.6}$$

where the vector $\alpha_i = \frac{M_{skin} - M_{lesion}}{O_{skin} - O_{lesion}}$ for $i = 1, 2, 3, ...28$ images.

Figure 4.3 shown the graphical procedure of the skin-lesion rate calibration.

4.4.3 Calibration with the Magnitude of the Reference Mean Vector Constant

The last calibration technique used (**method III**), is a calibration procedure with the magnitude of the reference mean vector constant. The assumption is that the magnitude of the reference mean vector is constant.

Consider the same notations as above. From the I_{ref} image the M_{skin} and M_{lesion}, for all the feature space are extracted. For each image in the resulting set, also extracts O_{skin} and O_{lesion}.

The mean of calibrated images are determined by the intersection of:

- Circumference centred in $(0, 0)$ with radius determined by the magnitude of

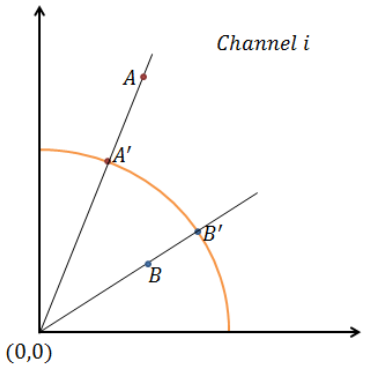


Fig. 4.4: Representative scheme of the technique of calibration with magnitude of reference constant.

$\sqrt{O^2_{skin,j} + O^2_{lesion,j}}$, where $j = 1, 2, 3$ corresponds to a color channel, and

- Line through the origin and the mean point observed.

In Figure 4.4 the representative scheme of the new mean vector found is shown.

4.5 Color Analysis

The signature of each color region extracted and stored can be represented by **color models**. As default, the dermoscopic color images usually deliver full color images, in RGB *color space* (or color model). Most color models can be obtained from the RGB color space.

The RGB Color Space

An RGB *color image* is an $M \times N \times 3$ array of color pixels, where each color pixel is a triplet corresponding to the red, green, and blue components of a RGB image at a specific spatial location (*spatial domain*) - Figure 4.5. An RGB image can be viewed as a "stack" of three gray-scale images that, when fed into the red, green, and blue inputs of a color monitor, produce a color image on the screen. By convention, the three images forming a RGB color image are referred to as the red, green, and blue *component images* (or *bands* or *channels*). The data class of the component images determines their range of values [32]. Since the acquisition of dermoscopic images delivers 8 bit per component image, the range of values is $[0, 255]$. The number of bits used to represent the pixel values of the component images determines the *bit depth* of a RGB image. In the dermoscopy case, each component image is an 8-bit

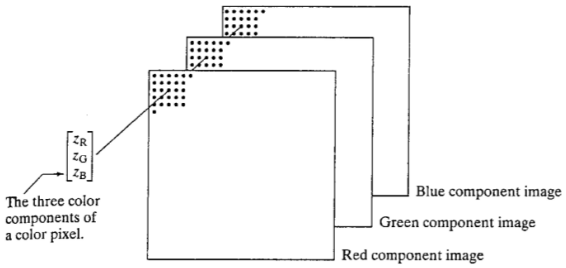
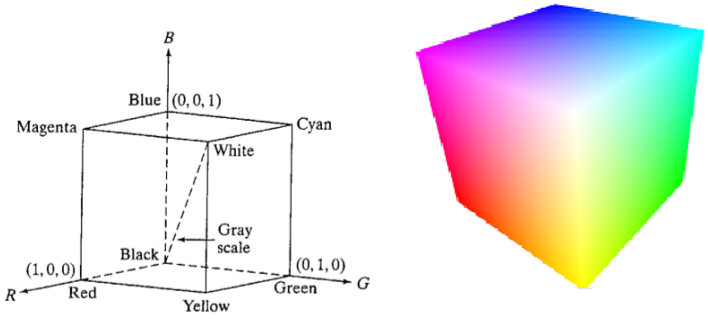


Fig. 4.5: Schematic showing how pixels of a RGB color image are formed from the corresponding pixels of the three component images [32].



(a) (b)

Fig. 4.6: (a) Schematic of the RGB color cube showing and secondary colors of light at the vertices. Points along the main diagonal have gray values from black at the origin to white at point $(1, 1, 1)$. (b) The RGB color cube [32].

image and the corresponding RGB images is said to be 24 bits deep (Section 2.3.2). Generally, the number of bits in all component images is the same. In this case, the number of possible colors in an RGB image is $(2^b)^3$, where b is the number of bits in each component image (case $b = 3$: $16,777,216$ colors).

The RGB *color space* usually is shown graphically as a RGB color cube, as depicted in Figure 4.6. The vertices of the cube are the *primary* (red, green, and blue) and secondary (cyan, magenta, and yellow) colors.

The HSI Color Space

Another representation for the signature of the dermoscopic color regions is the HSI (hue, saturation, intensity) color space. This color space *describes* colors in terms

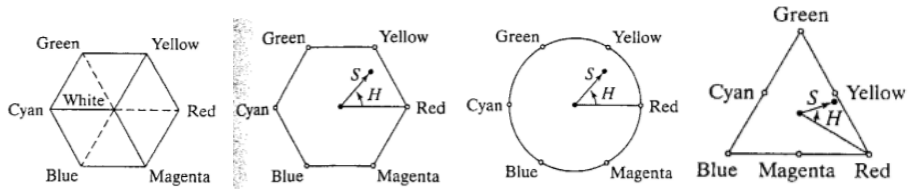


Fig. 4.7: Hue and saturation in the HSI color model. The dot is an arbitrary color point. The angle from the red axis gives the hue, and the length of the vectors is the saturation. The intensity of all colors in any of these planes is given by the position of the plane on the vertical intensity axis [32].

that are practical for human interpretations [32].

When humans view a colored object, we tend to describe it by its hue, saturation, and brightness. *Hue* is an attribute that describes a pure color (e.g., pure yellow, orange, or red), whereas *saturation* gives a measure of the degree to which a pure color is diluted by white light. *Brightness* is a subjective description of *intensity* and is a key factor in describing color sensation [32]. The HSI *color space* decouples the intensity component from the color-carrying information (hue and saturation) in a color image.

Figure 4.7(b) shows the hexagonal shape and an arbitrary color point (shown as dot). The hue of the point is determined by an angle from some reference point. Usually (but not always) an angle of 0° from the red axis designates 0 hue, and the hue increases in counterclockwise manner from there. The saturation (distance from the vertical axis) is the length of the vector from the origin to the point. Note that the origin is defined by the intersection of the color plane with the vertical intensity axis. The important components of the HSI color space are the vertical intensity axis, the length of the vector to a color point, and the angle this vector makes with the red axis. Therefore, it is usual to see the HS plane defined in terms of the hexagon just discussed, a triangle, or even a circle, as Figure 4.7(c) and (d) show. The shape chosen is not important because any one of these shapes can be warped into one of the other two by a geometric transformation [32].

The XYI Colors Space

A simple transformation that can be applied to the HSI color space is one that converts the hue and saturation components to cartesian coordinates, such that:

$$\begin{cases} X = S\cos H \\ Y = S\sin H \end{cases} \tag{4.7}$$

with $X, Y \in [-1, 1]$ and I remains the same.

The Opponent Color Space

The Opponent Color Space (OPP) is motivated by retinal processing [13]. Contemporary science of human vision states that color photo receptors at the retina (i.e., cones) are sensitive to long (L-cone), middle (M-cone) and short (S-cone) wavelengths. A single cone is color blind since its activation depends on both the wavelengths and intensity of the stimulus. A comparison of the signals from different classes of photo receptors is therefore the most basic computational requirement of a color vision system. The existence of cone-opponent retinal ganglion cells that perform such comparisons is well established for human vision [5].

In particular, opponent process theory postulates that yellow-blue and red-green information is represented by two parallel channels in the visual system that combine cone signals differently. It is now accepted that at an early stage in the red-green opponent pathway, signals from L and M cones are opposed, and in the yellow-blue pathway signals from S cones oppose a combined signal from L and M cones. In addiction, there is a third luminance or achromatic mechanisms in which retinal ganglion cells receive L- and M-cone input. Thus, L, M and S belong to a first layer of it, forming the basis of chromatic input to the primary cortex. Note also that this mechanism is not random since human color vision evolved for increasing the probability of subsistence [5].

Seeing the RGB space, the *components* (O_1, O_2, O_3) of Opponent color space are obtained using the equations [5]:

$$\begin{cases} red - green: & O_1 = (R - G)/\sqrt{2} \\ yellow - blue: & O_2 = ((R + G) - 2B)/\sqrt{6} \\ luminance: & O_3 = (R + G + B)/\sqrt{3} \end{cases} \tag{4.8}$$

for R, G and B running on values in $[0, 1]$.

4.6 Separability Measures

According to the proposed method (Figure 4.1), in the feature selection stage, the storing of the color set is also performed. For this stage it is convenient that each color class is *separable* from each other. In that case, the classifier can identify each cluster and label it correctly.

The color analysis section (4.5) introduces some of the features that can be the input data of the dermoscopic classifier. The feature selection is performed selecting the *dimensionality of the feature space* with appropriate and compact subsets of vital features to the accuracy and efficiency of subsequent classification step.

We have 6 color classes with N features to represent them. These features can be: red (R), green (G), and blue (B) from the RGB color space. But, for illustration, consider two dimensional multispectral space with two spectral classes as depicted in Figure 4.8. Suppose we wish to see whether the classes could be separated using only one feature - either x_1 or x_2. Of course it is not known which feature offers the best prospects *a priori*. This is what it has to be determined by a measure of separability. Consider an assessment of x_1. The spectral classes in the x_1 "subset" or subspace are shown in Figure 4.8 whereupon some overlap of the single dimensional distributions is indicated. If the distributions are well separated in the x_1 dimension then clearly the overlap will be small and it would be unlikely that a classifier would make an error in discriminating between them. The usefulness of x_1 feature subset therefore can be assessed in terms of the overlap of the distributions of that domain, or more generally, in terms of the *similarity of the distributions* in that domain as a function of x_1 alone [66].

Consider now an attempt to quantify the separation between a pair of probability distributions (as models of spectral classes) as an indication of the degree of overlap. Clearly distance between means is insufficient since overlap will also be influenced by the standard deviations of the distributions. Instead, a combination of both the distance between means and a measure of standard deviation is required. Moreover this must be vector-based measure in order to be applicable to multidimensional subspaces. Several such measures are available and are commonly encountered in connection with dermoscopy data [66].

4.6.1 Divergence

In order to obtain a quantitative measure of how separable two classes are, a distance measure is required. With the assumption of underlying distributions, a probabilistic

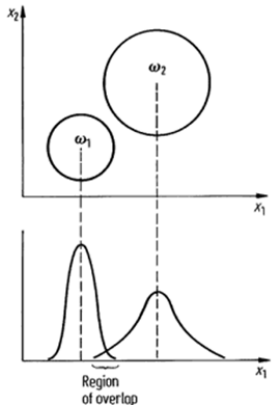


Fig. 4.8: Two dimensional multispectral space showing a hypothetical degree of separation possible in a single dimension subspace (in which class densities are shown) [41].

distance can be easily extracted from some parameters of the data.

Divergence is a measure of the separability of a pair of probability distributions that has its basis in their degree of overlap. It is defined in terms of the likelihood ratio [66]:

$$L_{ij}(x) = \frac{p(x|w_i)}{p(x|w_j)} \tag{4.9}$$

Where $p(x|w_i)$ and $p(x|w_j)$ are the values of the i^{th} and j^{th} spectral class probability distributions at the position x. These are shown in the overlapping region in Figure 4.9 whereupon it is evident that $L_{ij}(x)$ is a measure of "instantaneous" overlap. Clearly for very well separable spectral classes $L_{ij}(x) = 0$ or ∞ for all x.

It is useful to choose the logarithm of the likelihood ratio, in other words [66]:

$$L'_{ij}(x) = \ln\left(p(x|w_i)\right) - \ln\left(p(x|w_j)\right) \tag{4.10}$$

by means of which the *divergence* of the pair of class distribution is defined as

$$d_{ij}(x) = E\left\{L'_{ij}(x)|w_i\right\} + E\left\{L'_{ij}(x)|w_j\right\} \tag{4.11}$$

where $E\{\}$ is the expectation operator defined for continuous distributions as

$$E\left\{L'_{ij}(x)|w_i\right\} = \int_x L'_{ij}(x)p(x|w_i)dx \tag{4.12}$$

This is the average or expected value of the likelihood ratio with respect to all patterns in the i^{th} spectral class.

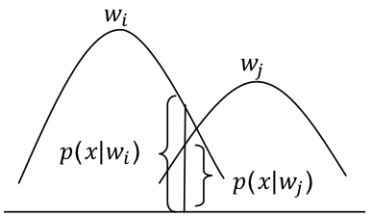


Fig. 4.9: *Divergence*: definition of the probabilities used in the likelihood ratio [66].

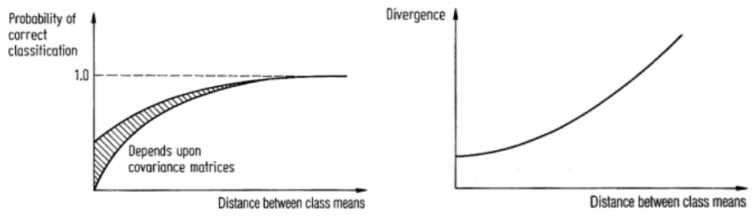
Since spectral classes in dermoscopy data are modeled by multidimensional normal distributions it is of particular interest to have available the specific form of Equation 4.11 when $p(x|w_i)$ and $p(x|w_j)$ are normal distributions with means and covariances represented by m_i, $\sum_i$ and m_j, $\sum_j$ respectively. By substitution of the full expressions for the normal distributions, it can be shown that [66]:

$$d_{ij}(x) = \tfrac{1}{2} tr\left\{\left(\sum\nolimits_i - \sum\nolimits_j\right)\left(\sum{}^{-1}{}_i - \sum{}^{-1}{}_j\right)\right\} + tr\left\{\left(\sum{}^{-1}{}_i + \sum{}^{-1}{}_j\right)(m_i - m_j)(m_i - m_j)^t\right\} = Term_1 + Term_2 \tag{4.13}$$

where $tr\{\}$ is the trace of the subject matrix. Note that $Term_1$ involves only covariances whereas $Term_2$ is the square of a normalized (by covariance) distance between the means of the distributions.

Equation 4.13 gives the divergence between a pair of spectral classes that are normally distributed. In the present case, there are more than two spectral classes (6 spectral classes, corresponding to the color regions) and all pairwise divergences need to be checked to see whether a particular feature subset gives sufficiently separable data.

As spectral classes become further removed from each other in multidimensional space, the probability of being able to classify a pattern at a particular location moves asymptotically to 1.0 as depicted in Figure 4.10(a). If divergence is similarly plotted it will be seen from its definition that it increases quadratically with separation between spectral class means, as depicted in Figure 4.10(b). This behavior unfortunately is quite misleading if divergence is to be used as an indication of how successfully patterns in the corresponding spectral classes could be mutually discriminated or classified. It implies, for example, that at large separations, further small increases will lead to vastly better classification accuracy whereas in practice this is not the case, as observed by Figure 4.10(a).



(a) (b)

Fig. 4.10: (a) Probability of correct classification as a function of spectral class separation and, (b) divergence as a function of spectral class separation [66].

4.6.2 Bhattacharyya Distance

Kailath [41] compared the *Bhattacharyya distance* and the *divergence* (Kullback-Leibler), and observed that Battacharyya yields better results in some cases while in other cases they are equivalent. In its simplest formulation, the Bhattacharyya distance between two classes can be calculated from the variance and mean of each class in the following way [2]:

$$D_B(c_1, c_2) = \frac{1}{4} \ln \left\{ \frac{1}{4} \left(\frac{\sigma^2 c_1}{\sigma^2 c_2} + \frac{\sigma^2 c_2}{\sigma^2 c_1} + 2 \right) \right\} + \frac{1}{4} \left\{ \frac{(\mu_{c_1} - \mu_{c_2})^2}{\sigma^2 c_1 + \sigma^2 c_2} \right\} \tag{4.14}$$

where $D_B(c_1, c_2)$ is the Bhattacharyya distance between c_1^{th} and c_2^{th} classes, σ_{c_1} is the variance of the c_1^{th} class, μ_{c_1} is the mean of the c_1^{th} class, and c_1 , c_2 are two different training classes.

For the multidimensional distance, the variances are replaced by covariance matrices and the means become vectors [66]:

$$D_B(c_1, c_2) = \frac{1}{8} (\mu_{c_1} - \mu_{c_2})^T \left[\sum_{c_1} + \sum_{c_2} \right]^{-1} (\mu_{c_1} - \mu_{c_2}) + \frac{1}{2} \ln \left\{ \frac{\left| \left(\sum_{c_1} + \sum_{c_2} \right) / 2 \right|}{\left| \sum_{c_1} \right|^{1/2} \left| \sum_{c_2} \right|^{1/2}} \right\} \tag{4.15}$$

4.6.3 Jeffries-Matusita

The *Jeffries-Matusita* (JM) distance measure of separability is used to assess how well two classes may be separated [81]. Between a pair of probability distributions

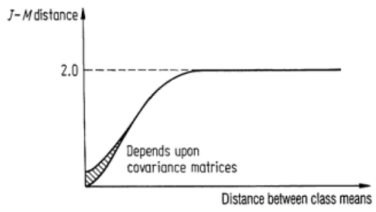


Fig. 4.11: Jeffries-Matusita measure as a function of separation between spectral class means [66].

(spectral classes) the JM distance is defined as [66]:

$$JM_{ij}(x) = \int_x \left\{ \sqrt{p(x|w_i)} - \sqrt{p(x|w_j)} \right\}^2 dx \tag{4.16}$$

which is seen to be a measure of the average distance between the two class density functions. For normally distributed classes this becomes [66]

$$JM_{ij} = 2\left(1 - \exp^{-D_B}\right) \tag{4.17}$$

in which D_B is referred to as the Bhattacharyya distance in Equation 4.15. For multidimensional spectral classes, it is of interest to note that the first term in 4.15 is akin to the square of the normalized distance between the class means. The presence of the exponential factor in Equation 4.17 gives an exponentially decreasing weight to increasing separations between spectral classes. If plotted as a function of distance between class means it shows a saturating behavior not unlike that expected for the probability of correct classification, as seen in Figure 4.11 [66].

It is asymptotic to 2.0 so that a JM distance of 2.0, the signatures can be totally separable with 100% accuracy. Generally, for results equal or greater than 1.8 the classes can be considered as separated: bellow 1.8, poor separation; between 1.8 and 1.9, the separation is fairly good; greater than 1.9 the separation is very good, we can say that the classes are separable [73].

JM distance performs better as a feature selection criterion for multivariate normal classes than divergence for the reasons given above. However, it is computationally more complex and thus expensive to use as can be seen by comparing Equations 4.13 and 4.15. Suppose a particular problem involves M spectral classes. Then consider the cost of computing all pairwise divergences and all pairwise JM distances. This computation involves matrix inverses and determinants. In the case of divergence it is necessary to compute an M number of inverse matrices. However

for the JM distance it is necessary to compute $\begin{pmatrix} M \\ 2 \end{pmatrix} + M$ equivalent matrix inverses since the individual covariances appear as pairs which have to be added and then inverted. It may be noted that $\begin{pmatrix} M \\ 2 \end{pmatrix} = \frac{1}{2}M(M+1)$ so that divergence is a factor of $\frac{1}{2}(M+1)$ more economical to use [66].

4.6.4 Transformed Divergence

A useful modification of divergence becomes apparent by noting the algebraic similarity of divergence to the parameter D_B in JM distance, as defined in Equation 4.15. Since both involve terms which are functions of the covariance alone, and terms which appear as normalised distances between class means, it should be possible to make use of an heuristic transformed divergence measure of the form (Swain and Davis, 1978) [66]:

$$d_{ij}^T = 2\left(1 - \exp^{-d_{ij}/8}\right) \tag{4.18}$$

Because of its exponential character it will have saturating behavior with increasing class separation, as does JM distance, and yet it is computationally more economical.

4.6.5 Kolmogorov-Smirnov Normality Test

For the computational implementation of 4.13 and 4.15, the data must be Gausian. As such, we can use the Kolmogorov-Smirnov (KS) normality test to determine whether or not the sample can be considered as originating from a normal distribution, whose parameters (mean and standard deviation) are obtained from the sample data. The hypothesis of KS test are [43]:

$$H_0 : Z \sim N(0,1) \tag{4.19}$$

$$H_1 : Z \nsim N(0,1) \tag{4.20}$$

where $Z = \frac{x_i - \mu_i}{s_i}$ with $i = 1,2,3,4,5,6$ corresponds to the color class number, μ_i and s_i are the sample mean and sample standard deviation, respectively. We want to know if a Gaussian distribution can be used to predict the behavior of the observed data. As a result of the KS test the *p-value* is obtained, which corresponds to the probability of the data from a given color class to have a normal distribution.

4.7 Inter- and Intra-Class Tests and Grouping Data

An a priori study can be performed in order to expect a particular outcome of classification step. This study involves inter- and intra-class tests. With these tests we can assess the variability and compactness for inter- and intra-class analysis, respectively. Some qualitatively measures, as boxplot graphs, will be exposed in order to explain some behavior of color classes.

The input data of the classifier is contained in the *feature space* where samples are collected, after assessing the separability of the clusters to be evaluated with the JM and TD metrics. However, the sample data of each color cluster can be assessed in different ways, as a *grouping data* attending to the instance of the color class per image or grouping color classes according behavioral/spatial similarities.

Single Image Training

For this test, a single image is used to train each color class [73]. The color of healthy skin and the depth of the lesion are some factors that extend the range of values of a given class (color).

Multiple Image Training

On this test, the training areas for each color class were identified in several images [73]. This increases the dispersion of the signatures in the *feature space* assigned for each class, thus increasing the overlap between the various color classes. In order to evaluate the behavior of each color class, a comparison of the feature space assigned to each color class from the different images used for training was performed.

Data Groups with Similar Behaviors

This test of grouping data with similar behavior is intuitive: if two classes are confused in the feature space (large rate of overlapping, the classes are correlated, etc.), they are grouped in a new class. If a third class maintains a good separability metric (JM and TD), the class is not grouped with any other class of existing group.

4.8 Non-Linear Technique for Increasing Clusters Separability

In general, classification algorithms work well when the classes are well separated in the feature space, which requires that the underlying data distributions have adequately large inter-cluster distances between patterns of different classes, and small intra-cluster distance between patterns of the same class. For most real-world problems, as the dermoscopic color identification, this is rarely the case, as overlapping class patterns in the feature space are usually the norm rather then the exception.

After the feature selection step (Figure 4.1), in which the classifier identifies the dermoscopic colors, there is a problem that needs to be addressed: the patterns generated by dermoscopy images overlap considerably in the feature space with small inter-cluster and large intra-cluster distances.

Poor separability problems are common among pattern recognition applications. However, the Nonlinear Cluster Transformation (NCT), introduced by Polikar in 1999 ([59], [60]), can be used for increasing pattern separability. Nonlinear cluster transformation is a three-step supervised procedure that attempts to increase the inter-cluster distances and reduce the intra-cluster distances, while preserving the dimensionality of the pattern vectors. NCT has no limitations in terms of dimensionality, number of classes, or the total number in the database [60].

In [60], Polikar et al. suggest firstly to eliminate the outliers in order to reduce the intra-cluster distances. The idea of NCT is to translate the clusters appropriately in order to physically separate them. Conceptually, all clusters are thought of as like charged particles: the magnitude and direction of the translation vector are then derived using the concept of a repulsive force by each cluster i on other clusters.

Consider a two-class problem with (possibly) overlapping clusters, whose centers are located at m_1 and m_2. The distance between these two clusters can be increased if class I patterns are translated along the vector $S_1 = -(m_2 - m_1)$, and class II patterns are translated along $S_2 = -S_1 = (m_2 - m_1)$. This idea can be extended to multi-class problems of arbitrary dimensionality, where patterns of class C_i can be translated along S_i, where the optimal direction S_i can be computed as [59]

$$S_i = -\sum_{j \neq i}^{C} (m_j - m_i) \tag{4.21}$$

and where m_i and m_j are the cluster centers of class i and class j, respectively, and C is the number of classes. The resultant vector for class i is $S_i = -M_i$, where

$$M_i = \sum_{j \neq i} (m_j - m_i) \tag{4.22}$$

All patterns in class i are moved along the direction of $-M_i$, and the translated patterns can be obtained by

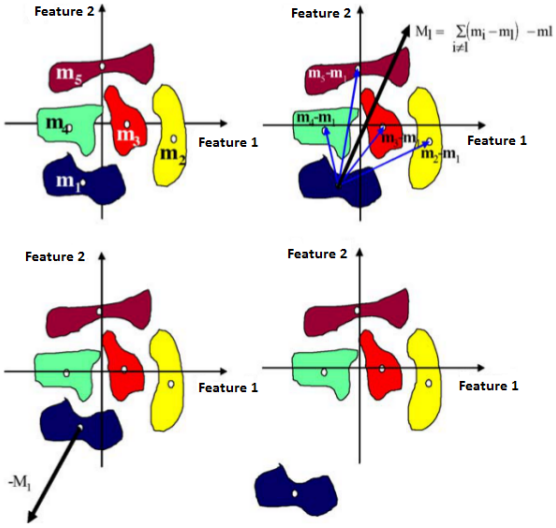
$$x_{S_i} = x_i + \left(-\frac{M_i}{\|M_i\|} \right) dist_i \tag{4.23}$$

where x_i is a pattern from the class i, $dist_i = 1/|m - m_i|$ is a normalizing constant that controls the amount of translations, x_{S_i} is the new location of the pattern x_i, m_i is the mean of all class i instances, and m is the mean of all instances from all classes. If the overlap in the data from different classes is high, $dist_i$ should be correspondingly high. For data with up to three dimensions of the feature space, this parameter can be determined simply by plotting the data and determining how much translations is necessary. It is straightforward to show mathematically that these directions of translation maximize inter-cluster distances [60]. Note that S_i points in the opposite direction of the resultant vector that combines the cluster center of class i to the centers of all other classes, that is, it points away from all clusters. The procedure is conceptually illustrated in Figure 4.12.

The cluster transformation described here can also be expressed in a matrix form. Let $i = 1, 2, ..., C$, where C is the number of classes ($C = 6$), $n = 1, 2, ..., N_i$ where N_i is the number of patterns in class i, x_n^i be the d-dimensional n^{th} pattern of the i^{th} class, y_n^i be the corresponding d-dimensional pattern after translation. Then:

$$\begin{bmatrix} {y_1}^i \\ {y_2}^i \\ . \\ . \\ . \\ {y_{N_i}}^i \end{bmatrix}_{N_i x d} = -dist_i \begin{bmatrix} 1 & . & . & . & 1 & 1 \\ 1 & & & & 1 & 1 \\ . & . & & & . & . \\ . & & . & & . & . \\ . & & & . & . & . \\ 1 & . & . & . & 1 & 1 \end{bmatrix}_{N_i x C} \begin{bmatrix} m_1 - m_i \\ m_2 - m_i \\ . \\ . \\ . \\ m_C - m_i \end{bmatrix}_{C x d} + \begin{bmatrix} {x_1}^i \\ {x_2}^i \\ . \\ . \\ . \\ {x_{N_i}}^i \end{bmatrix}_{N_i x d} \tag{4.24}$$

This equation can be implemented on the training data set a sufficient number of times so that it allows the separability between color classes (the separability is evaluated with JM and TD metrics). Each transformation generates a new dataset to



(a)

Fig. 4.12: Non-linear cluster transformation (adapted from [59]).

be used in the next transformation. When the classes are now sufficiently separate, the resulting dataset can be used to train a neural network to learn the overall transformation function.

4.9 Classification

The difficulty of standardizing the diagnostic criteria and the wide variability of color classes, as well as the discrimination of certain colors of a lesion, remains problematic. A system that allows analysis of tumours would be useful, especially for general practitioners who do not often observe different colors in the same lesion. Such system is introduced in Figure 4.13, which presents a general methodology based on the selection of six dermoscopic colors, non-linear transformations and cluster surfaces learning.

The previous steps allow a set of values (spectral signature of the color) to be calculated so that the tumour can be described by the color signature established by dermatologists. In order to classify the lesion/tumour as melanoma (or suspicious) or benign, several multilayer neural networks with a supervised learning algorithm are used and placed in competition.

In the schematic representation of Figure 4.13, each sample x, of the test set is subjected to the Polikar transformation until it stays at the space of the trained data.

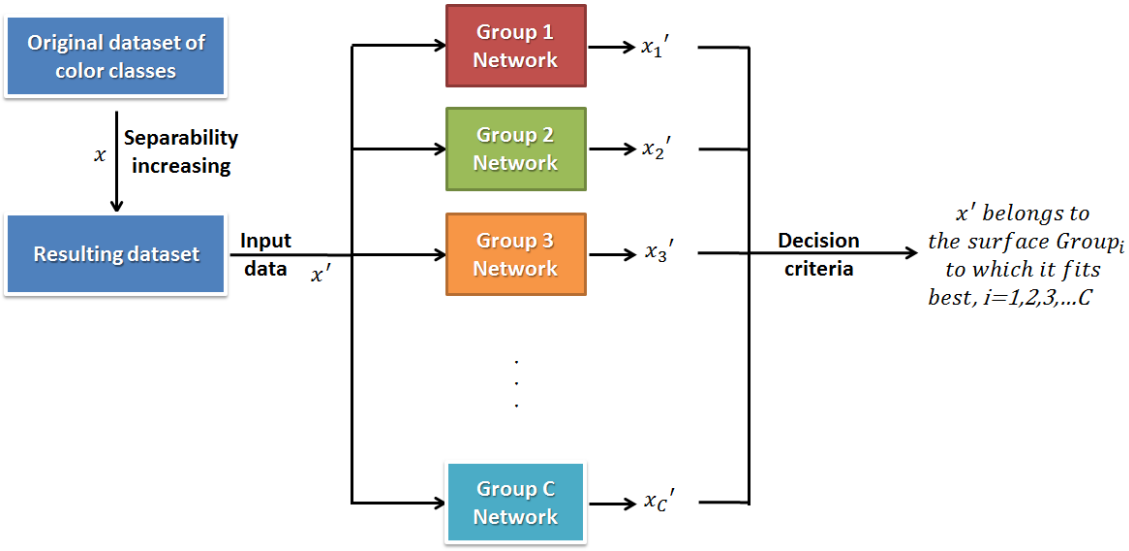


Fig. 4.13: Proposed approach for the classification task.

The point x meets the dataset parameters when applying the composed functions that correspond to the number of Polikar transformations, e.g., if $N = 3$ correspond to the number of transformations that allows the best separability between color classes. The network input data is the result of composition of functions f_1, f_2 and f_3, i.e., $x^{'} = f_3(f_2(f_1(x)))$ where f_i corresponds to the function transformation fixed with the training data. The resulting point, $x^{'}$, passes through each of the networks corresponding to each class and will be labeled as belonging to the class that best fits the point on the surface.

By analogy with the biological systems (Figure 4.14), the training of a neural network is intended to constrain the internal parameters so that the network develop their function in a desired way, and it finds that function as a result of repeated exposition to pattern examples. In general, the training of neural networks can be *supervised* or *unsupervised* (Section 2.5.3). In the case of the proposed classification, the training is supervised, and the data involve a description of a situation and the desired behavior of the network in each situation, directly influencing their parameters.

In the artificial neural networks there is neurons as input, as output and as internal non-linear processing units which transform the input data in the outcome. The neurons are organized in layers, with an unidirectional flow of information from the entry to the exit of the network. The connections between neurons are associated to weights $w_{kx,pz}$, so that the neuron signal emitted is multiplied by the weight of connection before the entering the next neuron. A neuron with index x from the

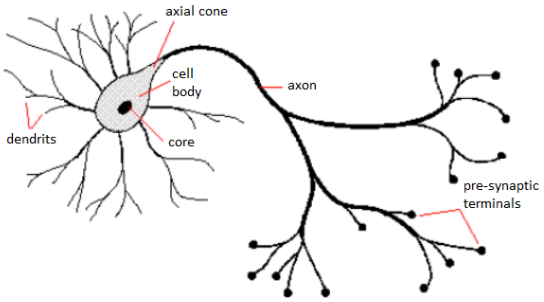


Fig. 4.14: Biological neuron scheme.

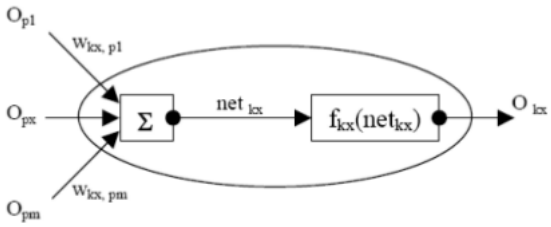


Fig. 4.15: Operational techniques in a generic neuron, designed as the x order in the layer k, receiving signals of the previous layer p, after being affected by the weights w.

layer k receives signals O_{pz} from neurons of the previous layer p and performs the transformations represented in Figure 4.15, before producing an output signal O_{kx} which, in turn, activates neurons in the following layer:

$$net_{kx} = \sum_{z=1}^{m} w_{kx,pz}.O_{pz} \tag{4.25}$$

where $O_{kx} = f_{kx}(net_{kx})$. The f function is called "activation function" of the neuron. The current functions used are linear, step or sigmoid [11].

If we want the network to learn a suitable mapping from a given dataset (color features in dermoscopic images, that correspond to a set of values $x^{'}$ in Figure 4.13), the learning will be based on the definition of a suitable error function, which is then minimized with respect to the weights in the network. The algorithm for evaluating the derivatives of the error function is known as *backpropagation* since it corresponds to a propagation of errors backwards through the network. This technique is based on the mathematical method of steepest descent, which is a local search method and so the neural networks training suffers the same problems of local gradients methods. These methods are more likely to become trapped in local optimums, the gradient always points in the "wrong direction" and does not allow the current direction

to evolve in another direction. Likewise, "valleys" in the objective function are elongated traps for the gradient algorithms, leading to stagnation of progress toward the great expenditure of effort with numerous iterations with reduced efficiency.

An heuristic method with good results involves the introduction of the momentum criterion, that corresponds to an attempt to establish the iterations progress in the founded productive direction, as if we were "jumping over" the "roughness" of the local objective function. The momentum term is introduced in the algorithm in the weights update step as follows:

$$\Delta w_t = -\eta \frac{\partial E}{\partial w} + w \Delta w_{t-1} \tag{4.26}$$

The weights of the neural networks are adjusted so that they minimize (or maximize) a cost function. All of cost functions are based on the error ($e_i = y_i - d_i$) made by the neural network in the training examples. In what follows will be used the Mean Square Error (MSE) cost function (for more information about it refer to [11]):

$$MSE = \frac{1}{N} \sum_{i=1}^{N} e_i^2 \tag{4.27}$$

The MSE function has some advantages: the error surface is softy and continuously derivable; it is a convex quadratic function, as so, it has just one local minima; it does not involve the use of a high computational effort when applied to the training of adaptive systems, such the present system.

After training a network for each cluster, we will be able to decide to what cluster a new point $x^{'}$ belongs. Applying the networks to the point $x^{'}$ and analysing the reproducibility of each one of them, the point $x^{'}$ will be given the label of the cluster to which it fits the best.

A frequent problem in the estimation of the minimum allowable error is to estimate a threshold of acceptance. In the case of competing networks, we do not determine the value of the error. Instead, we determine which of the networks generated the smallest error. This way, we are dismissing thresholds of acceptance and so reducing the arbitrariness.

Generally, if the training set is sufficiently representative of the data, the reproducibility of each class tends to the best approach of the classification task.

This classification method also preserves the local adaptation of the distance metric and the shape of the cluster.

1	3	3	1
1	3	3	1
3	3	3	3
3	3	3	3
1	1	1	1

(a)

3	3	3	3
3	3	3	3
3	3	3	3
3	3	3	3
3	3	3	3

(b)

Fig. 4.16: Illustration of filter applies in a pseudo-region of classified image. For a region as belong to class 3 and some isolated points classified inside the region as belonging to class 1: (a) original set of classified points and (b) mode filter result.

4.10 Post-Processing

According to the proposed flowchart shown in Figure 4.1, the following step refers to the post-processing, which involves corners removal and filtering.

Corners removal is a procedure in which the black corners in the classified image are removed in order to avoid classifying those areas of the image in the classification results. *Filtering* is applied in order to remove isolated points resulting from the classification. The goal is not to present a classified image, but to find regions of color classes. Sometimes, inside of a certain color class (consider the object class), the classification marks isolated points as belonging to other classes due to the noise in the image (Figure 2.7). Applying a mode filter in the classified image, the misunderstood points are eliminated.

In the present case, we want to remove some isolated points inside a class region (object). So, this region (illustration in Figure 4.16) has mostly points of a certain class ($class = 3$) with some isolated points misclassified ($class = 1$). The filter that results best for the elimination of the isolated points inside the region of class 3 is the mode filter, which "normalizes" the labels for the more dominant class (class 3).

In addition, some morphological operations can be carried out when a correctly class is confused with another class. For instance, suppose that Figure 4.16(a) represents a classified image. If we know, *a priori*, that class 1 is confused with class 3, i.e., points that belongs to class 1 are labelled as being from class 3, we can apply a morphological erosion operation in order to reduce the number of points labelled as class 3, increasing the number of points correctly labelled (class 1). If

we have changed the labels in Figure 4.16(a), the operation would be the morphological dilation in order to increase the number of points labelled as class 1. Both operations, dilation and erosion, use a structuring element for expanding or eroding (respectively) the shapes contained in the input image.

4.11 Validation

To validate the results, the test/validation set is used. Since the dataset is composed by 28 images, only two sets were considered: a training set and a test set. If the number of images available were higher, we could consider another set, the validation set, which is independent of the results in the training and test set.

The classifier learns the clusters surface with NNs for each cluster. The classification of the new points (points in the images of the test set) are validated with a *ground truth*. The ground truth is obtained by medical observation and annotation of the colors present in each image of the test set (Table 4.2). In this case, we can check if the classifier mimics the human classification, since it was trained by human annotations.

For the validation task, the usual measures of specificity (SP) and sensibility (SE) were used. These measures can indicate quantitative and qualitatively the performance of the proposed algorithm.

Chapter 5

Results and Results Discussion

This chapter presents the main results achieved using the proposed method for identification of dermoscopic colors in Pigmented Skin Lesion (PSLs). It is structured as follows: in Section 5.1 experimental data description and intra-classes tests in the first section; pre-processing and feature selection. In Section 5.2, which includes results of the calibration methods implementation and the selection of the features that characterize each cluster that is used to train the classifier. Results of the non-linear cluster transformation technique that allows the increase of color classes (clusters) separability in Section 5.4. Classification and post-processing results in Sections 5.5 and 5.6. Finally, in Section 5.7, the results validation that includes performance measures, such as sensibility (SE) and specificity (SP).

All sections include results by the Jeffries-Matusita and Transformed Divergence separability measures to quantify the separability of the clusters in the feature space. Only the most relevant figures and tables for the presentation of results is include in this chapter; other illustrations can be accessed in appendix to this thesis.

5.1 Experimental Data Description

The experimental data was already described in Section 4.2 and more particularly in Table 4.2 where the colors presence in each of the 28 dataset images were presented.

Pre-processing and feature selection will be proceed by primarily tests that allows the conduct of each color class. The RGB, the color space by default in dermoscopic image acquisition (Section 2.3.2) will be use to illustrate several data measures and parameters. For this primarily tests two groups were considered: the SIT and the MIT group.

As mentioned in Section 4.7, the SIT group use a single image to train each color class, and for the MIT group, the more realistic one, the training areas for each color class were identified in all 28 test images.

For the SIT group the images were selected randomly: IMD17 (for white), IMD27 (red), IMD05 (light brown), IMD01 (dark brown), IMD12 (blue gray) and IMD20 (black) - see Table 4.2.

5.1.1 Intra-Class Tests

When we move from the SIT group to the MIT group, by identifying the training areas for each color class using all 28 images, the number of images used to train each color class varies considerably, from only a few (1 for Re and 2 for Wh) up to 17 (for both LB and DB), as in Table 4.2. The large number of images used to train these classes (DB and LB) greatly increases the dispersion of their signatures in the RGB feature space, thus increasing the overlap between the various color classes.

In fact, the same color can have different spectral behavior when it comes from different images. This may be due to the differently skin tone or to a different color perception by the human validation. The same lesion could be interpreted differently if the background (skin) is different. As such, it is important to study the tonality variation of the skin along with the presence of each class and the spectral range variability of each of the six color classes. In order to do that, intra-class tests can be carried out.

In Figures 5.1, 5.2 and 5.3 boxplot graphs are presented for the class BG (Blue Gray) in the Red, Green and Blue color channels of the RGB color space, respectively. There are considerable differences in the three color channels (RGB). The same qualitative test was performed for the other color classes, which are presented in appendix B.

In Figures 5.4, 5.5 and 5.6 the variability of skin class are shown, illustrated with the same qualitative graphs - boxplots, which are very useful when we compare the spectral range between images. The corresponding plotted samples correspond to a small portion of healthy skin in each of 28 images for Red, Green and Blue channel, respectively.

The identification of some degree of correlation between color channels (in the RGB feature space) is noted in Figures 5.1-5.6 and, most obviously, the disparity in the range values in different images for the same class.

The signature range for the BG color class is approximately (Figures 5.1, 5.2

and 5.3): $[50, 200]$ for the Red channel, $[40, 150]$ for the Green channel and $[30, 160]$ for the Blue channel. For instance, consider the Red channel for images IMD20 and IMD28 (Figure 5.1): the image IMD28 has a BG spectral range of approximately $[100, 200]$, on the other hand, the image IMD20 has a BG spectral range of $[50, 200]$, approximately, which when plotted together perform a spectral range of $[50, 200]$. The BG color class have different signatures when coming from different images (e.g. IMD20 and IMD28), increasing the spectral range and the overlapping with other color classes.

For the skin, the small area of healthy skin collected in each of the 28 dermoscopic images perform a similar behavior as the BG color class, in the sense of the non-existence of a band in which all the skin samples falls. In the Red channel (Figrue 5.4), almost all the skin samples falls in the spectral range of $[200, 250]$, observing the proliferation of some skin samples outside this range. In the Green channel (Figure 5.5) the skin spectral range is more dense between $[150, 200]$ intensity values, but having a spectral range of, approximately, $[50, 250]$. In the Blue channel (Figure 5.6), the density in noted between $[100, 200]$ with intensity values outside this range.

Although the boxplot graphs represent qualitative measures it is a sufficient measure, since the variation is evident.

This problem of classes variability can be partially solved by performing image calibration, in order to normalize the range values inside the same class as we will see in the next section.

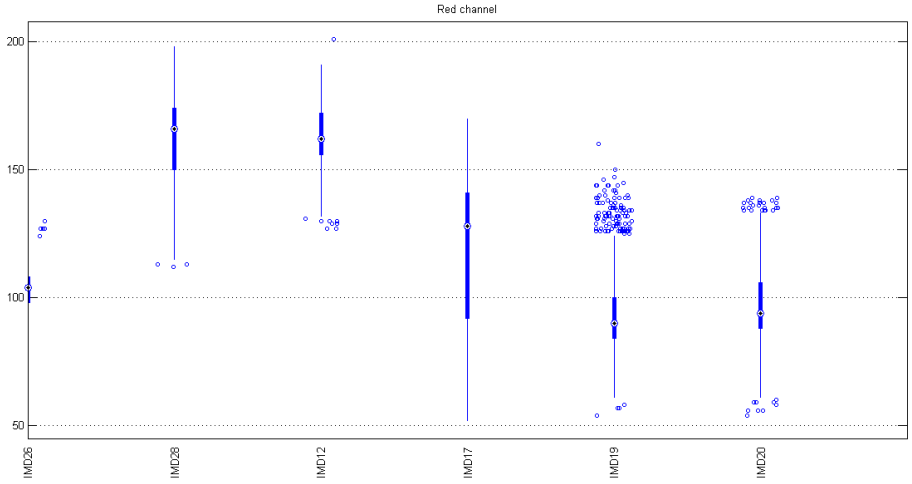


Fig. 5.1: Boxplot for subset of images that have the BG (Blue Gray) color class: IMD26, IMD28, IMD12, IMD17, IMD19 and IMD20, in the Red color channel of RGB feature space.

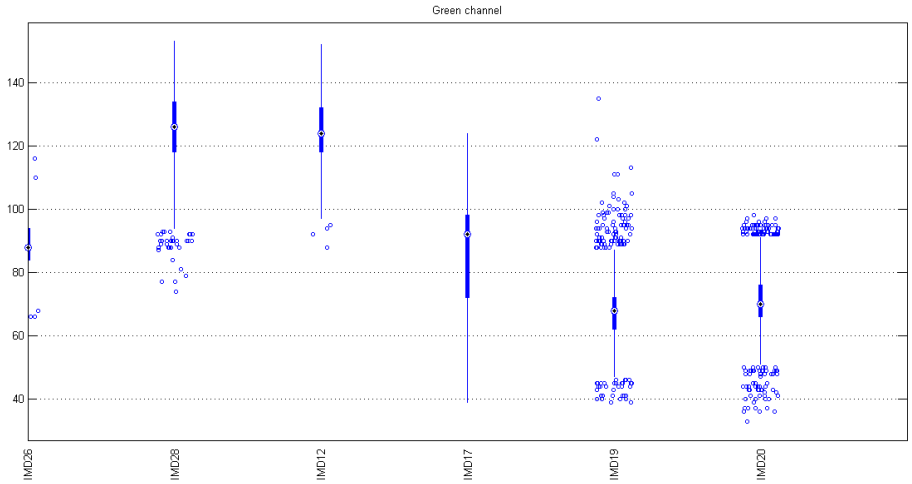


Fig. 5.2: Boxplot for subset of images that have the BG (Blue Gray) color class: IMD26, IMD28, IMD12, IMD17, IMD19 and IMD20, in the Green color channel of RGB feature space.

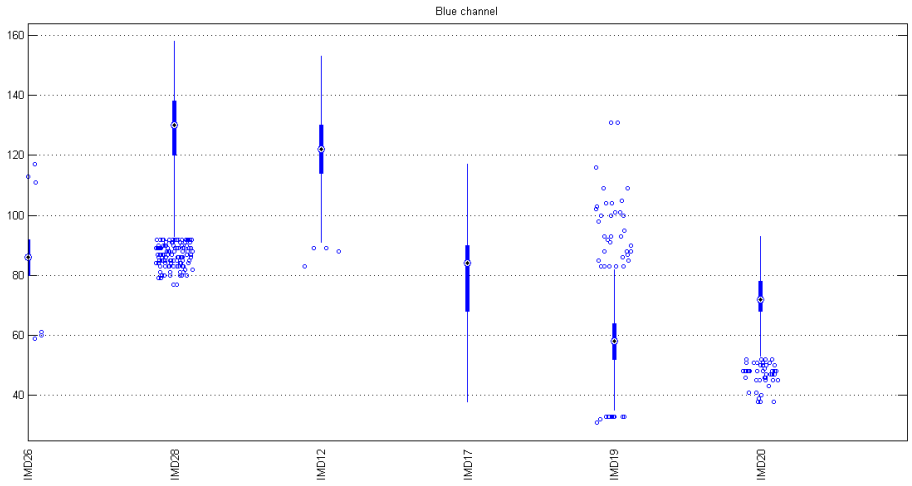


Fig. 5.3: Boxplot for subset of images that have the BG (Blue Gray) color class: IMD26, IMD28, IMD12, IMD17, IMD19 and IMD20, in the Blue color channel of RGB feature space.

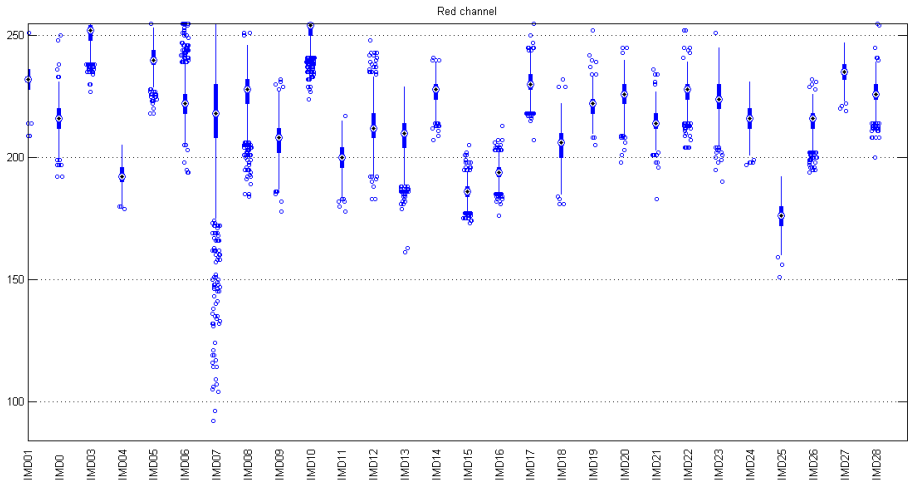


Fig. 5.4: Boxplot of all (28) test images for skin class - evaluation of the Red color channel.

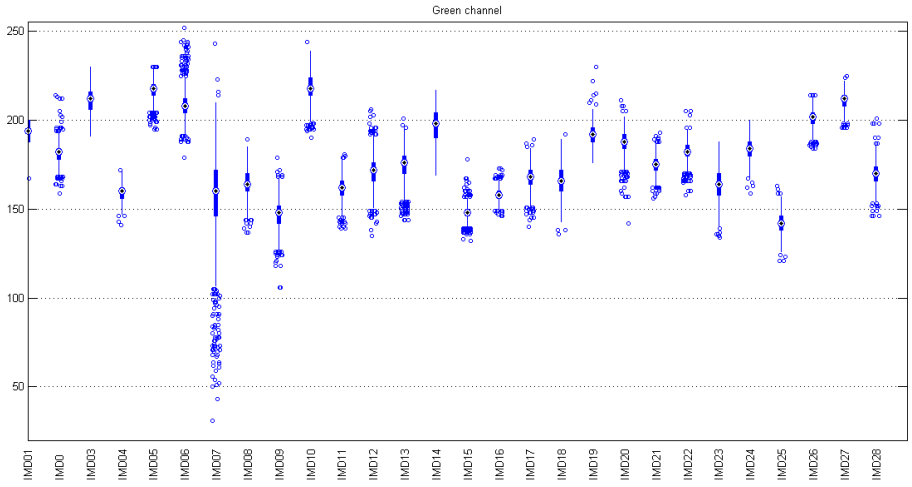


Fig. 5.5: Boxplot of all (28) test images for skin class - evaluation of the Green color channel.

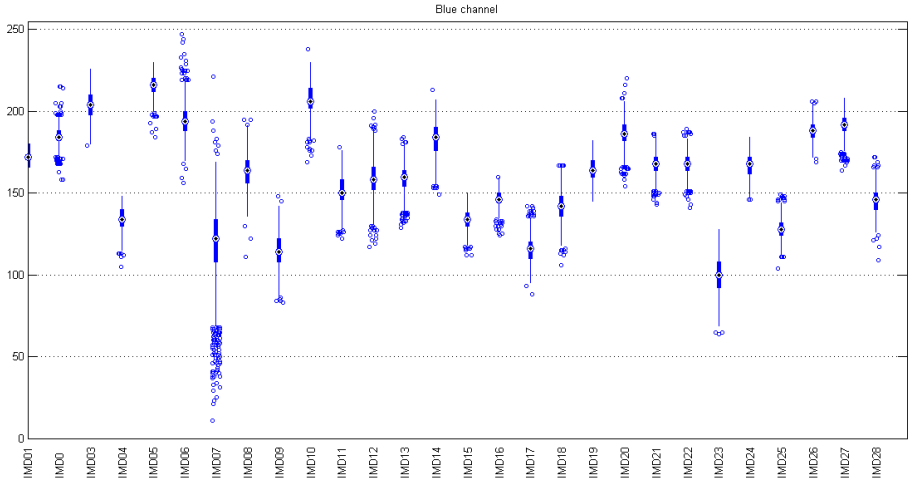


Fig. 5.6: Boxplot of all (28) test images for skin class - evaluation of the Blue color channel.

5.2 Pre-Processing and Feature Selection

The pre-processing step refers to the image calibration task applied in the proposed method (Figure 4.1). To perform image calibration three different methods (method I, II, II - Section 4.4) were applied. For all calibration techniques the same reference sample was considered, consisting by the same one image, IMD25 (Appendix A). From now on, this image will be considered as reference or reference sample or I_{ref}.

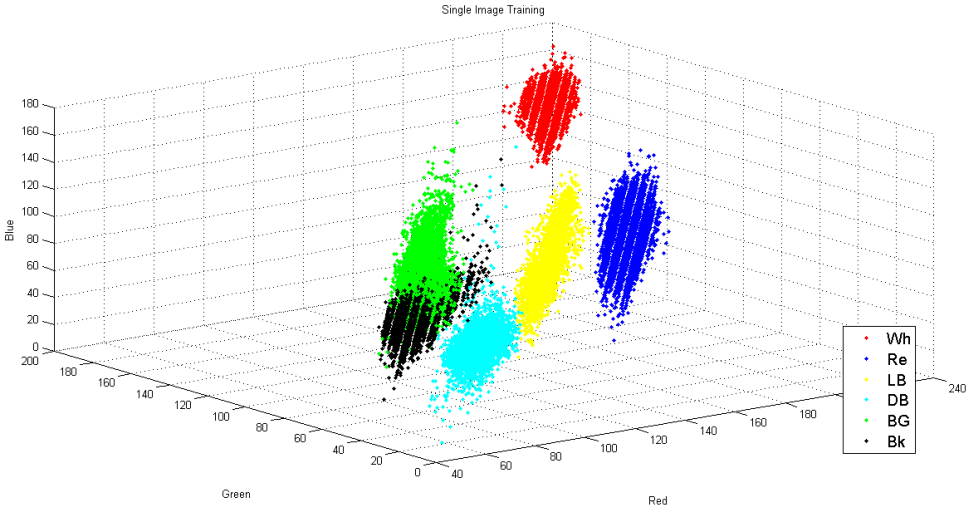
The calibration method that collects the best results is the "skin intensity calibration", or method I (Subsection 4.4.1). From now on, the calibration technique refers to this technique, which can be also stated as "method I".

Feature selection step refers to the selection of the significant features that describe each dermoscopic color class. In this case, the RGB (Red, Green and Blue channels) chromatic colors to represent each cluster were chosen. Comprising an N-dimensional feature space, $N = 3$. Other color spaces were used to explore the separability increasing, such as HSI, XYI, and OPP color spaces (see Section 4.5). The HSI is indicated as the most used color space and the one that achieves best results. However, in this case of dermoscopic image classification approach, the HSI color space does not improve results relative to the RGB color space. Even when both color spaces are used simultaneously ($N = 6$), the separability of clusters do not improve. So, in order to reduce the computational effort only the original RGB color space was considered as the feature space to characterize each color cluster. The results of separability distances between classes for other color models can be seen in Appendix C.

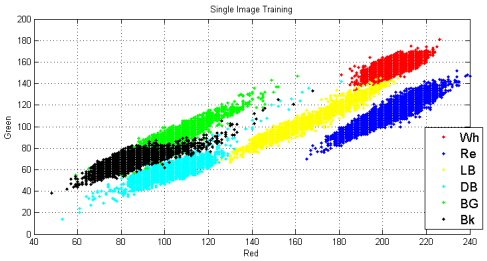
Once that the whole training samples and the vital features which describe each cluster are collected, as well as, the best calibration method, the separability measures and data behavior are presented using the SIT and MIT groups.

5.2.1 Single Image Training

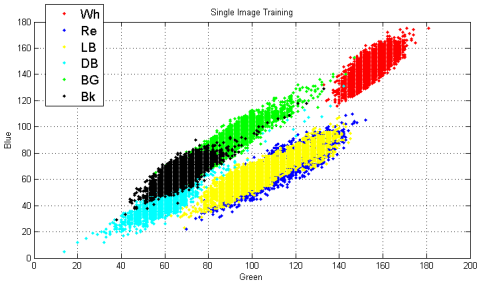
The 3D scatterplot and the projection in the three 2D planes (Red/Green, Green/Blue, Red/Blue) for SIT group are presented in Figure 5.7. As expected, the feature space is scarcely used and there is a certain amount of overlap between some classes.



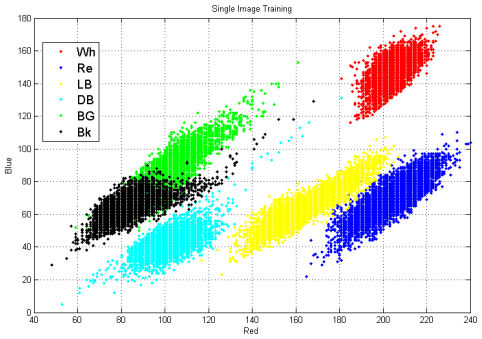
(a)



(b)



(c)



(d)

Fig. 5.7: Scatterplots for the single image training test without calibration: (a) on the RGB feature space, (b) a projection on the 2D Red, Green plane, (c) 2D Green, Blue plane and (d) 2D Red, Blue plane.

Tab. 5.1: Original data behavior for single image training and multiple image training groups without calibration, presenting minimum (Min), maximum (Max), mean (μ), standard deviation (σ) and p-value with 5% of significance for each class.

Single Image Training															
Classes	**Min**	**Max**	μ	σ	**p-value**	**Min**	**Max**	μ	σ	**p-value**	**Min**	**Max**	μ	σ	**p-value**
	Red channel					**Green channel**					**Blue channel**				
Wh	181	226	205	6.1	0.99	133	181	156	5.8	0.99	116	171	148	8.5	0.96
Re	165	240	205	10.4	0.95	70	152	114	11.1	0.94	22	110	70.4	10.7	0.93
LB	117	207	164	15.3	0.91	63	145	104	14.4	0.90	23	107	64.5	12.5	0.94
DB	53	181	102	9.4	0.96	14	142	59.2	7.5	0.98	5	131	42.9	7.9	0.95
BG	59	161	101	9.5	0.95	51	147	86.0	10.6	0.95	44	153	84.4	12.1	0.92
Bk	48	168	83	8.4	0.96	38	133	68.1	6.0	0.98	29	129	64.8	6.4	0.99
Multiple Image Training															
	Red channel					**Green channel**					**Blue channel**				
Wh	121	241	181	20.6	0.84	84	191	136	16.1	0.87	77	182	127	17.3	0.83
Re	137	255	208	13.2	0.83	62	196	122	16.4	0.80	5	158	75.9	13.1	0.82
LB	62	255	172	28.1	0.53	0	255	101	28.4	0.48	0	223	60.8	32.8	0.46
DB	26	221	113	25.5	0.49	7	183	72	18.2	0.62	0	181	58.4	17.1	0.59
BG	52	201	123	33.2	0.64	27	159	93.6	24.9	0.76	22	158	90.2	26.8	0.71
Bk	31	224	94.8	15.5	0.70	11	194	74.7	12.4	0.80	10	192	71.5	13.8	0.81

Wh - White, Re - Red, LB - Light Brown, DB - Dark Brown, BG - Blue Gray, Bk - Black

Tab. 5.2: Jeffries-Matusita - JM (top right triangular matrix) and Transformed Divergence - TD (bottom left) distance values, for the single image training test.

Classes	Wh	Re	LB	DB	BG	Bk
Wh	-	2.00	2.00	2.00	2.00	2.00
Re	2.00	-	2.00	2.00	2.00	2.00
LB	2.00	2.00	-	1.98	2.00	2.00
DB	2.00	2.00	2.00	-	1.99	1.99
BG	2.00	2.00	2.00	2.00	-	1.02
Bk	2.00	2.00	2.00	1.99	1.01	-

Wh - white, Re - Red, LB - Light Brown, DB - Dark Brown, BG - Blue Gray, Bk - Black

The Kolmogorov-Smirnov (KS) test was used to determine if the samples can be considered as originating from a normal (Gaussian) distribution (Subsection 4.6.5). The KS test provides the mean (μ_i) and standard deviation (σ_i) parameters for each sample i, as well as the probability of the data to have a normal distribution (p-value). The color classes follow Gaussian distributions, as demonstrated by the p-values results for significance level $\alpha = 5\%$ presented in Table 5.1 (all above 0.90). The SIT corresponding matrix, in Table 5.1, also presents the minimum, maximum, average (μ) and standard deviation (σ) values for each color class. The minimum and maximum parameters allows the assessment of the spectral range of each color class.

Considering that the color classes have a normal distribution, it is thus possible to apply the simplified forms of the divergence (Equation 4.13) and Bhattacharyya distance (Equation 4.15) distances to evaluate the class separabilities. Table 5.2 presents the two triangular matrices (TD and JM) together, the JM distance as a top right triangular matrix and the TD as a bottom left triangular matrix. For most class pairs, both TD and JM have either a perfect separability (2.00), or very high values (above 1.9). The only major difficulty is to distinguish between color classes BG and Bk, which have low values of both TD and JM distances, 1.01 and 1.02, respectively, which it is expected by the observation of the spectral ranges in Table 5.1: the BG and Bk color classes have an identical spectral range for all 3 color channel (Red, Green and Blue). According to this test, considering the RGB color features and training data (SIT group) used, these two classes are not distinguishable, but all other class pairs are well separable.

Tab. 5.3: Jeffries-Matusita - JM (top right triangular matrix) and Transformed Divergence - TD (bottom left) distance values, for the multiple-image training test.

Classes	Wh	Re	LB	DB	BG	Bk
Wh	-	2.00	1.74	1.97	1.69	1.99
Re	2.00	-	1.48	1.97	2.00	2.00
LB	1.98	1.23	-	1.47	1.87	1.99
DB	1.95	2.00	1.21	-	1.03	1.22
BG	1.70	2.00	1.94	0.75	-	0.55
Bk	1.99	2.00	2.00	0.94	0.10	-

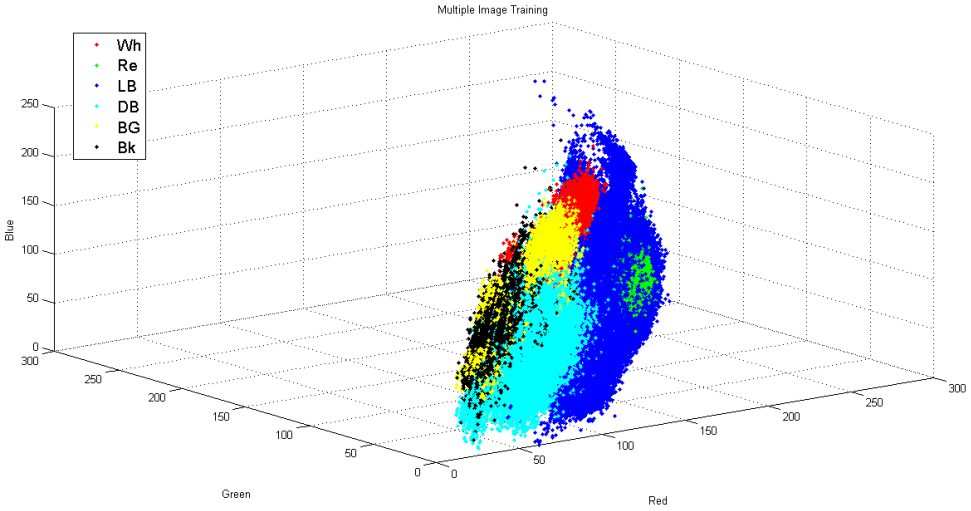
Wh - white, Re - Red, LB - Light Brown, DB - Dark Brown
BG - Blue Gray, Bk - Black

5.2.2 Multiple Image Training

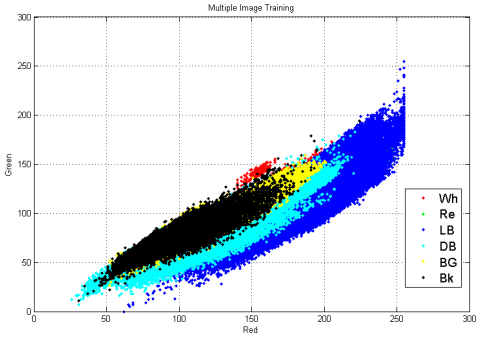
The MIT group is more challenging than SIT group, identifying the training areas for each color class using all the 28 images. Besides that, it is more realistic than the SIT group. Figure 5.8 show the 3D scatterplot and the projection in the three 2D planes (Red/Green, Green/Blue, Red/Blue) for MIT group. As expected, the amount of overlay information is immense.

Table 5.1 also present the minimum, maximum, average (μ) and standard deviation (σ) values for each color class for the MIT group. The p-values from the KS test are generally lower than those for the SIT group. For class Bk the p-values vary between 0.70 and 0.81 in the 3 color channels (R, G, B), while for classes Wh and Re the p-values are all above 0.80 with the maximum p-value of 0.87. For class BG the p-values vary between 0.64 and 0.76 for the three color channels. The worst classes are classes LB and DB, with p-values of 0.53 and 0.49, 0.48 and 0.62, 0.46 and 0.59 for the R, G and B color channels, respectively. The data was nevertheless still treated as having normal distribution, and the TD and JM distances were thus computed.

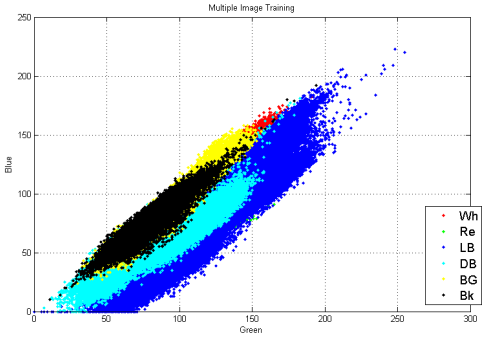
Table 5.3 presents the two triangular matrices for the MIT test (the JM distance at top right and the TD at bottom left). These results indicate that 8 out of 15 classes pairs are clearly distinguishable (for JM) or 9 out of 15 (based on TD values). There is a class pair (Wh-LB) with contradictory indications from JM and TD distances (1.74 and 1.98, respectively). The remaining 6 class pairs have both low JM and TD values (below 1.8) and cannot thus be considered separable. These class pairs are: Wh-LB, Wh-BG, Re-LB, LB-DB, DB-BG, DB-Bk, BG-Bk. The SIT and MIT tests are also presented in Silva et al.[73] (2012) with a sightly different



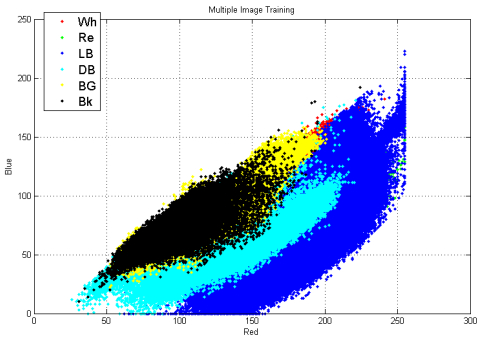
(a)



(b)



(c)



(d)

Fig. 5.8: Scatterplots for the multiple image training test without calibration: (a) on the RGB feature space, (b) a projection on the 2D Red, Green plane, (c) 2D Green, Blue plane and (d) 2D Red, Blue plane.

data samples.

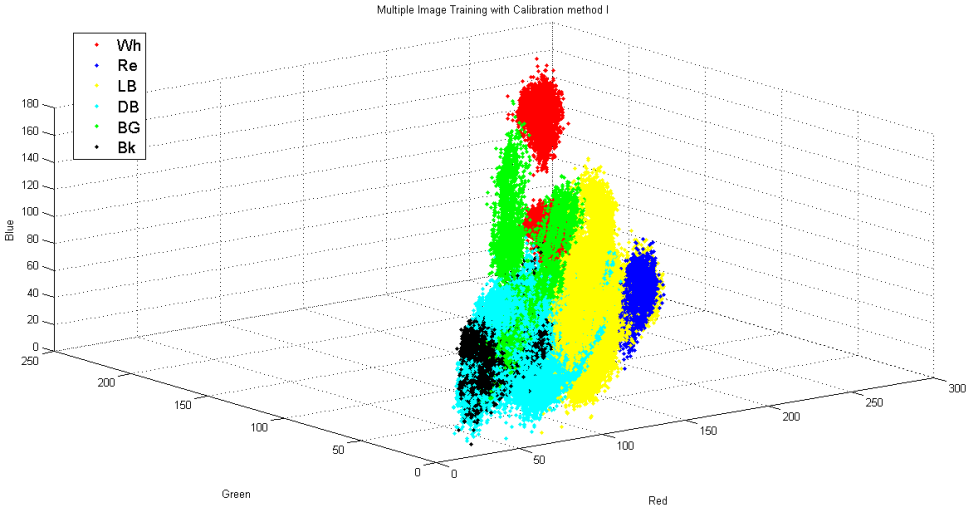
5.2.3 With Calibration

The resulting images from calibration (methods I, II and III) can be seen in Appendix D-F and a summary of separability results can be seen in Table 5.6.

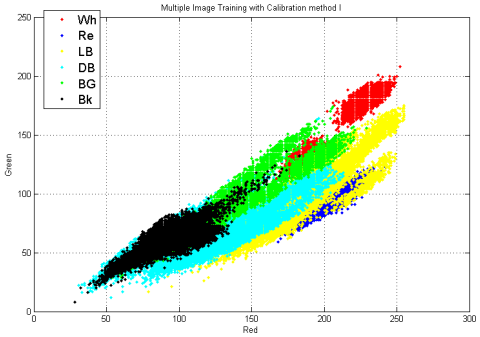
Image calibration method I (skin intensity calibration) allows better results comparing to the other two calibration methods and, since the MIT group represent the most realistic case, the results are presented just for this group using calibration method I.

Figure 5.9 show the 3D scatterplot and the projection in the three 2D planes (Red/Green, Green/Blue, Red/Blue) for MIT group with calibration method I, and Table 5.4 show the p-values as well as minimum (Min), maximum (Max), sample mean (μ) and sample standard deviation (σ) for each calibrated class in the MIT group. These results must be compared with Figure 5.8 and Table 5.1, which correspond to the original behavior (without calibration) of the MIT group.

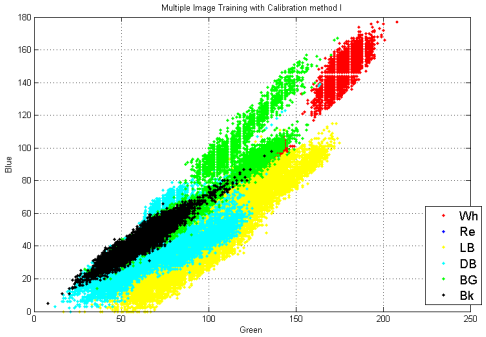
A simple image calibration technique (calibration method I) shortened the spectral range values of each color class in the MIT, as we can see comparing the calibrated spectral values in Table 5.4 with original spectral values in Table 5.1. Several calibrated classes have higher p-values than the respective p-values in non-calibrated classes, but other have lower p-values comparing to the original distribution, such as class Wh. For the class Wh, the original data generated p-values with an average of 0.85 and calibrated data produced p-values with an average of 0.63. However, all the color classes were treated as having Gaussian (or normal) distribution.



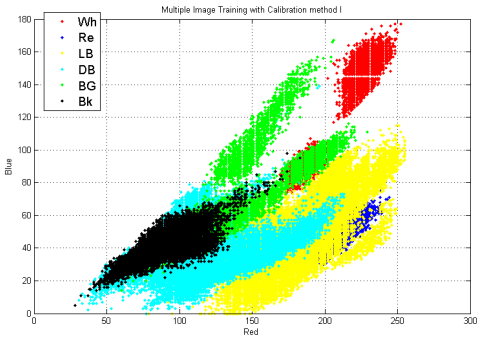
(a)



(b)



(c)



(d)

Fig. 5.9: Scatterplots for the single image training test with skin intensity calibration (method I): (a) on the RGB feature space, (b) a projection on the 2D Red, Green plane, (c) 2D Green, Blue plane and (d) 2D Red, Blue plane.

Tab. 5.4: Data behavior for multiple image training group with image calibration method I, presenting minimum (Min), maximum (Max), mean (μ), standard deviation (σ) and p-value with 5% of significance.

Classes	Min	Max	μ	σ	p-value	Min	Max	μ	σ	p-value	Min	Max	μ	σ	p-value
	Red channel					Green channel					Blue channel				
Wh	165	252	211	21.3	0.87	105	208	159	24.8	0.74	68	177	125	30.6	0.28
Re	168	244	209	10.6	0.91	59	128	95.7	9.34	0.97	15	75	48.0	7.30	0.98
LB	79	255	180	24.9	0.66	17	175	95.4	23.3	0.65	0	115	47.6	19.1	0.77
DB	31	213	105	28.0	0.54	12	164	60.3	15.8	0.80	0	139	36.8	12.1	0.91
BG	53	229	125	35.7	0.58	29	174	84.0	26.7	0.70	14	167	62.0	23.7	0.70
Bk	28	183	87.3	13.1	0.77	8	136	59.6	10.1	0.88	5	98	42.2	7.54	0.96

Wh - white, Re - Red, LB - Light Brown, DB - Dark Brown, BG - Blue Gray, Bk - Black

Tab. 5.5: Jeffries-Matusita - JM (top right triangular matrix) and Transformed Divergence - TD (bottom left) distance values, for the multiple-image training test with skin intensity calibration (method I).

Classes	Wh	Re	LB	DB	BG	Bk
Wh	-	2.00	1.86	1.96	1.73	2.00
Re	2.00	-	1.78	1.97	2.00	2.00
LB	1.83	1.77	-	1.48	1.84	2.00
DB	2.00	2.00	1.49	-	0.78	0.96
BG	1.98	2.00	1.78	0.23	-	1.11
Bk	2.00	2.00	(*)	(*)	(*)	-

Wh - white, Re - Red, LB - Light Brown, DB - Dark Brown
BG - Blue Gray, Bk - Black
(*) not possible to calculate

The JM and TD values, computed with the calibrated images, are presented in Table 5.5. There are a few cases where the indications from JM and TD distances are contradictory, and also some cases where it was not possible to calculate TD due to the fact that the data matrix was not symmetrical. The results presented in Tables 5.2, 5.3 and 5.5 are summarized in Table 5.6, as well as the results for the other two calibration methods, where the number of class pairs (15 in total) are grouped in well separable ($JM \geq 1.9$), moderately separable ($1.8 \leq JM < 1.9$) and non-separable or undistinguishable ($JM < 1.8$). As expected, class separability evaluation benefited slightly by performing image calibration to a reference. Once more, the best results were obtained by the calibration method I, as seen in Table 5.6. The results based on TD values are slightly better, but as it was not always possible to compute TD, the evaluation is mostly based on JM values.

5.2.4 Including the Skin

A final test was carried out, including an additional class: the skin. The skin can be seen as the background of the lesion in a dermoscopic image, but also as a class on its own. According to [72], the separability between skin class and non-skin classes was highest in RGB color space. However, this additional class increases the difficulty in the classification/discrimination problem. The JM and TD values were computed for the 7 classes case (21 pairs). A summary of the JM results for the various scenarios tested a presented in Table 5.7. This table shows the number of class pairs considered well separable in the feature space ($JM \geq 1.9$), moderately separable ($1.8 \leq JM < 1.9$) and undistinguishable ($JM < 1.8$). For the single-

Tab. 5.6: Summary of JM results for the various scenarios tested (number of class pairs).

Test scenario	$JM \geq 1.9$	$1.8 \leq JM < 1.9$	$JM < 1.8$
Single-Image Training	14	-	$1^{(a)}$
Multiple Image Training	7	1	$7^{(b)}$
M-image train. Calib. I	7	2	$6^{(c)}$
M-image train. Calib. II	8	-	$7^{(d)}$
M-image train. Calib. III	6	1	$8^{(e)}$

(a): BG-Bk

(b): Wh-LB, Wh-BG, Re-LB, LB-DB, DB-BG, DB-Bk, BG-Bk

(c): Wh-BG, Re-LB, LB-DB, DB-BG, DB-Bk, BG-Bk

(d): Wh-BG, LB-DB, LB-BG, LB-Bk, DB-BG, DB-Bk, BG-Bk

(e): Wh-LB, Wh-BG, Re-LB, LB-DB, LB-BG, DB-BG, DB-Bk, BG-Bk

Tab. 5.7: Summary of JM results for the various scenarios tested (number of class pairs) including the skin class.

Test scenario	$JM \geq 1.9$	$1.8 \leq JM < 1.9$	$JM < 1.8$
Single-Image Training	19	-	$2^{(a)}$
Multiple Image Training	11	1	9
M-image train. Calib. I	13	2	$6^{(b)}$
M-image train. Calib. II	10	-	11
M-image train. Calib. III	8	3	$10^{(c)}$

(a): Bk-BG, Wh-skin

(b): Wh-BG, Re-LB, LB-DB, DB-BG, DB-Bk, BG-Bk

(c): Wh-LB, Wh-BG, Wh-skin, Re-LB, LB-DB, LB-BG, LB-skin, DB-BG, DB-Bk, BG-Bk

image training case, most class pairs are distinguishable, even with the additional skin class. However, when the training uses all 28 test images, there are 9 class pairs that cannot be separable using the RGB feature space and only 11+1 distinguishable class pairs. When the images are calibrated to a reference (using method I), the number of distinguishable class pairs increases to 13+2 (out of 21).

5.2.5 Discussion

All the results exposed in this first section can be seen in [49] and in [73] with a slightly different dataset but based on the same 28 images.

The identification of color classes in dermoscopic images is a subjective task, which poses great challenges for an automatic implementation. The purpose of this first step of the proposed method was to evaluate the potential discrimination between the various Menzies color classes in dermoscopic RGB images. The tests

performed using the JM and TD separability metrics indicate that it is possible to identify and distinguish most color classes. In the most challenging case, where the class signature is obtained by multiple images (between 1 and 17 for each class, on a total of 28 test images), 7+1 of the 15 class pairs are distinguishable without image calibration. The result is slightly improved by using a simple calibration procedure, which normalizes the background intensity. Considering the skin as an additional class, a total of 13 class pairs are well separable and another 2 pairs are moderately separable (using calibration method I). Only 6 class pairs remain as non-separable in both cases (with or without the skin class).

Although most class pairs are distinguishable in RGB color images, at this point, the proposed method cannot yet be used since there are considerable confusion between some color classes. Section 5.4 show the results of an increase of separability distances using a non-linear technique explained in Section 4.8 but, firstly, the training and test sets are presented in the following section.

5.3 Train and Test Sets Description

Knowing already the classes behavior and assumed a Gaussian distribution to all, the MIT group must be divided in two sets: train and test sets. Actually, the MIT group refers to the all dataset images with a different denomination and because of that, we can divide it in train and test sets, as usual.

The necessity of the increasing of clusters separability distances in order to guarantee that the classifier will have no doubts when making a decision was already seen, i.e. to minimize the misclassification rates. Because of that, once more, the train and the test set must be fixed.

Both sets were chosen randomly, guaranteeing the representativeness of each color class (cluster). Table 5.8 shown which images are taken to represent each set. Overall, the training set contains 75% of dataset images and the test set contains 25% of dataset images (21 and 7 images, respectively).

The training set serves to collect representative samples of each cluster and consequent surface learning in the classification step. The test set is just used in the validation step in order to performed some accuracy measures, as SE and SP.

The spectral signatures of the training and test sets will be presented in the feature space chosen (RGB) with and without the calibration method I. In Table 5.9 the spectral values (minimum, maximum, sample mean, sample standard deviation) are presented, as well as the p-values resulted from the KS test with significant level

Tab. 5.8: Training and Testing sets with 75% and 25% of images dataset (28 images), respectively.

Colors	Training set (IMDxx)	Testing set (IMDxx)
Wh	17, 20	-
Re	27	-
LB	01, 03, 04, 05, 08, 09, 10, 11, 13, 22, 23, 25, 27	06, 14, 28
DB	01, 02, 03, 04, 07, 09, 11, 12, 16, 17, 22, 26	06, 15, 19, 21, 28
BG	12, 17, 20, 26	19, 28
Bk	13, 20	18
No. images	21	7

Wh - white, Re - Red, LB - Light Brown, DB - Dark Brown, BG - Blue Gray, Bk - Black

of $\alpha = 5\%$, with and without calibration method I. In Table 5.10 the same values are shown for the test set with and without calibration method I.

In the training set, the color classes follow Gaussian distributions, with minimum p-value value of 0.50 for color class LB in the Blue color channel without calibration, and with minimum p-value value of 0.59 for the color class DB in the Red color channel. Thus, it is possible to apply the equations of Bhattacharyya distance (Equation 4.15) and Divergence (Equation 4.13). In general, the spectral range values for each color class and for the skin class becomes smaller when the images are subject to calibration method I, which is favorable in view of the degree of overlapping. In Figure 5.10 the 3D scatterplot for all the train set without calibration in the RGB feature space is shown, as well as the projections in the three 2D planes. To compare, Figure 5.11 shows the calibration method I results.

Tab. 5.9: Training data with and without calibration, presenting minimum (Min), maximum (Max), mean (μ), standard deviation (σ) and p-value with 5% of significance.

TRAIN SET																
Without Calibration																
Classes		**Min**	**Max**	μ	σ	**p-value**	**Min**	**Max**	μ	σ	**p-value**	**Min**	**Max**	μ	σ	**p-value**
	No. of pixels	**Red channel**					**Green channel**					**Blue channel**				
Wh	10320	181	226	205	6.07	0.99	133	181	156	5.75	0.99	116	175	148	8.51	0.96
Re	7743	165	240	205	10.4	0.95	70	152	114	11.1	0.94	22	110	70.4	10.7	0.93
LB	47453	79	255	165	30.3	0.56	17	212	96.2	30.4	0.58	0	191	60.1	33.8	0.50
DB	41968	51	207	112	26.3	0.61	14	154	71.2	18.4	0.75	0	151	56.1	18.7	0.71
BG	16859	54	185	109	19.7	0.73	35	151	84.9	14.7	0.81	18	165	81.7	18.0	0.75
Bk	18548	47	168	85.0	13.6	0.87	31	133	67.2	7.93	0.96	29	129	62.9	7.78	0.96
Skin	63259	134	255	221	14.9	0.85	96	255	181	22.8	0.76	65	250	159	32.2	0.62
With Calibration Method I																
	No. of pixels	**Red channel**					**Green channel**					**Blue channel**				
Wh	10320	202	252	228	6.75	0.97	153	208	179	6.61	0.98	117	177	150	8.88	0.95
Re	7743	168	244	209	10.6	0.91	59	128	96.7	9.34	0.97	15	75	48.0	7.30	0.98
LB	47453	79	255	175	26.3	0.66	17	175	90.5	23.5	0.66	0	115	44.3	19.2	0.78
DB	41968	49	213	115	28.4	0.59	12	164	64.9	17.0	0.84	0	139	38.8	13.8	0.91
BG	16859	58	206	116	23.6	0.68	33	174	78.2	20.3	0.74	14	167	58.1	21.6	0.70
Bk	18548	46	183	87.2	13.8	0.89	24	132	59.4	10.5	0.95	17	95	41.4	7.58	0.97
Skin	63259	126	255	228	8.88	0.80	77	230	165	9.09	0.92	51	165	113	8.26	0.94

Wh - White, Re - Red, LB - Light Brown, DB - Dark Brown, BG - Blue Gray, Bk - Black

Tab. 5.10: Test data with and without calibration, presenting minimum (Min), maximum (Max), mean (μ), standard deviation (σ) and p-value with 5% of significance.

TEST SET																
Without Calibration																
Classes		**Min**	**Max**	μ	σ	**p-value**	**Min**	**Max**	μ	σ	**p-value**	**Min**	**Max**	μ	σ	**p-value**
	No. of Pixels	**Red channel**					**Green channel**					**Blue channel**				
Wh	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Re	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
LB	18224	114	245	189	14.4	0.78	26	194	116	23.8	0.72	0	187	77.1	35.0	0.58
DB	27810	28	184	86.0	20.0	0.76	16	141	56.4	10.3	0.87	3	127	48.3	10.6	0.88
BG	5094	53	201	139	39.3	0.11	33	153	104	30.0	0.43	38	154	105	27.5	0.70
Bk	25032	31	191	96.1	13.8	0.74	11	179	78.8	13.0	0.77	10	179	78.2	13.6	0.80
Skin	23939	178	255	226	16.4	0.91	132	240	184	20.5	0.81	97	233	166	28.3	0.71
With Calibration Method I																
	No. of Pixels	**Red channel**					**Green channel**					**Blue channel**				
Wh	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Re	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
LB	18224	117	249	193	14.2	0.78	26	168	108	16.8	0.82	0	112	56.1	15.9	0.84
DB	27810	31	187	90.5	19.8	0.80	16	122	53.4	10.6	0.86	2	79	33.7	7.91	0.93
BG	5094	53	229	154	50.4	0.00	39	157	103	35.1	0.08	23	116	75.1	25.5	0.26
Bk	25032	28	174	87.4	12.5	0.78	8	136	59.7	9.83	0.89	5	98	42.8	7.46	0.97
Skin	23939	185	255	233	8.66	0.95	126	227	171	10.2	0.92	77	172	116	8.52	0.91

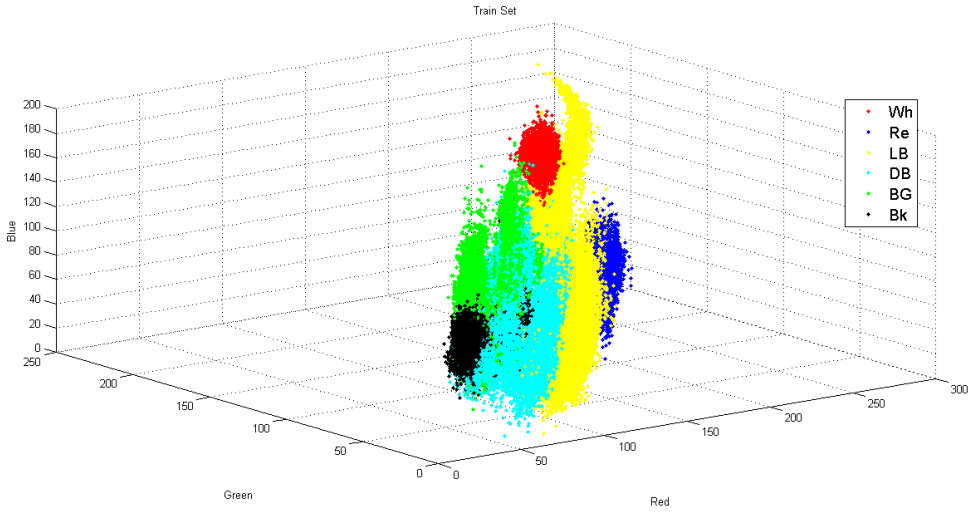
Wh - White, Re - Red, LB - Light Brown, DB - Dark Brown, BG - Blue Gray, Bk - Black

The test set contains no sample of some classes, including: Wh and Re. This is due to the low number of samples for each class. In Table 4.2 we saw that the class Re appears just in one image and the class Wh appears in two images of the dataset available. Since the classifier must learn each class, the image that contains the Re class belongs to the train set then is not part of the test set. The class Wh have different behaviors in both images (IMD17 and IMD20) - see Appendix B -, because of that, both images were assigned as belonging to the train set.

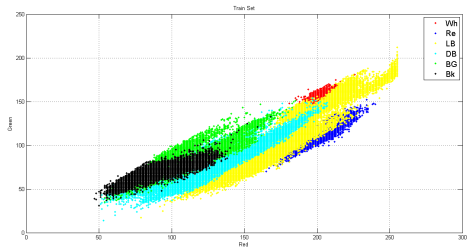
The test set is not used to generate data as classifier input, however the spectral values as well as the scatterplots for all classes are shown in Table 5.10 and Figures 5.12 and 5.13, respectively.

Table 5.11 shows the JM and TD separability measures for both sets. Once more, the test set was not used as input data to the classifier, the results only serve to illustrate the data behavior and separability inside the set. The JM and TD separability measures for the train set are also presented in Table 5.11, before and after the calibration process. Before the implementation of the calibration process, 14 out of 21 class pairs were separable in the RGB feature space (Table 5.11). After the calibration process, as expected, the separability results improved: JM and TD metrics indicate that 17 out of 21 pairs of classes are separable in the RGB feature space, the others 4 pairs of classes (LB-DB, DB-BG, DB-Bk and BG-Bk) have low JM and TD values.

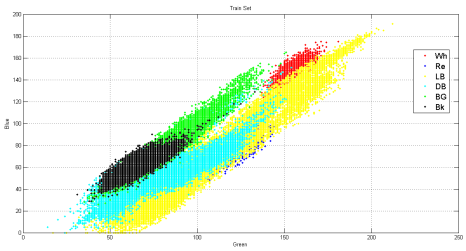
The Non-linear Cluster Transformation (NCT) must be now fixed in the training set in order to increase the clusters (color classes) separability. The next section reports the results of this step.



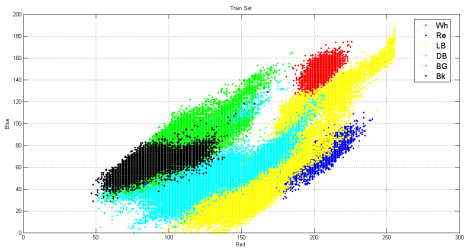
(a)



(b)

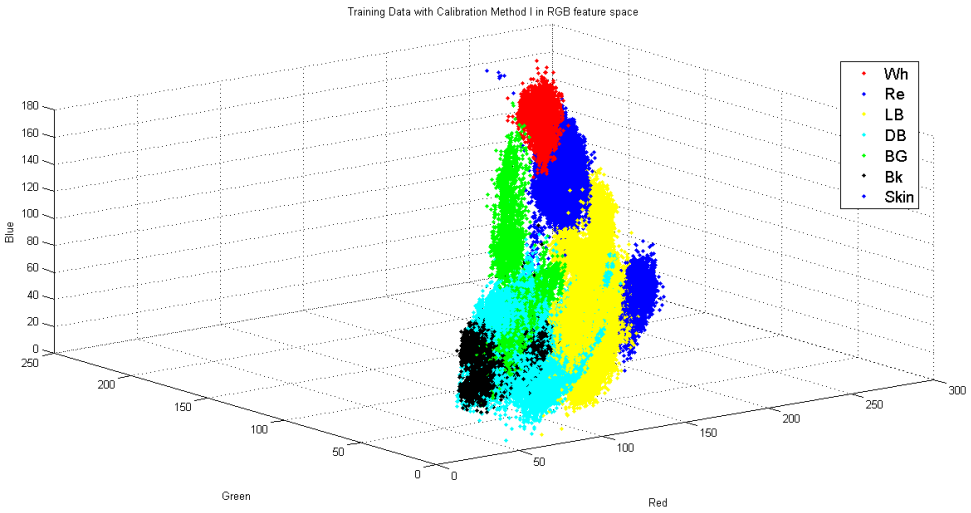


(c)

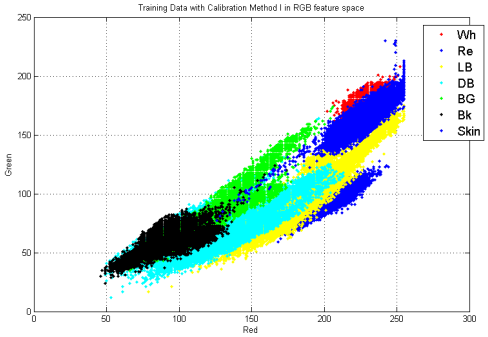


(d)

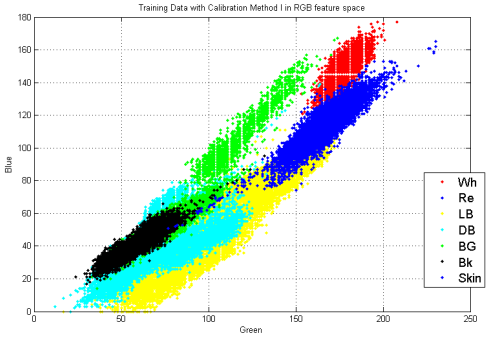
Fig. 5.10: Scatterplots for the training set: (a) on the RGB feature space; a projection on the (b) 2D Red, Green plane, (c) 2D Green, Blue plane and (d) 2D Red, Blue plane, without calibration.



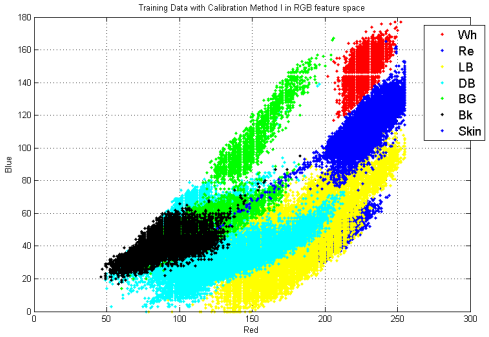
(a)



(b)

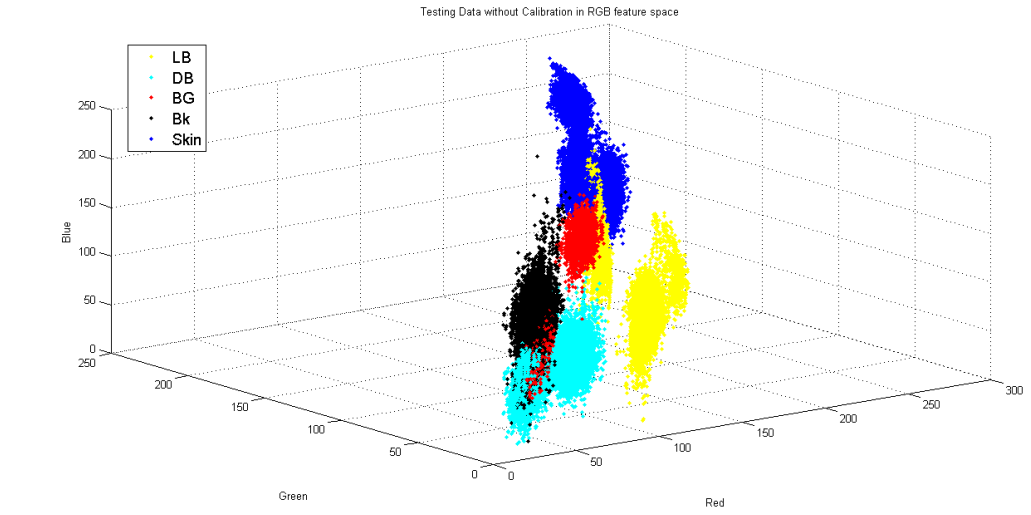


(c)

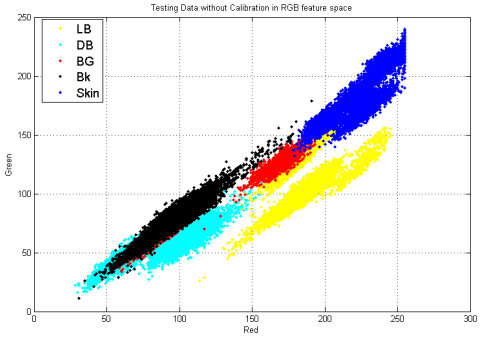


(d)

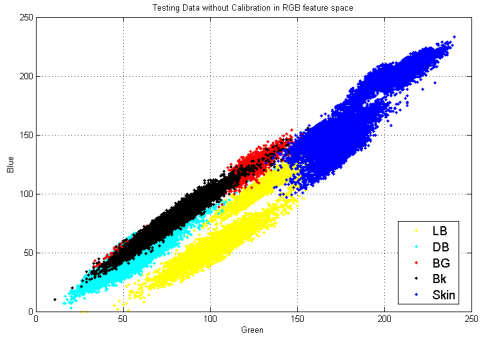
Fig. 5.11: Scatterplots for the training set: (a) on the RGB feature space; a projection on the (b) 2D Red, Green plane, (c) 2D Green, Blue plane and (d) 2D Red, Blue plane, with calibration method I.



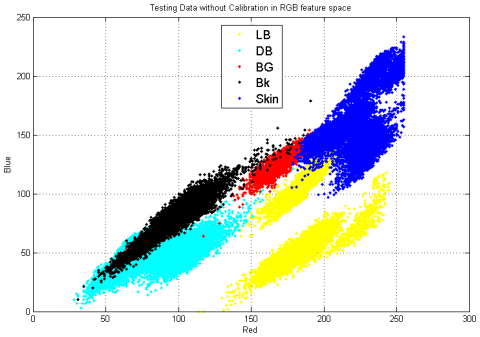
(a)



(b)

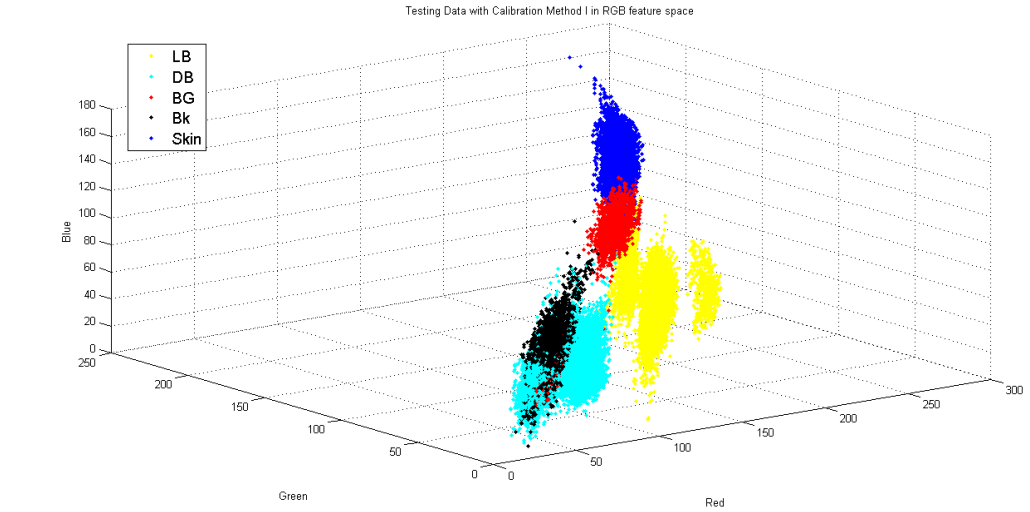


(c)

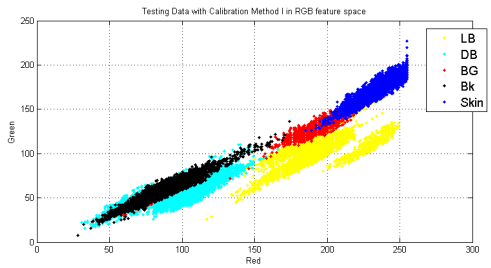


(d)

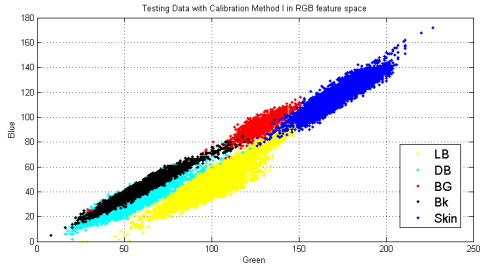
Fig. 5.12: Scatterplots for the test set: (a) on the RGB feature space; a projection on the (b) 2D Red, Green plane, (c) 2D Green, Blue plane and (d) 2D Red, Blue plane, without calibration.



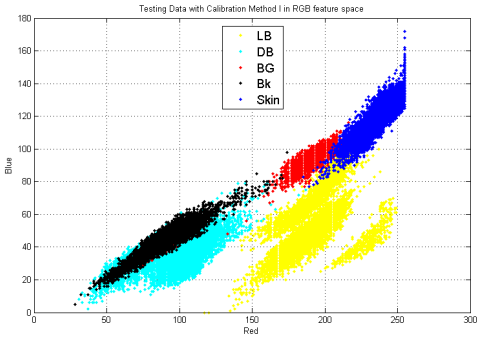
(a)



(b)



(c)



(d)

Fig. 5.13: Scatterplots for the test set: (a) on the RGB feature space; a projection on the (b) 2D Red, Green plane, (c) 2D Green, Blue plane and (d) 2D Red, Blue plane, with calibration method I.

Tab. 5.11: Jeffries-Matusita - JM (top right triangular matrices) and Transformed Divergence - TD (bottom left) distance values, for the training and test sets, with and without calibration method I.

WITHOUT CALIBRATION														
SET	**TRAIN**							**TEST**						
Classes	**Wh**	**Re**	**LB**	**DB**	**BG**	**Bk**	**Skin**	**Wh**	**Re**	**LB**	**DB**	**BG**	**Bk**	**Skin**
Wh	-	2.00	1.90	2.00	2.00	2.00	1.59	-	-	-	-	-	-	-
Re	2.00	-	1.60	1.97	2.00	2.00	1.93	-	-	-	-	-	-	-
LB	1.99	1.62	-	1.26	1.94	1.99	1.56	-	-	-	2.00	1.94	2.00	1.49
DB	2.00	2.00	0.99	-	1.02	1.27	1.97	-	-	2.00	-	1.70	1.56	2.00
BG	2.00	2.00	1.97	0.47	-	0.78	2.00	-	-	2.00	1.81	-	1.39	1.91
Bk	2.00	2.00	1.98	0.71	0.38	-	2.00	-	-	2.00	1.37	1.36	-	2.00
Skin	1.09	2.00	1.57	1.99	2.00	2.00	-	-	-	1.42	2.00	1.74	2.00	-
WITH CALIBRATION METHOD I														
SET	**TRAIN**							**TEST**						
Classes	**Wh**	**Re**	**LB**	**DB**	**BG**	**Bk**	**Skin**	**Wh**	**Re**	**LB**	**DB**	**BG**	**Bk**	**Skin**
Wh	-	2.00	2.00	2.00	2.00	2.00	1.98	-	-	-	-	-	-	-
Re	2.00	-	1.79	1.96	2.00	2.00	2.00	-	-	-	-	-	-	-
LB	2.00	1.77	-	1.22	1.94	1.99	1.96	-	-	-	1.99	1.98	2.00	1.92
DB	2.00	2.00	1.19	-	0.77	0.97	2.00	-	-	2.00	-	1.49	1.12	2.00
BG	2.00	2.00	1.92	(*)	-	1.03	2.00	-	-	2.00	1.45	-	1.25	1.90
Bk	2.00	2.00	2.00	0.03	(*)	-	2.00	-	-	2.00	0.38	1.47	-	2.00
Skin	1.99	2.00	1.99	2.00	2.00	2.00	-	-	-	1.97	2.00	1.99	2.00	-

(*) not possible to calculate

Wh - White, Re - Red, LB - Light Brown, DB - Dark Brown, BG - Blue Gray, Bk - Black, Skin - healthy Skin

5.4 Increasing Clusters Separability

Hitherto, in the train set, with a simple image calibration technique (method I), a total of 17 out of 21 pairs of clusters can be separable. The NCT technique, firstly introduced by Polikar et al.[59][60] and presented in Section 4.8, was applied to the data in order to increase the clusters separability distances results and guaranteeing that the classifier receives totally separable clusters.

Polikar's NCT perform a first step of outlier removal, however, for the present data it is not feasible to remove points since the dataset is very limited (28 images with medical inconsistencies). In fact, the classes have a nucleus around which there are scattered points (possibly outliers). However, a larger data set will have different behavior and as such, the outliers removal stage is not applied.

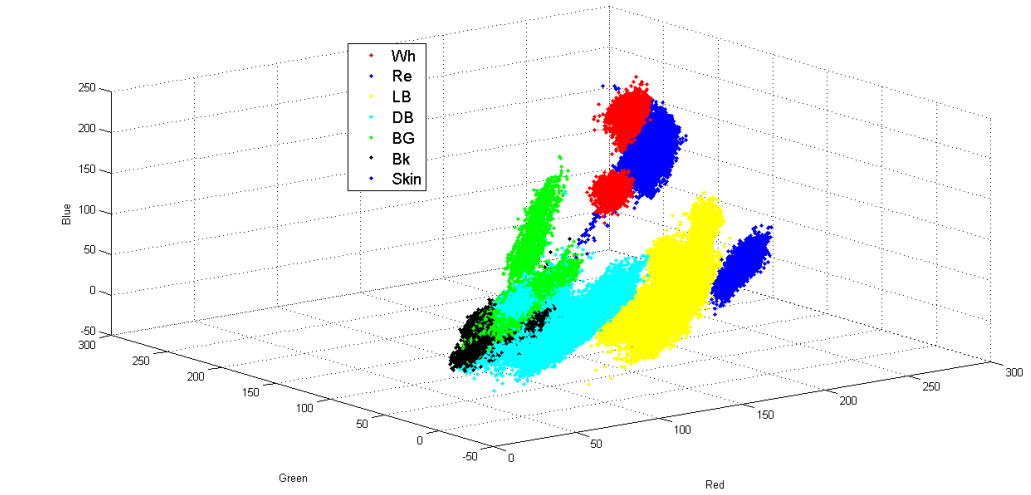
The cluster translation was only performed in order to increase the separability, maintaining the same internal structure (intra-class) of each cluster and increasing the inter-class distances.

In order to assess the increase of separability or to know if it is enough separable in the RGB feature space, we can simply plotting the data and determining how much translations is necessary.

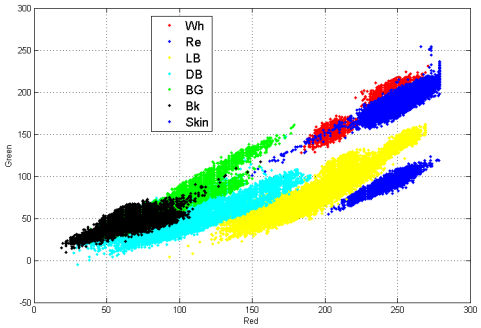
To the calibrated training data that ensured the separability of 17 out of 21 pairs separable, five Polikar's transformations were applied to increase inter-class distances between the four pairs LB-DB, DB-BG, DB-Bk and BG-Bk. Figure 5.14 shown the scatterplot of the resulting set of color classes after applied **five** Polikar transformations to the original.

The resulted dataset allowed the separation of just a couple of classes, from 17 to 18 out of 21 pairs of color classes separable in the RGB feature space. Table 5.12 show the JM and TD separability measures between color classes. The pair LB-DB ceased to be confused in the RGB feature space. However, three pairs of classes remains non-separable (DB-BG, BG-Bk and DB-Bk). This may be due to the low sampling of the classes BG and Bk and due to the similar behavior between them. Even with the Polikar's translations, the direction of progression is the same in these classes, which does not help in increasing the separability.

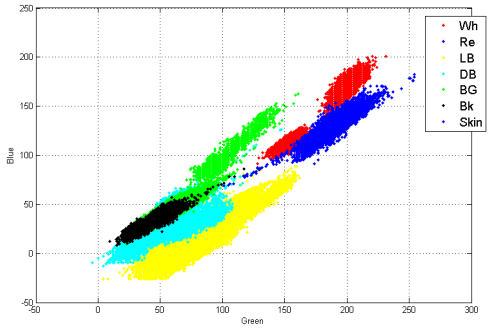
Now, the three classes with similar behavior (DB, BG and Bk) can be combined as one - Subsection 4.7 - and training a classifier to labelling new images in 5 possible groups: **group 1** with the Wh class, **group 2** with the Re class, **group 3** with the LB class, **group 4** with the DB, BG and Bk classes, and **group 5** with the skin class.



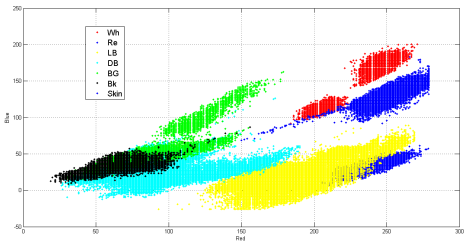
(a)



(b)



(c)



(d)

Fig. 5.14: Scatterplots for the test set with skin intensity calibration (method I) and a non-linear technique which increase the clusters separability distances: (a) on the RGB feature space; (b) a projection on the 2D Red, Green plane, (c) 2D Green, Blue plane and (d) 2D Red, Blue plane.

Tab. 5.12: .
Jeffries-Matusita and Transformed Divergence distance values, for the dataset resulting of five Polikar's transformations.

Classes	Wh	Re	LB	DB	BG	Bk	Skin
Wh	-	2.00	2.00	2.00	2.00	2.00	2.00
Re	2.00	-	1.98	2.00	2.00	2.00	2.00
LB	2.00	1.99	-	1.89	2.00	2.00	2.00
DB	2.00	2.00	1.95	-	1.07	1.15	2.00
BG	2.00	2.00	2.00	0.72	-	1.05	2.00
Bk	2.00	2.00	2.00	0.82	0.04	-	2.00
Skin	2.00	2.00	2.00	2.00	2.00	2.00	-

Wh - white, Re - Red, LB - Light Brown, DB - Dark Brown
BG - Blue Gray, Bk - Black, Skin - healthy skin

Tab. 5.13: Jeffries-Matusita and Transformed Divergence distance values, for the grouping dataset resulting of the 5th Polikar transformation.

Group	1	2	3	4	5
1	-	2.00	2.00	2.00	2.00
2	2.00	-	1.98	2.00	2.00
3	2.00	1.99	-	1.95	2.00
4	2.00	2.00	1.99	-	2.00
5	2.00	2.00	2.00	2.00	-

Group 1 - Wh, Group 2 - Re, Group 3 - LB,
Group 4 - DB, BG, Bk, Group 5 - Skin
Wh - white, Re - Red, LB - Light Brown, DB - Dark Brown
BG - Blue Gray, Bk - Black, Skin - healthy skin

In this case, the JM and TD separability measures are presented in Table 5.13. Considering these 5 groups to represent our 7 classes, according to JM and TD metrics, the 5 groups are now completely separable in the RGB feature space. The translational functions will be fixed in the train set and will be applied in the test set after report the classifier labelling.

5.5 Classification and Post-Processing - Step 1

Consider the five groups identified above: group 1 with the class Wh, group 2 with the class Re, group 3 with the class LB, group 4 with the classes DB, BG and Bk, and group 5 with the skin class. Neural networks (NNs) in competition (Figure 4.13) can now be used to learn each one of the five group surfaces. In Table 5.14 all the parameters used to all NNs are shown.

Tab. 5.14: Parameters of the neural network classifier.

Neural Networks Parameters	
Structure	3-4-3
Algorithm	Backpropagation
Learning criteria	MSE
Training algorithm	Batch Gradient Descent with Momentum
Transfer function	$tansig - tansig^a$
Momentum rate	0.15
Epochs	500
(a) see Figure 5.15	

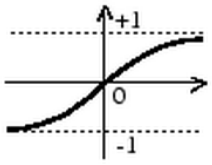


Fig. 5.15: Hyperbolic tangent sigmoid transfer function.

The hyperbolic tangent sigmoid (*tansig* in Figure 5.15) as transfer functions allows adjustment of the weights by backpropagation in the range of $[-1, 1]$, saturating all values greater than 1 or lower than -1. So, it is critical to normalize the data after train the network. The data were normalized to $[0, 1]$.

Since there are a large number of pixels for each group class, 500 epochs were used to train each of the five NNs (see Table 5.9). The momentum rate was fixed in 0.15.

All of the five NNs have the same structure: 3-4-3, 3 input neurons at the input layer, one hidden layer with 4 neurons, and an output layer with 3 neurons (Figure 5.14). Since each NN learns the group surface (with local and global behavior of the distribution function), we train the NNs in order to minimize the MSE error between the input and the output, where the output has a target value which corresponds to the data points in the input layer. In Figure 5.16, $t_R = x_R$, $t_G = x_G$ and $t_B = x_B$ and the difference between the output layer and the target data will be minimal (ideally zero) using the error backpropagation trough the NN using the MSE criterion.

For the classification of all images in the test set (Table 5.8) in five possible labels (group 1, 2, 3, 4 and 5), with NNs in competition. To present the result and explain the following procedures one image classification case (IMD06) will be used - Figure 5.17(a).

In Table 5.15 the validation of colors present in this first classification is shown. For the image IMD06 the presence of group labels 3, 4 and 5 was truly identified.

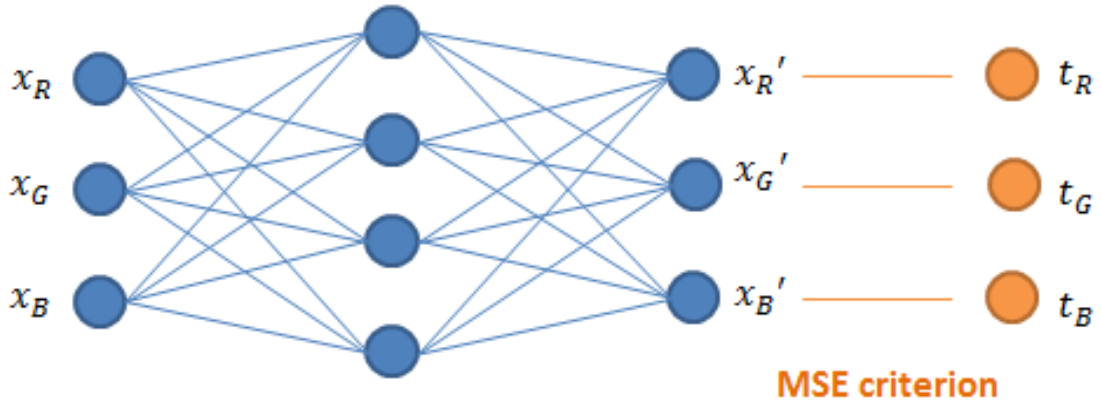


Fig. 5.16: Structure of the neural network classifier.

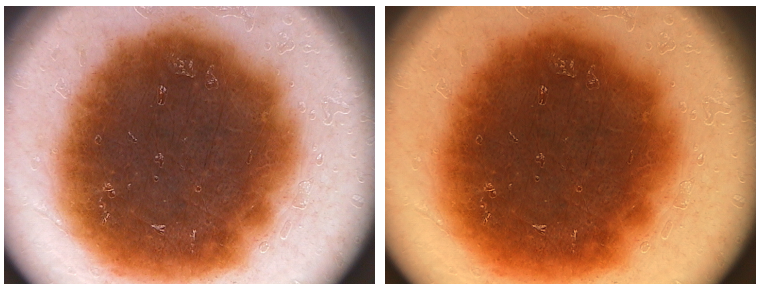
Tab. 5.15: Classification results - step 1: classification in 5 rearranged groups.

Presence of color	IMD06	IMD14	IMD15	IMD18	IMD19	IMD21	IMD28
Group 1	TN	TN	TN	FP	TP	FP	FP
Group 2	TN	TN	TN	TN	TN	TN	TN
Group 3	TP	TP	TN	FP	TN	TN	TP
Group 4	TP	FP	TP	TP	TP	TP	TP
Group 5	TP	TP	TP	TP	TP	TP	TP
SP (%)	100	67	100	33	100	67	50
SE (%)	100	100	100	100	100	100	100
TP - True Positives, FP - False Positive, TN - True Negative, FN - False Negative							
SE - Sensibility, SP - Specificity							

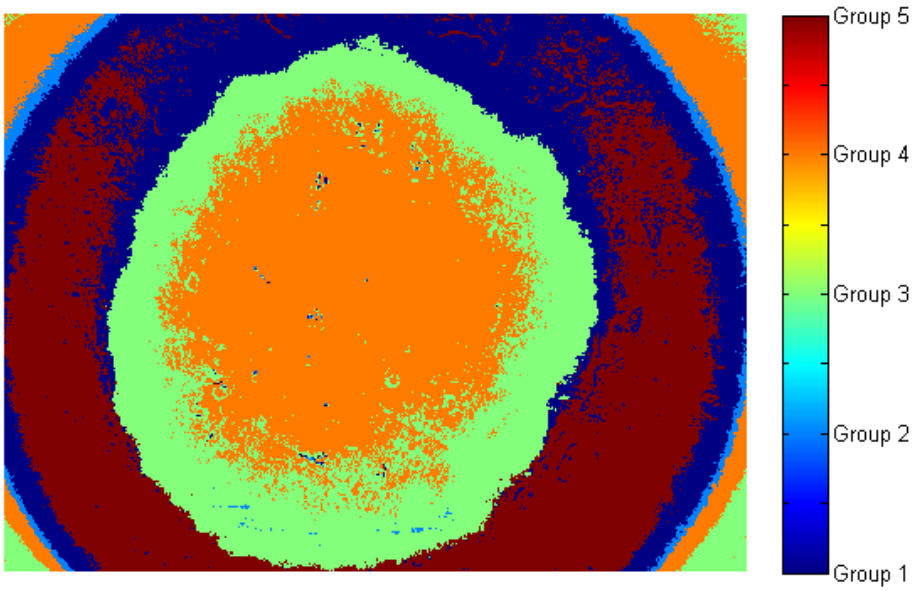
As we can seen in Table 4.2, image IMD06 has the colors LB, DB and healthy skin. Thus, in the rearranged groups it corresponds to the group labels: 3, 4 and 5. At the same time, the first classification step does not identify the group labels 1 and 2, which are correct, producing the TN (True Negatives) in that group labels. The performance for this image at this first classification step was 100% for sensibility and 100% for specificity.

The result for the other test images (IMD14, IMD15, IMD18, IMD19, IMD21 and IMD28 - Table 5.8) is also presented in Table 5.15. All of the test images had 100% of sensibility, which means that the classifier is sensible to the true positives classes. However, the specificity goal has an average of 57%, which means that the classifier identifies more colors present than the real colors that the PSL had. The resulted classified images are available in Appendix G.

As it is known, all the images has, at least, a small sample of healthy skin. In general, a large amount of skin class (group 5) was wrongly classified as belonging

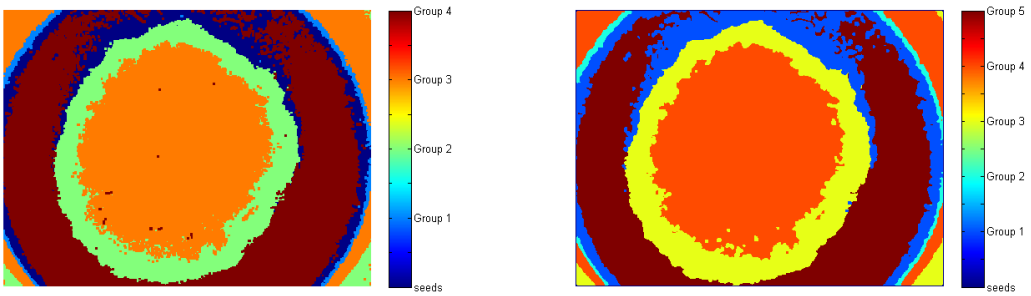


(a) (b)



(c)

Fig. 5.17: Results of the neural networks in competition classification considering class groups for image IMD06: (a) original image IMD06, (b) calibrated image, resulting for the calibration method I, and (c) classified image in 5 possible labels.



(a) (b)

Fig. 5.18: Results of morphological operators and filtering in the step 1 classified images (IMD06). The seeds legend correspond to the initialization of the morphological operators and filtering.

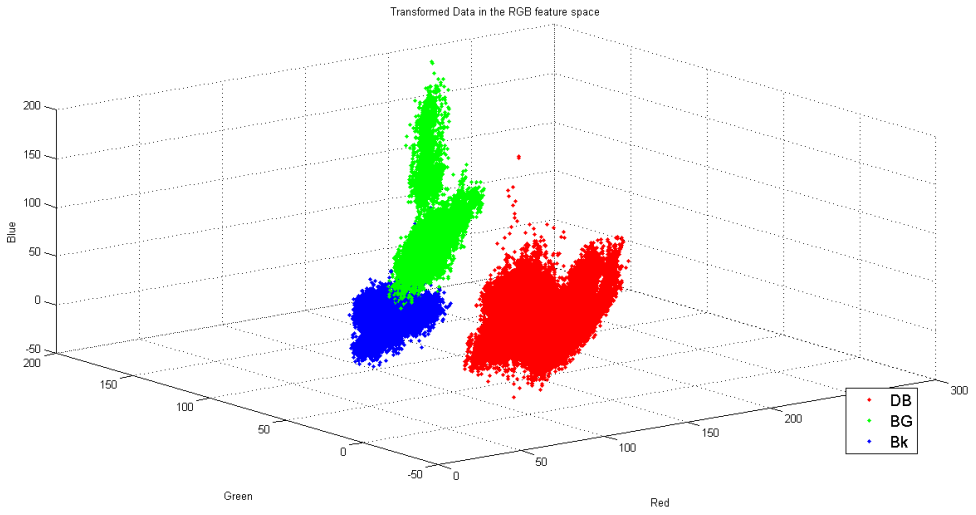
to group 1 (class Wh - White). This happens in all the test images, decreasing the specificity measure.

To solve the problem of misclassification of group label 5, some post-processing can be implemented in order to increase the skin classification area. Morphological operators were used, in particular, dilation with 5x5 square filter and then an mode filter to remove isolated points inside color regions. The result of these operations are presented as an example in Figure 5.18(a) and (b), respectively, for the test image IMD06. These images must be compared with results in Figure 5.17(c).

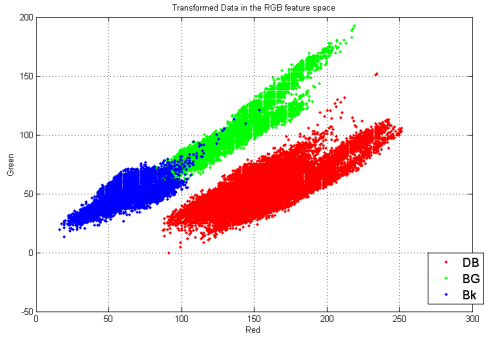
Now, with the problem of misclassification of group label 5 partially solved, we must divide the group label 4 in three possible colors: DB, BG and Bk. Appendix G also shows the rest of the test images post-processed.

5.6 Classification and Post-Processing - Step 2

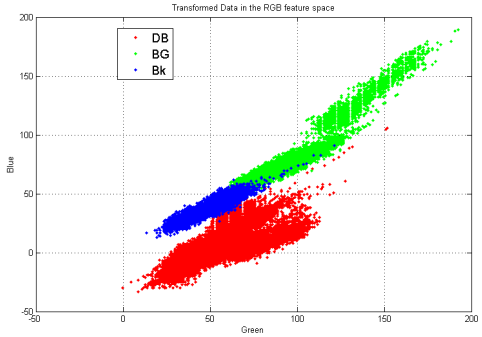
Once more, the division of group label 4 in three classes that compose it: DB, BG and Bk, must be performed. As presented in Table 5.4, the spectral signatures of the DB, BG and Bk color classes are very similar, i.e, there are a large amount of shared space in the RGB feature space. Thus, the Polikar technique can be applied to increase the color classes separability distances. After some experiments it was found that 12 Polikar transformations were enough to guarantee good separability of the three color classes, as is shown in Table 5.16 and Figure 5.19.



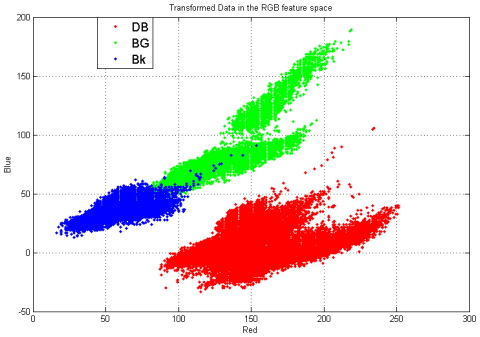
(a)



(b)



(c)



(d)

Fig. 5.19: Scatterplots for the DB, BG and Bk color classes after 12 Polikar transformation: (a) on the RGB feature space, (b) a projection on the 2D Red, Green plane, (c) 2D Green, Blue plane and (d) 2D Red, Blue plane.

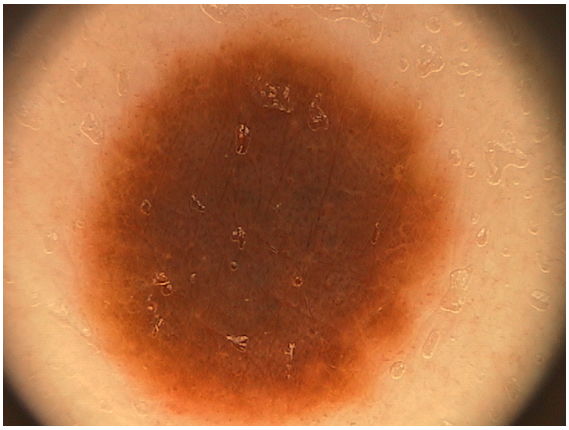
Tab. 5.16: Polikar transformations for increase separability distances in the three color classes that compose group label 4.

Classes	DB	BG	Bk
DB	-	1.99	1.99
BG	2.00	-	1.78
Bk	2.00	1.88	-

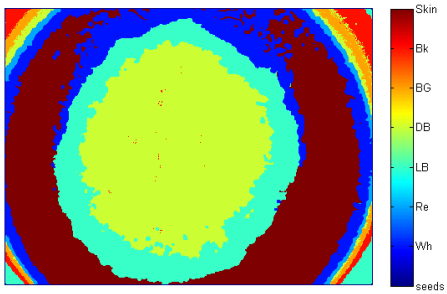
DB - Dark Brown, BG - Blue Gray, Bk - Black

For this transformed dataset, three NNs for each of the color classes DB, BG and Bk were also trained with the same parameters indicated in Table 5.14. The result of this second classification, for the image IMD06, is presented in Figure 5.20(b). For the others test images the results are presented in Appendix G also present the image results.

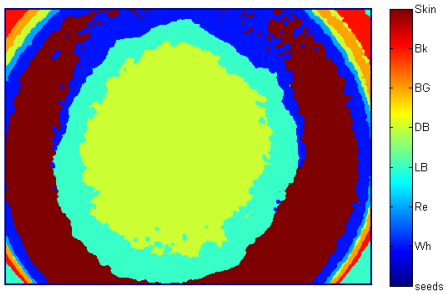
The output of the second classification will be subjected to a post-processing procedure, such as: filtering, to remove isolated points inside classified regions; and black corners removal. Figure 5.20 presents the result of the post-processing procedures, (c) and (d) for IMD06.



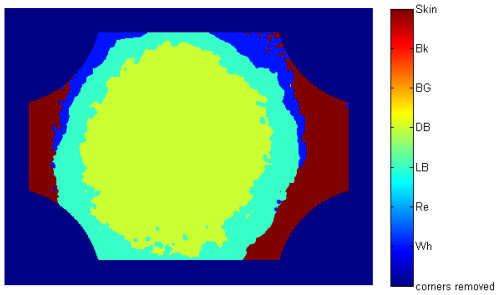
(a)



(b)



(c)



(d)

Fig. 5.20: Results of the second classification and post-processing for image IMD06: (a) calibrated image IMD06, (b) second classification result, (c) filtering and (d) corners removal.

5.7 Validation

Finally, the identification of the colors present in the test images is now possible. However, in order to remove some false positives, a color is indicated as present in the PSL if there are more than 3% (out of the total number of pixels) with the corresponding label. Establishing this criterion, Table 5.17 present the validated results of the post-processed images. For comparison with the human diagnosis of the PSLs in the test set, Table 4.2 can be used.

Tab. 5.17: Validation results.

Presence of color	IMD06	IMD14	IMD15	IMD18	IMD19	IMD21	IMD28
Wh	FP	TN	TN	FP	FP	FP	FP
Re	TN	TN	TN	TN	FP	TN	TN
LB	TP	TP	FP	TN	FP	TN	TP
DB	TP	TN	TP	TN	TP	FN	TP
BG	TN	TN	TN	FP	TP	FP	TP
Bk	TN	TN	TN	TP	TN	FP	FP
Skin	TP	TP	TP	TP	TP	TP	TP
SP (%)	75	100	80	60	25	60	33
SE (%)	100	100	100	100	100	50	100
Accuracy (%)	86	100	86	71	57	50	71
TP - True Positives, FP - False Positive, TN - True Negative, FN - False Negative							
SE - Sensibility, SP - Specificity							
Wh - White, Re - Red, LB - Light Brown, DB - Dark Brown, BG - Blue Gray, Bk - Black, Skin - Healthy skin							

The image that collects best performance measures was IMD14 with 100% of accuracy in the identification of the dermoscopic colors. The test images IMD06 and IMD15 produced 86% of accuracy, both with sensibilities of 100%. The images IMD18 and IMD28 obtained 71% of accuracy. The worst cases were assigned for the images IMD19 and IMD21 with accuracy of 57% and 50%, respectively. This bad diagnosis is due to the hairs present in the image IMD19 and the poor quality of the image IMD21. In general, the test images obtained high scores of sensibility (except for image IMD21) and intermediate values for specificity (except for images IMD19 and IMD28).

Using Equation 4.1, the diagnose of each PSL image in the test set as benign or suspicious is made just counting the colors present in images. The machine diagnosis said that the image IMD14 is benign and the other 6 images (IMD06, IMD15, IMD18, IMD19, IMD21 and IMD28) are suspicious of being malignant PSL

(or melanomas). The machine diagnosis indicates more benign PSL as suspicious ones than the human diagnosis.

Chapter 6

Concluding Remarks

An alternative approach to the standard computer based analysis of dermoscopic images is offered by the Menzies diagnosis method. The method consists on the identification of the presence of colors, out of 6 possible color classes, in a dermoscopic image. This method has some advantages comparing to the standard approaches, since it does not require the lesion segmentation. However, the identification of color classes in dermoscopic images is a subjective clinical task, which poses great challenges for an automatic implementation.

The calibration technique improve significant pairs of separable classes, at the same time, standardizing the range values of each color class. The Polikar's non-linear technique presents itself as an useful transformation and allows the separability of all rearranged groups.

The neural networks methods works well in the 1^{st} step. In the 2^{nd} step, some misclassification occurred. Nevertheless, the discrimination between the various Menzies color classes in dermoscopic RGB images performed 93% of sensibility, 62% of specificity and 74% of accuracy averaged measures. This result indicates that it is possible discriminate the different colors inside a dermoscopic lesion, mimicking the human diagnosis.

In order to confirm the preliminary results presented here, a collection of images should be tested. The dataset used is very small and does not contains sufficient number of sample data.

Future Research

There are several lines of work that can be followed after this work, such as those listed below:

- Analysis of the subjectivity in the human perception of color. In order to evaluate this aspect, a more extensive collection of images and clinical evaluation is needed, preferably with more than one medical evaluation for each dermoscopic image.

- A statistical study must be performed in order to weight the colors that has the malignancy more evident than other colors. For instance, color classes LB and DB are common dermoscopic colors, appearing in almost the entirely dataset and do not imply, itself, the PSL malignancy. This statistical evidence must be taken into account.

- Another procedure in the 2^{nd} classification step could evolve the use of different classifiers or using different metrics to label the test pixels. A clustering technique that will possibly decrease the misclassification ratio is the supervised Gustafson-Kessel algorithm. Its main feature is the local adaptation of the distance metric to the shape of the cluster, using different metrics for different classes, generating local adaptation without loss of generalization.

- Sub-classes within a color class could be considered, as the separability measures indicate. For instance, in the present dataset, the Wh color class could be divided in 2 sub-classes, but that was not necessary since the test set does not contain the White (Wh) color.

- The Neural Networks classifier could have more input neurons, as the Hue, Saturation and Intensity signatures of the HSI color space. However, the use of the MSE criterion, which is based in the Euclidean metric, should be carefully applied the feature Hue taking into account it is a circular measure (an angle). For instance, an angle (or Hue assignment) of 0° and other of 359° are very close, however, with the MSE criterion, the points will be considered far from each other.

- For the implementation of other features, different criteria could be used to the training of the NNs classifier, including criteria based on Information Theoretic Learning (ITL), such as: maximization of the mutual information between the

input and the output layer, or minimization of the quadratic entropy between the input and the output layer. Moreover, ITL criteria allows the classes have different distribution functions.

Bibliography

[1] Abramowitz M.; Davidson M. W. Introduction to microscopy. *Molecular Expressions. Retrieved 2012-06-11*, 2007.

[2] Aldasoro, C. C. R.; Bhalerao, A. The bhattacharyya space for feature selection and its application to texture segmentation. *Pattern Recognition Society*, 39:812–26, 2006.

[3] ALERT Life Sciences Computing. [Online]. *http://www.alert-online.com/pt/medical-guide/melanoma*.

[4] Annessi, G.; Bono, R.; Sampogna, F.; et al. Sensitivity, specificity, and diagnostic accuracy of three dermoscopic algorithmic methods in the diagnosis of doubtful melanocytic lesions: the importance of light brown structureless areas in differentiating atypical melanocytic nevi from thin melanomas. *Journal of American Academy of Dermatology*, 56(5):759–67, 2007.

[5] Anwer R. M.; Vázquez D.; López A. M. Opponent colors for human detection. *J. Vitrià; J.M. Sanches; M. Hernández (Eds.): IbPRIA, Lecture Notes on Computer Science 6669. Springer-Verlag Berlin Heidelberg*, pages 363–70, 2011.

[6] Argenziano, G.; Fabbrocini, G.; Carli, P.; et al. Epiluminescence microscopy: Criteria of cutaneous melanoma progression. *Journal of American Academy of Dermatology*, 37(1):68–74, 1997.

[7] Argenziano, G.; Fabbrocini, G.; Carli, P.; et al. Epiluminescence microscopy for the diagnosis of doubtful melanocytic skin lesions. comparison of the abcd rule of dermatoscopy and a new 7-point checklist based on pattern analysis. *Archives of Dermatology*, 134(12):1563–70, 1998.

[8] Argenziano, G.; Ferrara, G.; Francione S.; et al. Dermoscopy: the ultimate tool for melanoma diagnosis. *Seminars in Cutaneous Medicine and Surgery*, 28(3):142–8, 2009.

[9] Baldi, A.; Quartulli, M.; Murace, R.; et al. Automated dermoscopy image analysis of pigmented skin lesions. *Cancers*, 2(262-73), 2010.

[10] Binder, M.; Schwarz, M.; Winkler, A.; et al. Epiluminescence microscopy: A useful tool for the diagnosis of pigmented skin lesions for formally trained dermatologist. *Archives of Dermatology*, 131(3):286–91, 1995.

[11] Bishop, C. M. *Neural Networks for Pattern Recognition*. Oxford, University Press, 1995.

[12] Blum, A.; Luedtke, H.; Ellwanger, U.; et al. Digital image analysis for diagnosis of cutaneous melanoma: development of a highly effective computer agorithm based on analysis of 837 melanocytic lesions. *British Journal of Dermatology*, 151(5):1029–38, 2004.

[13] Borer S.; Süsstrunk, S. Opponent color space motivated by retinal processing. *Proceedings IST First European Conference on Color in Graphics, Imaging and Vision (CGIV)*, 1(187-9), 2002.

[14] Braun, R. P.; French, L. E.; Saurat, J. H. Dermoscopy of pigmented lesions: a valuable tool in the diagnosis of melanoma. *Swiss Medical Weekly*, 134:83–90, 2004.

[15] Braun, R. P.; Rabinovitz, H. S.; Oliviero, M.; et al. Pattern analysis: a two-step procedure for the dermoscopic diagnosis of melanoma. *Clinic of Dermatology*, 20(3):236–39, 2002.

[16] Braun, R. P.; Rabinovitz, H. S.; Oliviero, M.; et al. Dermoscopy of pigmented skin lesions. *Journal of American Academy of Dermatology*, 52(1):109–21, 2005.

[17] Burroni, M.; Corona, R.; Dell'Eva, G.; et al. Melanoma computer-aided diagnosis: Reliability and feasibility study. *American Association for Cancer Research*, 10(1881-6), 2004.

[18] Celebi, M. E.; Iyatomi, H.; Stoecker, W. V.; et al. Automatic detection of blue-white and related structures in dermoscopy images. *Computerized Medical Imaging and Graphics*, 32(670-7), 2008.

[19] Celebi, M. E.; Kingravi, H. A.; Uddin, B.; et al. A methodological approach to the classification of dermoscopy images. *Computarized Medical Imaging and Graphics*, 31(362-73), 2007.

[20] Checovich, B.; Puchalski, A. A non-invasive imaging device for the early detection of skin cancer. Technical report, Coulter Translational Partnership, University of Winsconsin, 2010.

[21] de Giorgi, V.; Trez, E.; Salvini, C.; Duquia, R.; de Villa, D.; Sestini, S.; Gervini, R.; Lotti, T. Dermoscopy in black people. *Clinical and Laboratory Investigations*, 155:695–9, 2006.

[22] Dermocopy.org. [Online]. http://dermoscopy.org/.

[23] Deserno, T. M. *Biomedical Image Processing - Biological and Medical Physics, Biomedical Engineering*. Springer, 2011.

[24] Duda, R. O.; Hart, P. E.; Stork, D. G. *Pattern Classification*. Wiley-Interscience, 2nd edition, 2001.

[25] Elter, M.; Horsch, A. Cadx of mammographic masses and clustered microcalcifications: a review. *Medical Physics*, 36(6):2052–68, 2009.

[26] English, D. R.; Mar, C. D.; Burton, R. C. Factors influencing the number needed to excise: excision rates of pigmented lesions by general practitioners. *Medical Journal of Australia*, 180(1):16–9, 2004.

[27] Ferreira, P. M., Mendonça, T.; Rocha, P.; et al. A new interface for manual segmentation of dermoscopic images. In *Proceedings of the III ECCOMAS Thematic Conference on Computational Vision and Medical Image Processing: VipIMAGE 2011*, October 2011.

[28] Ganster, H.; Pinz, A.; Röhrer, R.; et al. Automated melanoma recognition. *IEEE Transactions on Medical Imaging*, 20(233-9), 2001.

[29] Garbe, C.; Eigentler, T. K. Diagnosis and treatment of cutaneous melanoma: state of the art 2006. *Melanoma Research*, 17(2):117–27, 2007.

[30] Garbe, C.; Leiter, U. Melanoma epidemiology and trends. *Clinics in Dermatology*, 27(1):3–9, 2009.

[31] Gerger, A.; Pompl, R.; Smolle, J. et al. Automated epiluminescence microscopy: tissue counter analysis using cart and 1-nn in the diagnosis of melanoma. *Skin Research Technology*, 9(2):105–10, May 2003.

[32] Gonzalez, R. C.; Woods, R. E.; Eddins, S. L. *Digital Image Analysis using MATLAB*. Pearson, 2009.

[33] Hansen, C.; Wilkinson, R. P.; Hansen M.; et al. How good are skin cancer clinics at melanoma detection? number needed to treat variability across a national clinic group in australia. *Journal of American Academy in Dermatology*, 61(4):599–604, 2009.

[34] Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning - Data Mining, Inference and Prediction*. Springer-Verlag, 2001.

[35] Hoffmann, K.; Gambichler, T.; Rick, A.; et al. Diagnostic and neural analysis of skin cancer (danaos). a multicentre study for collection and computer-aided analysis of data from pigmented skin lesions using digital dermoscopy. *British Journal of Dermatology*, 149(4):801–9, 2003.

[36] Horsch, A.; Blank R.; Eigenman, D. The efmi reference image database initiative: concept, state and related work. *Proceedings of Computer Assisted Radiology and Surgery*, pages 447–52, 2005.

[37] Horsch, A.; Stolz, W.; Neiss, A.; et al. Improving early recognition of malignant melanomas by digital image analysis in dermatology. *Studies in health technology and informatics*, 43 Pt B:351–35, 1997.

[38] Iyatomi, H.; Oka, H.; Celebi, M. E.; et al. Parameterization of dermoscopic findings for the internet-based melanoma screening system. *Proceedings of IEEE Symposium on Computational Intelligence in Image and Signal Processing (CIISP)*, pages 189–93, April 2007.

[39] Iyatomi, H.; Oka, H.; Celebi, M. E.; et al. An improved internet-based melanoma screening with dermatologist-like tumor area extraction algorithm. *Computerized Medical Imaging and Graphics*, 32(566-79), 2008.

[40] Joel, G.; Schmid-Saugeon, P.; Guggisberg, D.; et al. Validation of segmentation techniques for digital dermoscopy. *Skin Research Technology*, 8(4):240–9, 2002.

[41] Kailath, T. The divergence and bhattacharyya distance measures in signal decision. *IEEE Transactions in Communication Technology*, 15(1):52–60, 1967.

[42] Kittler, H.; Pehamberger, H.; Wolff, K.; Binder, M. Diagnostic of dermoscopy. *The Lancelot Oncology*, 3(159-65), March 2002.

[43] Kolmogorov, A. Sulla determinazione empirica di una legge di distributione. *Giornale dell Instituto Italiano degli Attuari*, 4(83-91), 1993.

[44] Lakhani, S. Early clinical pathologists 4: John hunter. *Journal of Clinical Pathology*, 44(621-3), 1991.

[45] Lamarão, P.; Cabral, H. C. *The Menzies algorithm and a proposal online screening with support for acral lesions.* Meeting of Grupo Português de Dermatologia, February 2011.

[46] Liga Portuguesa Contra o Cancro. [Online]. http://www.ligacontracancro.pt/.

[47] Liu, H.; Motoda, H. *Computational Methods of Feature Selection.* Taylor and Francis Group, 2008.

[48] Losina, E.; Walensky, R. P.; Geller, A.; et al. Visual screening for malignant melanoma: a cost-effectiveness analysis. *Archives of Dermatology*, 143(1):21–8, 2007.

[49] Marçal, A. R. S.; Silva, C. S. P.; Mendonça, T.; et al. Evaluation of the menzies method potential for automatic dermoscopic image analysis. *CompImage - Proceedings on Computational of Objects Presented in Images: Fundamentals, Methods and Applications. 3rd ed. (in press)*, 2012.

[50] Marghboob, A. A.; und Alfred, W.; Kopf, R. B. *An atlas of dermoscopy.* Encyclopedia of Visual Medicine Series. London: Taylor and Francis, 2004.

[51] Masoomeh, B.; Ghaninezhad, H.; Mansoori, P.; et al. Computer-aided dermoscopy for diagnosis of melanoma. *BMC Dermatology*, 5(8):1–4, July 2005.

[52] MedicineWorld.Org. [Online]. *http://medicineworld.org.*

[53] Menzies, S. W.; Ingvar, C.; Crotty, K. A.; et al. Frequency and morphologic characteristics of invasive melanomas lacking specific surface microscopic features. *Archives of Dermatology*, 132(1178-82), 1996.

[54] Nachbar, F.; Stolz, W.; Merkle, T.; et al. The abcd rule of dermatology. high prospective value in the diagnosis of doubtful melanocytic skin lesions. *Journal of American Academy of Dermatology*, 30(4):551–9, 1994.

[55] Orborne, J. E.; Bourke, J. F.; Graham-Brown, R. A. C.; Hutchinson, P. E. False positive clinical diagnoses of malignant melanoma. *British Journal of Dermatology*, 140(5):902–8, May 1999.

[56] Otsu, N. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man and Cybernetics*, 9(1):62–6, January 1979.

[57] Pehamberger, H.; Binder, M.; Steiner, A.; et al. In vivo epiluminescence microscopy: improvement of early diagnosis of melanoma. *Journal of Investigative Deramatology*, 100(356S-62S), 1993.

[58] Pehamberger, H.; Steiner, A.; Wolff, K. In vivo epiluminescence microscopy of pigmented skin lesions. i. pattern analysis of pigmented skin lesions. *Journal of American Academy of Dermatology*, 17(4):571–83, 1987.

[59] Polikar, R.; Jahan, K.; Healy, B. A combined pattern separability and two-tiered classification approach for identification of binary mixtures of vocs. *Sensors and Actuators B*, 116(174-82), 2006.

[60] Polikar, R.; Udpa, L.; Udpa, S. S. Nonlinear cluster transformation for increasing pattern separability. *Conference on Neural Networks. IJCNN '99. International Joint.*, 6:4006–11, 1999.

[61] Press, W. H.; Teukolsky, S. A.; Vetterling, W. T.; Flannery, B. P. *Numerical Recipes: The Art of Scientific Computing.* New York: Cambridge University Press, 3rd edition, 2007.

[62] [Psaty, E. L.; Halpern, A. C. Current and emerging technologies in melanoma diagnosis: the state of the art. *Clinic of Dermatology*, 27(1):35–45, 2009.

[63] Pusey, W. A. The history of dermatology. *American Journal of Public Health*, 23(4):394–5, April 1933.

[64] Rajpara, S. M.; Botello, A. P.; Townend, J.; et al. Systematic review of dermoscopy and digital dermoscopy: artificial intelligence for the diagnosis of melanoma. *British Journal of Dermatology*, 161(3):591–604, 2009.

[65] Rallan, D.; Harland, C. C. Skin imaging: is it clinically useful? *Clinical and Experimental Dermatology*, 29(5):453–9, 2004.

[66] Richards, J. A.; Jia, X. *Remote Sensing Digital Image Analysis: An introduction*. Springer, 4th edition, 2006.

[67] Rosado, B.; Menzies, S.; Harbauer, A.; et al. Accuracy of computer diagnosis of melanoma: a quantitative meta-analysis. *Archives of Dermatology*, 159(3):361–7, 2003. discussion 366.

[68] Rubegni, P.; Cevenini, G.; Burroni, M; et al. Automated diagnosis of pigmented skin lesions. *International Journal of Cancer*, 101:576–80, 2002.

[69] Ruiz, D.; Berenguer, V.; Soriano, A.; et al. A decision support system for the diagnosis of melanoma: a comparative approach. *Expert Systems and Applications*, 38(12):15217–23, November-December 2011.

[70] Scott, H.; Dusza, S. W.; Wang, S. Q.; et al. The cash algorithm (color, arquitecture, symmetry, and homogeneity) algorithm for dermoscopy. *Journal of American Academy of Dermatology*, 56(45-52), 2007.

[71] Shapiro, L. G.; Stockman, G. C. *Computer Vision*. New Jersey, Prentice-Hall, 2001.

[72] Shin, M. C.; Chang, K. I.; Tsap, L. V. Does colorspace transformation make any difference on skin detection? *Proceedings of Sixth IEEE Workshop on Aplications of Computer Vision*, pages 275–9, 2002.

[73] Silva, C. S. P.; Marcal, A. R. S.; Pereira, M. A.; et al. Separability analysis of color classes on dermoscopic images. *A. Campilho, M. Kamel (eds): ICIAR 2012, Part II, Lecture Notes on Computer Science 7325. Springer-Verlag Berlin Heidelberg*, pages 268–77, 2012.

[74] Silveira, M.; Nascimento, J. C.; Marques, J.; et al. Comparison of segmentation methods for melanoma diagnosis in dermoscopy images. *IEEE Journal of Selected Topics in Signal Processing*, 3(1):35–45.

[75] SkinCancerNet. [Online]. *http://www.skincarephysicians.com/skincancernet/glossary.html*.

[76] Sonka, M.; Fitzpatrick, J. M. *Handbook of Medical Imaging. Medical Image Processing and Analysis*, volume 2. SPIE Press, 2000.

[77] Soyer, H. P.; Argenziano, G.; Zalaudek, I.; et al. Three-point checklist of dermoscopy. a new screening method for early detection of melanoma. *Dermatology*, 208(27-31), 2004.

[78] Stolz, W.; Braun-Falco, O.; Bilek, P.; et al. *Color Atlas of Dermatology.* London: Wiley-Blackwell, 2nd edition, 2002.

[79] Stolz, W.; Riemann, A.; Cognetta, A. B.; et al. Abcd rule of dermatoscopy: a new practical method for early recognition of malignant melanoma. *European Journal of Dermatology*, 4(521-7), 1994.

[80] Theodoridis, S.; Koutroumbas, K. *Pattern Recognition.* Academic Press, Elsevier, 4th edition, 2009.

[81] Torkkola, K. Feature extraction by non-parametric mutual information maximization. *Journal of Machine Learning Research*, 3:1415–38, 2003.

[82] Vestergaad, M. E.; Macaskill, P.; Holt P. E.; et al. Dermoscopy compared with naked eye examination for the diagnosis of primary melanoma: a meta-analysis of studies performed in a clinical setting. *British Journal of Dermatology*, 159(3):669–76, 2008.

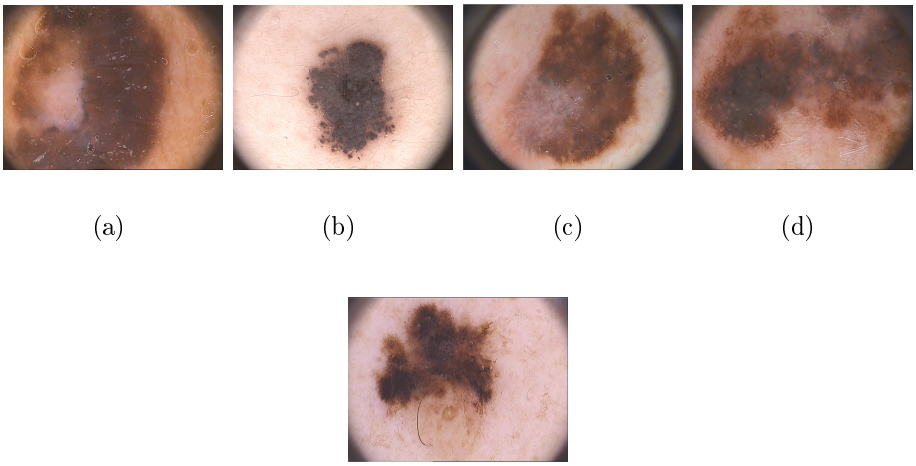
[83] Wang, S. Q.; Rabinovitz H.; Kopf, A W.; et al. Current technologies for the in vivo diagnosis of cutaneous melanoma. *Clinic of Dermatology*, 22(3):217–22, 2004.

[84] Wirth, M.; Fraschini, M.; Masek, M.; et al. Performance evaluation in image processing. *EURASIP Journal on Advances in Signal Processing*, pages 1–4, 2006.

Appendix A

Dataset of Dermoscopy Images

The dataset used is represented by 28 melanocytic dermoscopic images: 5 melanomas (Figure A.1) and 23 *nevus* (Figure A.2).



(a) (b) (c) (d)

(e)

Fig. A.1: Pigemented skin lesions of melanocytic melanomas. IMD: (a) 17, (b) 18, (c) 19, (d) 20, and (e) 28.

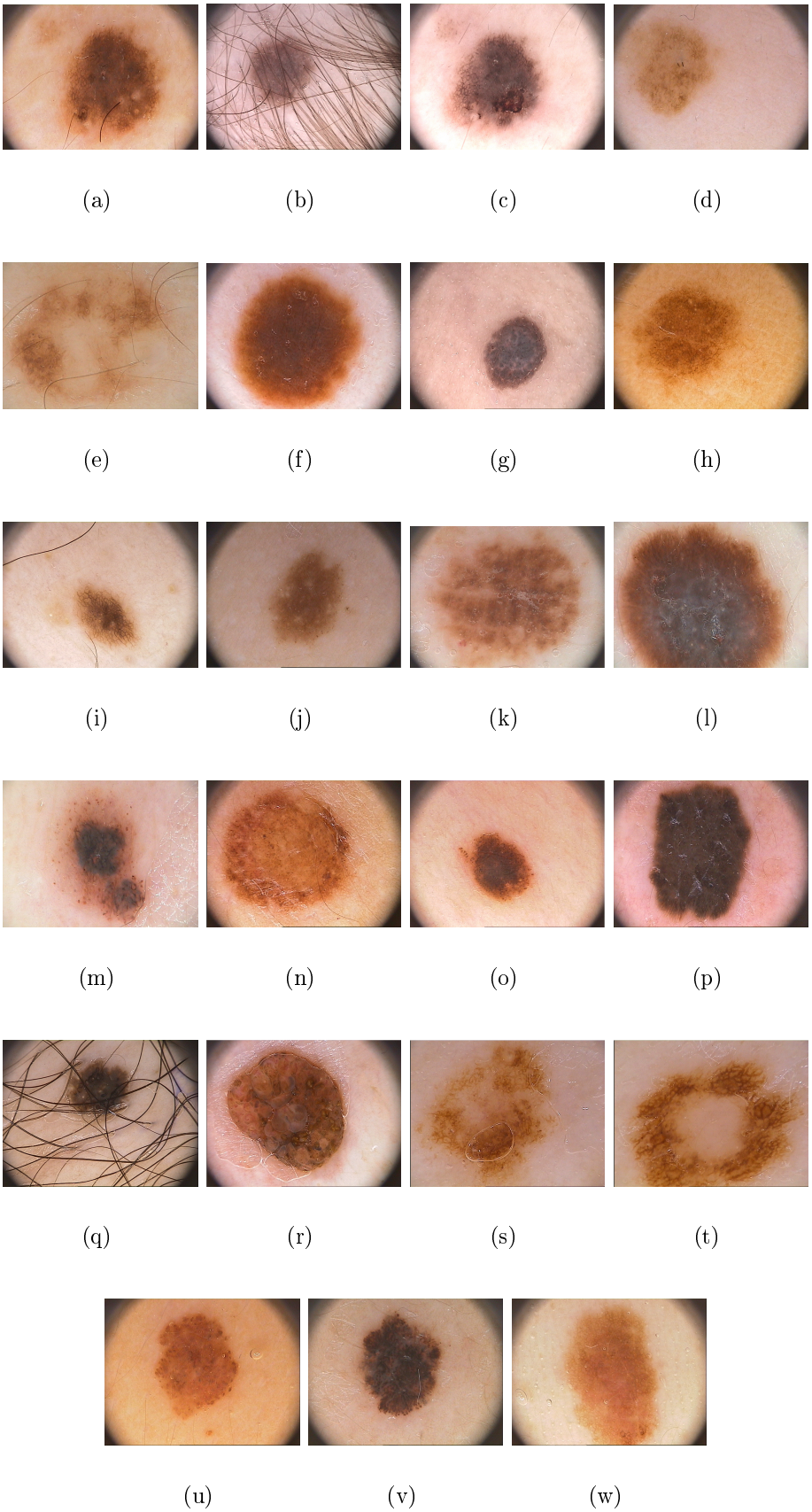


Fig. A.2: Pigmented skin lesions of melanocytic nevus. IMD: (a) 01, (b) 02, (c) 03, (d) 04, (e) 05, (f) 06, (g) 07, (h) 08, (i) 09, (j) 10, (k) 11, (l) 12, (m) 13, (n) 14, (o) 15, (p) 16, (q) 21, (r) 22, (s) 23, (t) 24, (u) 25, (v) 26 and (w) 27.

Appendix B

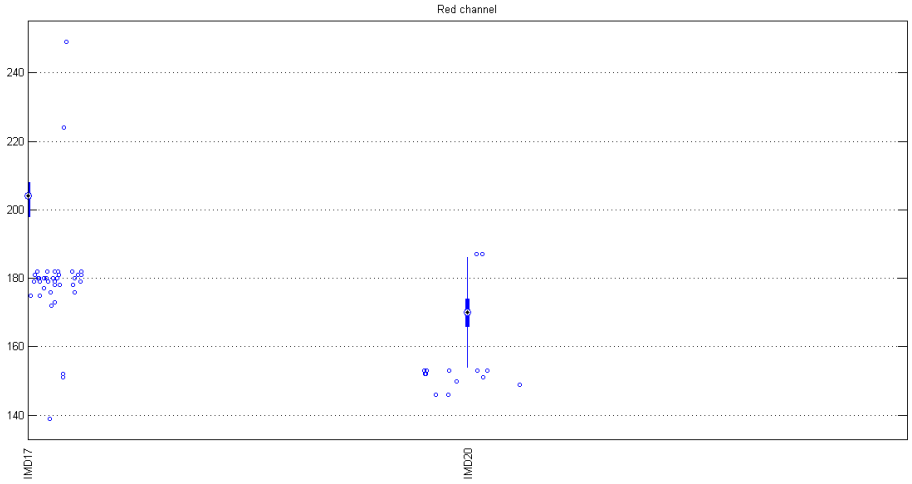
Intra-Classes Tests

In the original RGB color space, when we pass from the SIT group to the MIT group, the overlap between color classes is highly evident. This is due to the large spectral range of each color class. The following figures (B.1-B.15) show the boxplot graphs for each color class, in each color channel (Red, Green and Blue) that have more than one sample image (Wh, LB, DB, BG and Bk), reporting qualitatively the difference in the spectral ranges in different images.

The presence of each color class can be seen in Table 4.2 and, summary, the number of its appearance of each color in the dataset images can be seen in Table 4.2.

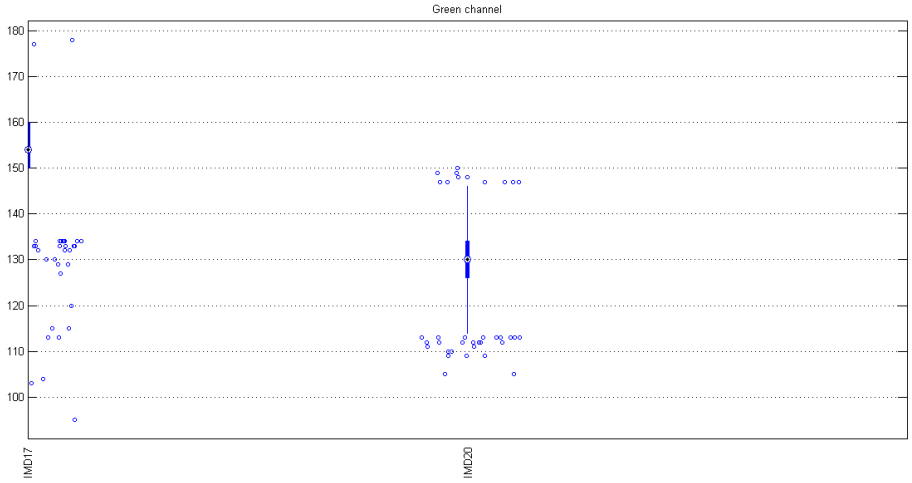
For the Wh (White), Re (Red), LB (Light Brown), DB (Dark Brown) and BG (Blue Gray) color classes, as reported by the boxplot graphs in Figures B.1-B.12, for the Red, Green and Blue color channels, do not exists "a band" of spectral range values, which causes the overlapping in the RGB feature space.

For the Bk (Black) color class, in Figures B.13-B.15 the boxplot graphs in the Red, Green and Blue channel, respectively, are shown. For this color class it is evident "a band" of spectral range values, i.e., the Bk color has a low rate of variability in all of its appearances and that it is a good expected behavior.



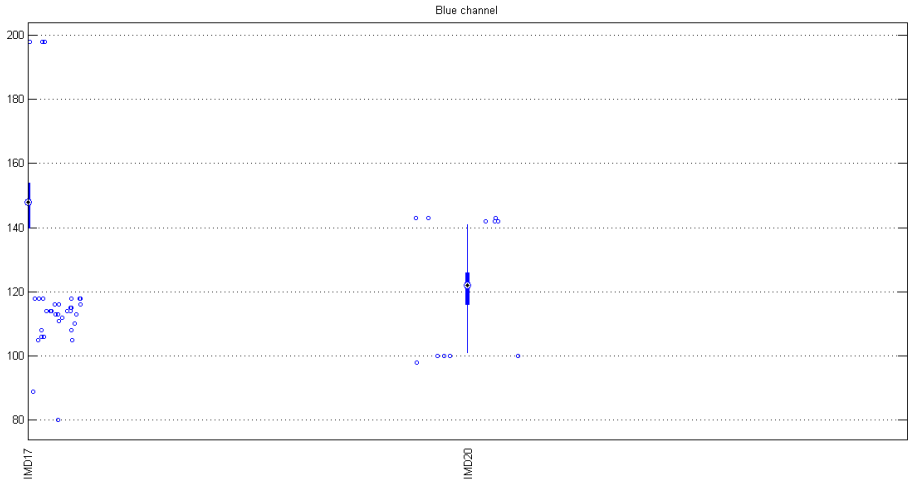
(a)

Fig. B.1: Boxplot for subset of images that have the Wh (White) color class: IMD17 and IMD20, in the Red color channel of RGB feature space.



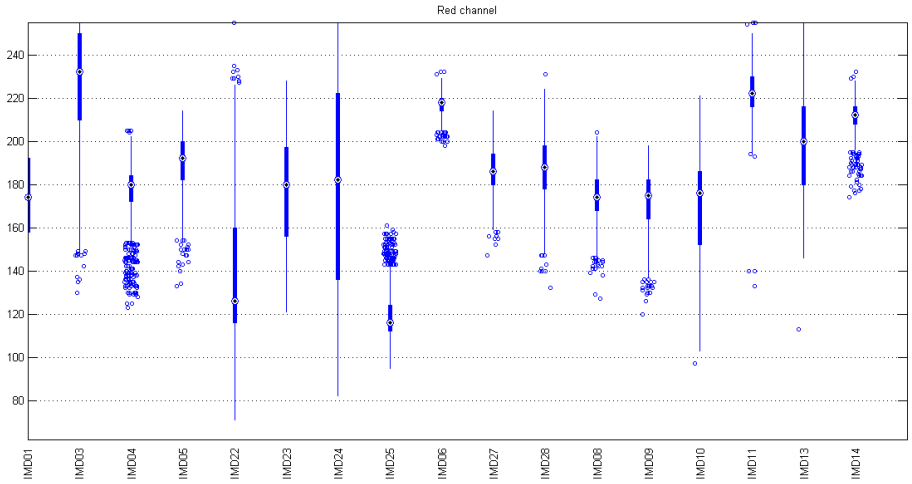
(a)

Fig. B.2: Boxplot for subset of images that have the Wh (White) color class: IMD17 and IMD20, in the Green color channel of RGB feature space.



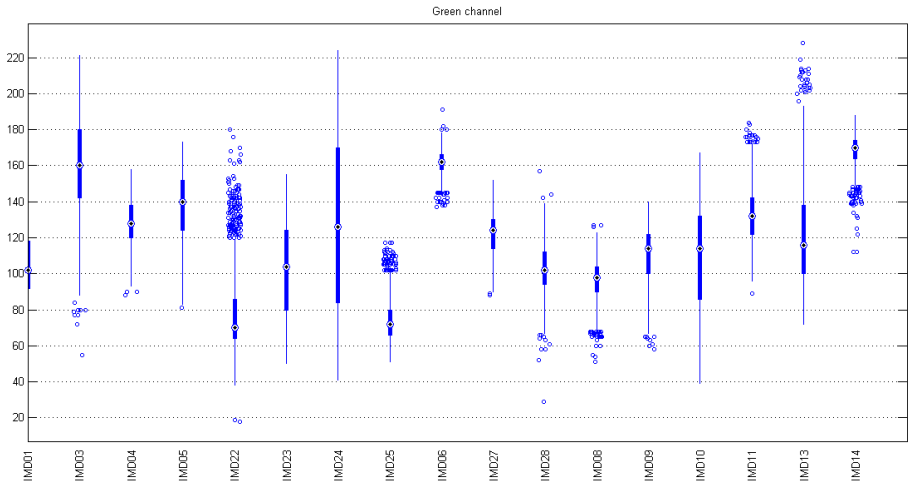
(a)

Fig. B.3: Boxplot for subset of images that have the Wh (White) color class: IMD17 and IMD20, in the Blue color channel of RGB feature space.



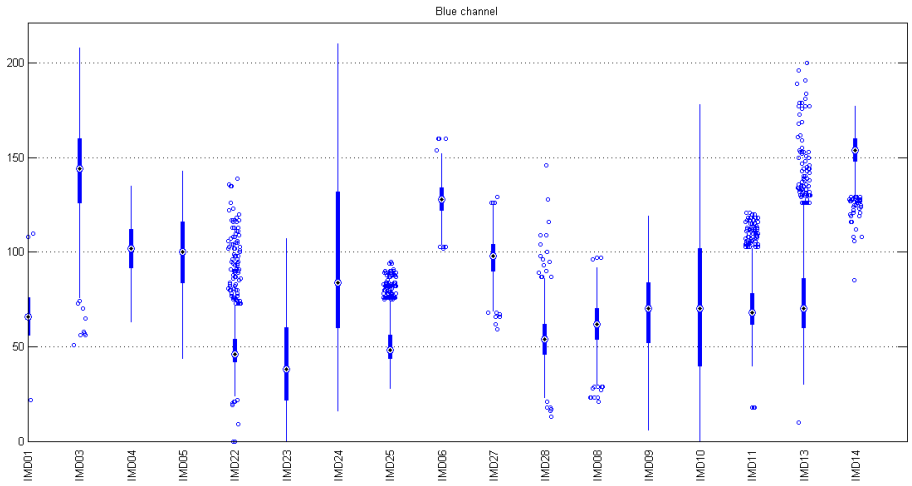
(a)

Fig. B.4: Boxplot for subset of images that have the LB (Light Brown) color class: IMD01, IMD03, IMD04, IMD05, IMD06, IMD08, IMD09, IMD10, IMD11, IMD13, IMD14, IMD22, IMD23, IMD24, IMD25, IMD27 and IMD28 in the Red color channel of RGB feature space.



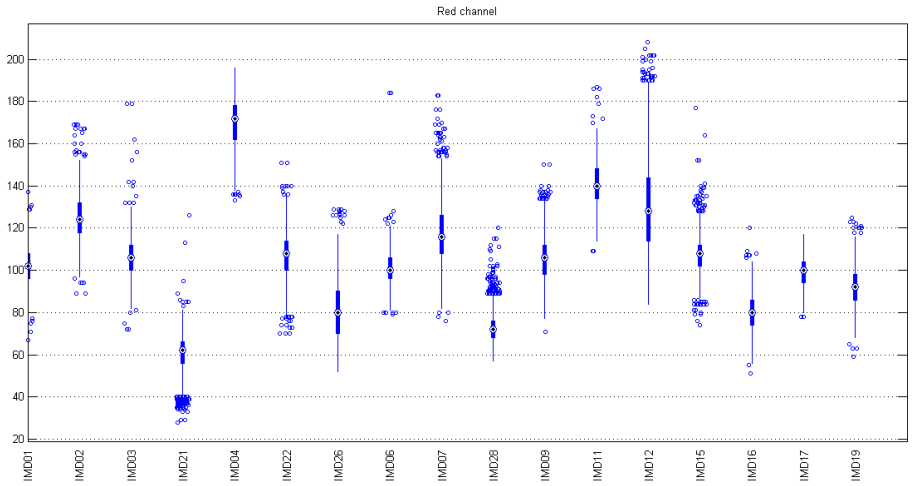
(a)

Fig. B.5: Boxplot for subset of images that have the LB (Light Brown) color class: IMD01, IMD03, IMD04, IMD05, IMD06, IMD08, IMD09, IMD10, IMD11, IMD13, IMD14, IMD22, IMD23, IMD24, IMD25, IMD27 and IMD28 in the Green color channel of RGB feature space.



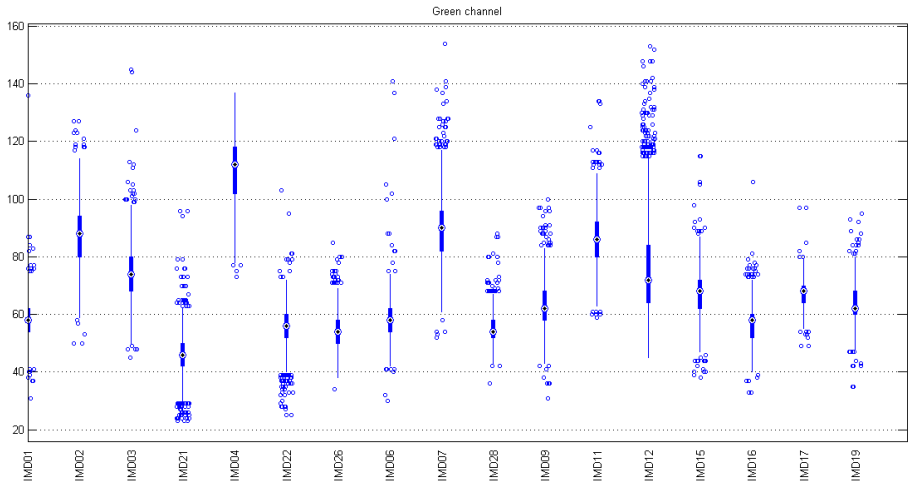
(a)

Fig. B.6: Boxplot for subset of images that have the LB (Light Brown) color class: IMD01, IMD03, IMD04, IMD05, IMD06, IMD08, IMD09, IMD10, IMD11, IMD13, IMD14, IMD22, IMD23, IMD24, IMD25, IMD27 and IMD28 in the Blue color channel of RGB feature space.



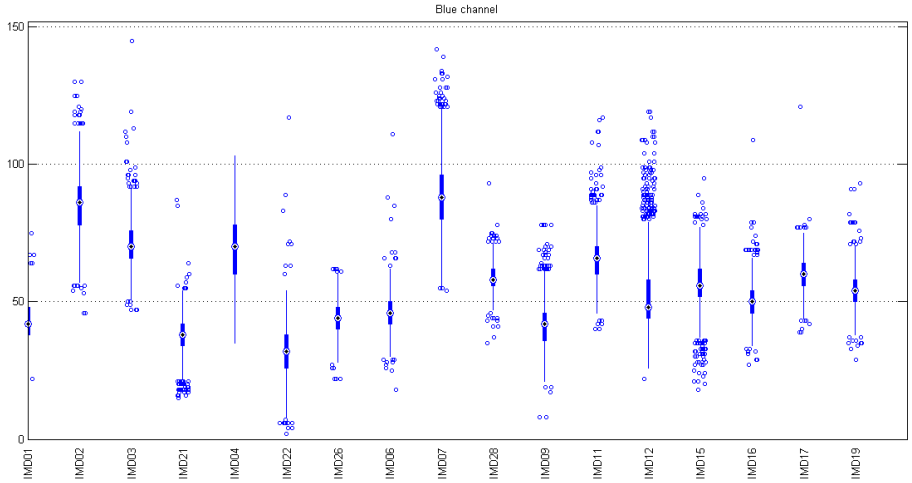
(a)

Fig. B.7: Boxplot for subset of images that have the DB (Dark Brown) color class: IMD01, IMD02, IMD03, IMD04, IMD06, IMD07, IMD09, IMD11, IMD12, IMD15, IMD16, IMD17, IMD19, IMD21, IMD22, IMD26 and IMD28 in the Red color channel of RGB feature space.



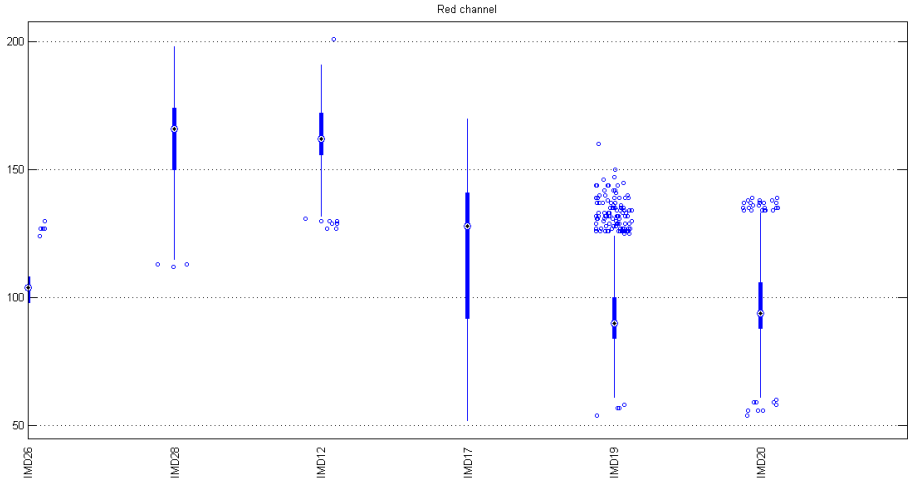
(a)

Fig. B.8: Boxplot for subset of images that have the DB (Dark Brown) color class: IMD01, IMD02, IMD03, IMD04, IMD06, IMD07, IMD09, IMD11, IMD12, IMD15, IMD16, IMD17, IMD19, IMD21, IMD22, IMD26 and IMD28 in the Green color channel of RGB feature space.



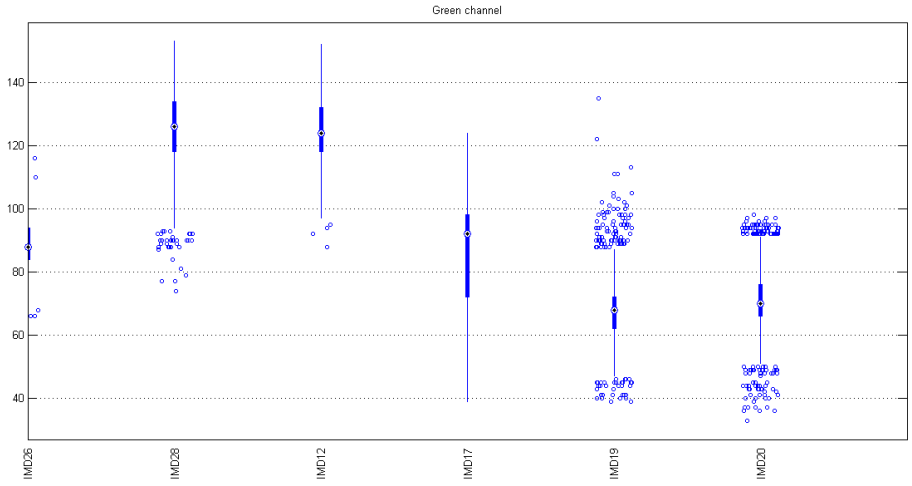
(a)

Fig. B.9: Boxplot for subset of images that have the DB (Dark Brown) color class: IMD01, IMD02, IMD03, IMD04, IMD06, IMD07, IMD09, IMD11, IMD12, IMD15, IMD16, IMD17, IMD19, IMD21, IMD22, IMD26 and IMD28 in the Blue color channel of RGB feature space.



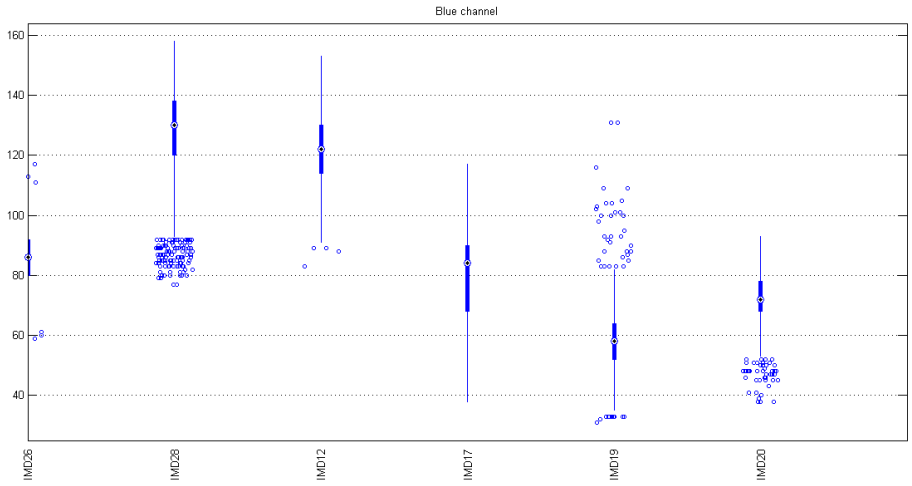
(a)

Fig. B.10: Boxplot for subset of images that have the BG (Blue Gray) color class: IMD12, IMD17, IMD19, IMD20, IMD26 and IMD28 in the Red color channel of RGB feature space.



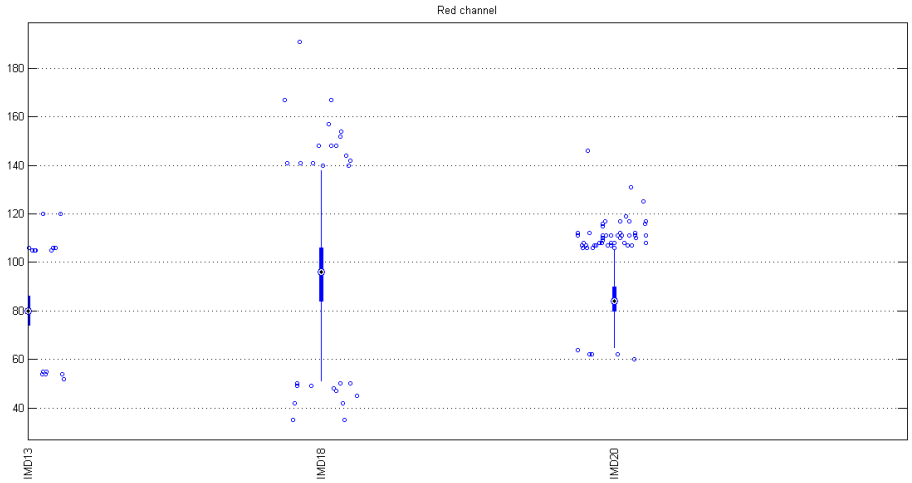
(a)

Fig. B.11: Boxplot for subset of images that have the BG (Blue Gray) color class: IMD12, IMD17, IMD19, IMD20, IMD26 and IMD28 in the Green color channel of RGB feature space.



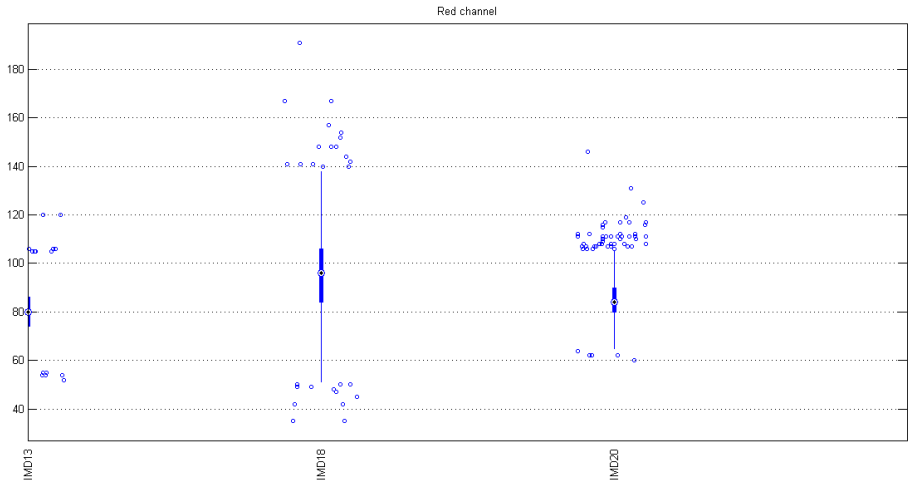
(a)

Fig. B.12: Boxplot for subset of images that have the BG (Blue Gray) color class: IMD12, IMD17, IMD19, IMD20, IMD26 and IMD28 in the Blue color channel of RGB feature space.



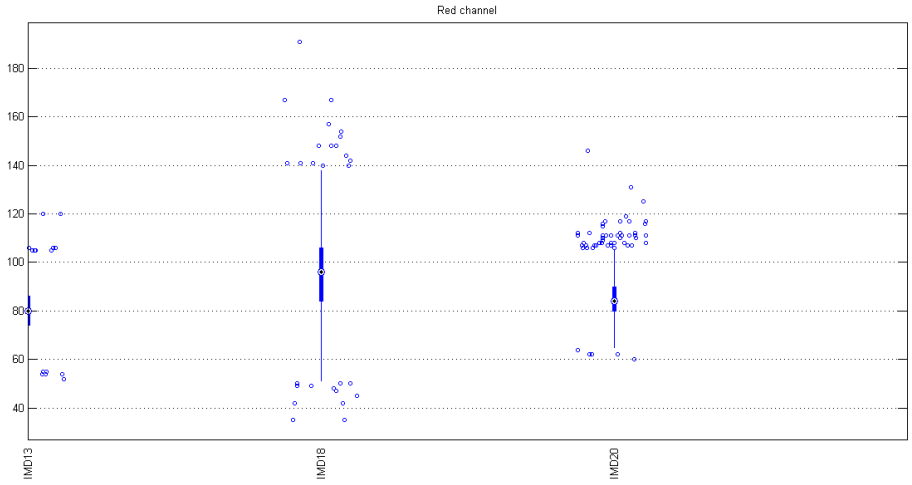
(a)

Fig. B.13: Boxplot for subset of images that have the Bk (Black) color class: IMD13, IMD18 and IMD20 in the Red color channel of RGB feature space.



(a)

Fig. B.14: Boxplot for subset of images that have the Bk (Black) color class: IMD13, IMD18 and IMD20 in the Green color channel of RGB feature space.



(a)

Fig. B.15: Boxplot for subset of images that have the Bk (Black) color class: IMD13, IMD18 and IMD20 in the Blue color channel of RGB feature space.

Appendix C

Color Analysis

The JM and TD separability measures are presented in Tables C.1, C.2 and C.3 for the HSI, OPP and XYI color spaces, respectively. These tables must be compared with Table 5.3 which present the JM and TD measures for the MIT group in the RGB feature space.

For the RGB feature space was guaranteed that 9 out of 15 pairs of color classes are separable. For the HSI color space, 10 out of 15 pairs are considered separable; for the OPP color space, 8 out of 15 pairs are considered separable; and for the XYI color space also 8 out of 15 pairs are considered separable. The treatment in the OPP and XYI color spaces do not present significant improvements. In the HSI color space there were an improvement of the separability of one more pair of color classes. However, this slight improvement do not compensate the computational effort. For this reason was chosen the RGB feature space.

Tab. C.1: Jeffries-Matusita - JM (top right triangular matrix) and Transformed Divergence - TD (bottom left) distance values, for the multiple image training using the HSI color space.

Classes	Wh	Re	LB	DB	BG	Bk
Wh	-	2.00	1.88	1.80	1.21	1.91
Re	(*)	-	1.14	1.97	2.00	2.00
LB	(*)	(*)	-	1.81	1.95	2.00
DB	1.65	(*)	(*)	-	1.03	1.05
BG	0.74	(*)	(*)	0.47	-	0.53
Bk	1.89	(*)	(*)	0.75	0.18	-

(*) not possible to calculate

Wh - White, Re - Red, LB - Light Brown,

DB - Dark Brown, BG - Blue Gray, Bk - Black

Tab. C.2: Jeffries-Matusita - JM (top right triangular matrix) and Transformed Divergence - TD (bottom left) distance values, for the multiple image training using the Opponent color space.

Classes	Wh	Re	LB	DB	BG	Bk
Wh	-	2.00	1.72	1.80	1.12	1.90
Re	2.00	-	1.23	1.96	2.00	2.00
LB	1.96	0.56	-	1.37	1.82	1.96
DB	1.71	2.00	1.14	-	0.96	1.11
BG	1.05	2.00	1.91	0.65	-	0.54
Bk	1.91	2.00	1.99	0.88	0.10	-

(*) not possible to calculate

Wh - White, Re - Red, LB - Light Brown,

DB - Dark Brown, BG - Blue Gray, Bk - Black

Tab. C.3: Jeffries-Matusita - JM (top right triangular matrix) and Transformed Divergence - TD (bottom left) distance values, for the multiple image training using the XYI color space.

Classes	Wh	Re	LB	DB	BG	Bk
Wh	-	2.00	1.75	1.81	1.22	1.90
Re	2.00	-	1.19	1.94	2.00	2.00
LB	2.00	(*)	-	1.63	1.90	1.99
DB	1.63	2.00	(*)	-	0.96	0.96
BG	0.71	2.00	1.57	0.57	-	0.55
Bk	1.89	2.00	1.98	0.85	0.17	-

(*) not possible to calculate

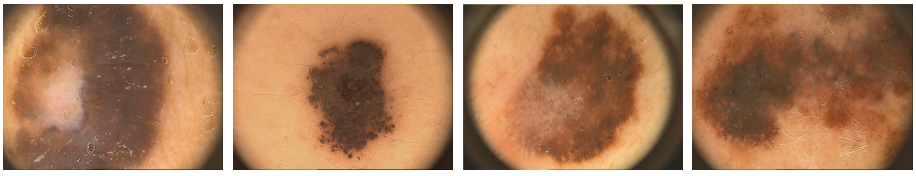
Wh - White, Re - Red, LB - Light Brown,

DB - Dark Brown, BG - Blue Gray, Bk - Black

Appendix D

Resulted Images of Image Intensity Calibration

The calibrated images resulted from image calibration technique method I (image intensity calibration) are presented in Figures D.1 and D.2, where Figure D.1 shown the 5 melanomas and Figure D.2 show the 23 *nevus* of total of 28 images in the dataset.



(a) (b) (c) (d)



(e)

Fig. D.1: Pigmented skin lesions of melanocytic melanomas resulting from Calibration method I. IMD: (a) 17, (b) 18, (c) 19, (d) 20, and (e) 28.

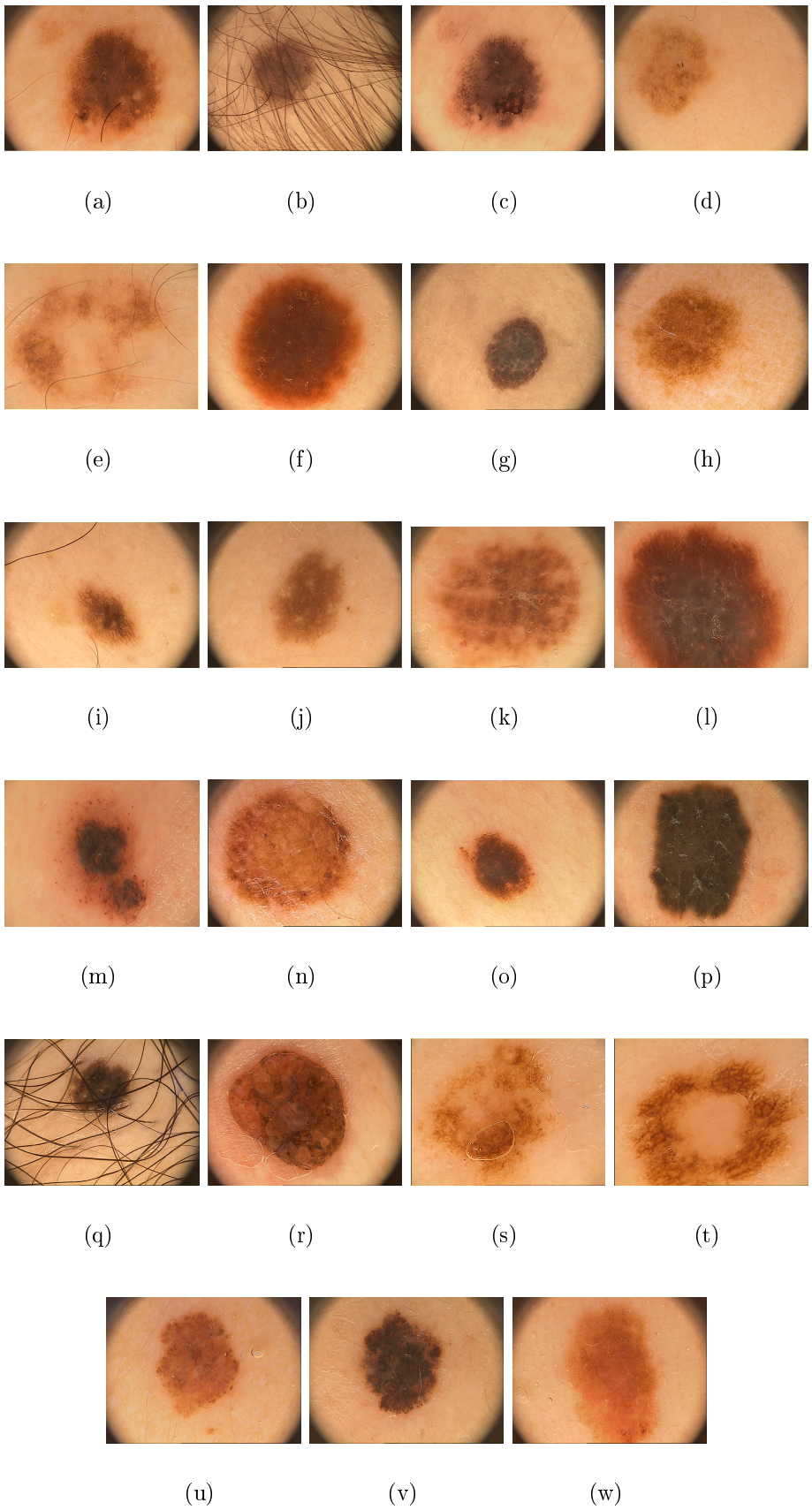
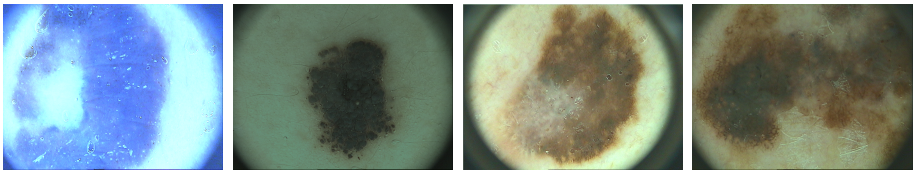


Fig. D.2: Pigemented skin lesions of melanocytic nevus resulting from Calibration technique method I. IMD: (a) 01, (b) 02, (c) 03, (d) 04, (e) 05, (f) 06, (g) 07, (h) 08, (i) 09, (j) 10, (k) 11, (l) 12, (m) 13, (n) 14, (o) 15, (p) 16, (q) 21, (r) 22, (s) 23, (t) 24, (u) 25, (v) 26 and (w) 27.

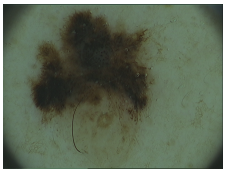
Appendix E

Resulted Images of Calibration based on Skin-Lesion Rate

The calibrated images resulted from image calibration technique method II (calibration based on skin-lesion rate) are presented in Figures E.1 and E.2, where Figure E.1 shown the 5 melanomas and Figure E.2 show the 23 *nevus* of total of 28 images in the dataset.



(a) (b) (c) (d)



(e)

Fig. E.1: Pigmented skin lesions of melanocytic melanomas resulting from Calibration method II. IMD: (a) 17, (b) 18, (c) 19, (d) 20, and (e) 28.

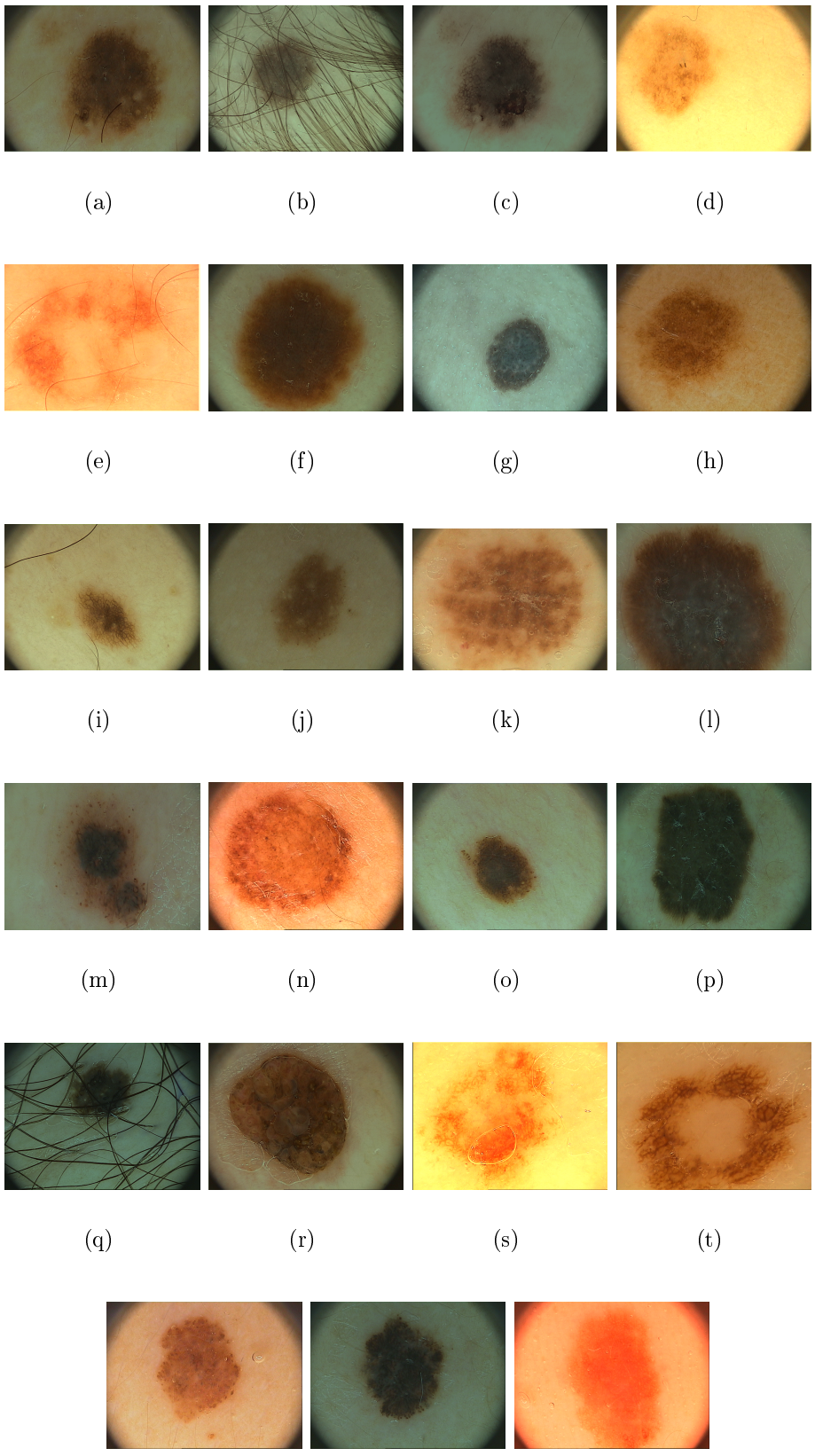
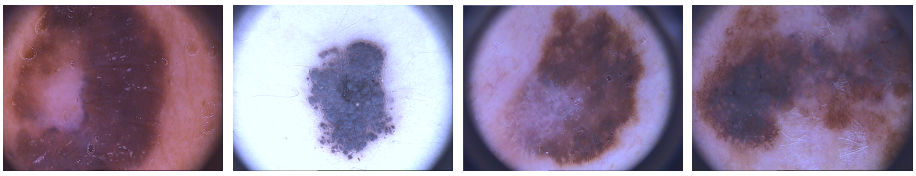


Fig. E.2: Pigemented skin lesions of melanocytic nevus resulting from Calibration technique method II. IMD: (a) 01, (b) 02, (c) 03, (d) 04, (e) 05, (f) 06, (g) 07, (h) 08, (i) 09, (j) 10, (k) 11, (l) 12, (m) 13, (n) 14, (o) 15, (p) 16, (q) 21, (r) 22, (s) 23, (t) 24, (u) 25, (v) 26 and (w) 27.

Appendix F

Resulted Images of Calibration with the Magnitude of the Reference Mean Vector Constant

The calibrated images resulted from image calibration technique method III (calibration with the magnitude of the reference mean vector constant) are presented in Figures F.1 and F.2, where Figure F.1 shown the 5 melanomas and Figure F.2 show the 23 *nevus* of total of 28 images in the dataset.



(a) (b) (c) (d)



(e)

Fig. F.1: Pigmented skin lesions of melanocytic melanomas resulting from Calibration method III. IMD: (a) 17, (b) 18, (c) 19, (d) 20, and (e) 28.

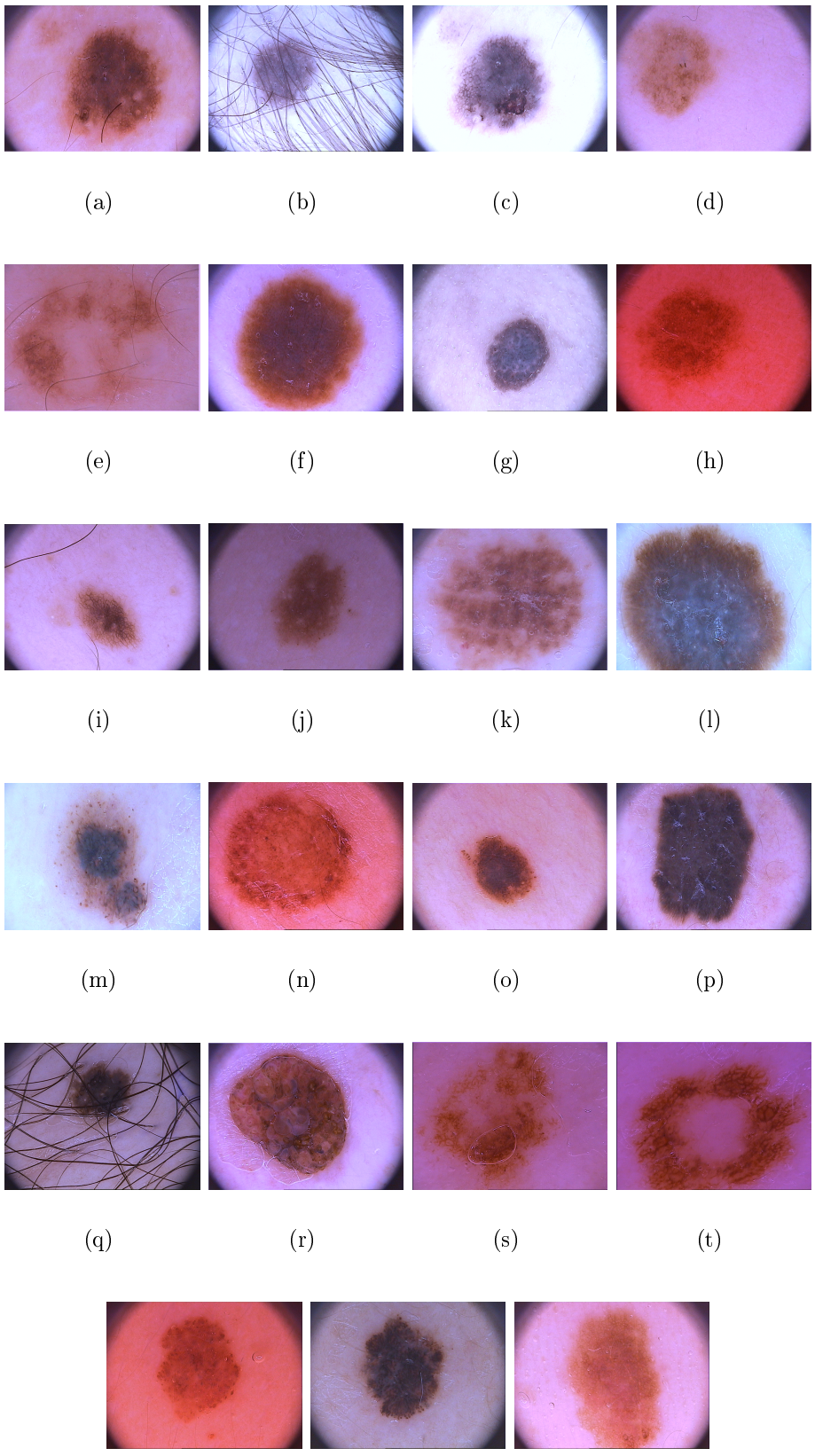


Fig. F.2: Pigemented skin lesions of melanocytic nevus resulting from Calibration technique method III. IMD: (a) 01, (b) 02, (c) 03, (d) 04, (e) 05, (f) 06, (g) 07, (h) 08, (i) 09, (j) 10, (k) 11, (l) 12, (m) 13, (n) 14, (o) 15, (p) 16, (q) 21, (r) 22, (s) 23, (t) 24, (u) 25, (v) 26 and (w) 27.

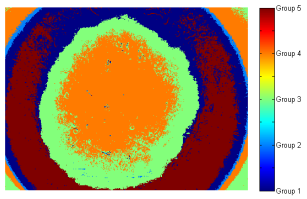
Appendix G

Classified Images

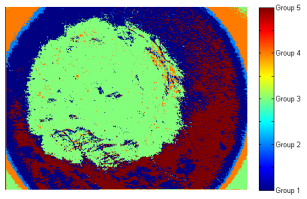
For the classification step, firstly, all the seven classes (6 color classes + skin class) were arranged in 5 groups and each of the test images were classified in 5 possible labels. The group 1 has the class Wh, group 2 the class Re, group 3 the class LB, group 4 the classes DB, BG and Bk, and group 5 the skin class. In Figure G.1 are shown the results for the 1^{st} classification step and in Figures G.2 and G.3 are shown the results for the post-processing procedures, dilation and filtering, respectively.

In the 2^{nd} classification step, we must divide the group 4 in 3 possible classes which it contains: DB, BG and Bk. To this procedure, the classification results are shown in Figure G.4 and the following post-processing procedure is shown in Figure G.5.

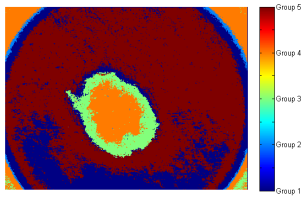
Finally, Figure G.6 show the post-processed images with black corners removed.



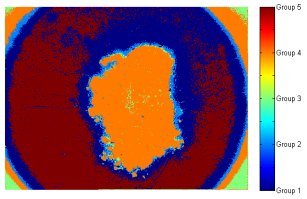
(a)



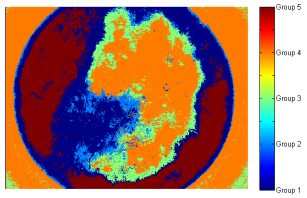
(b)



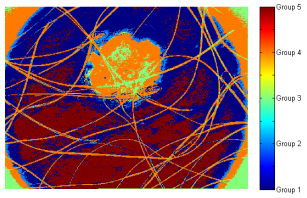
(c)



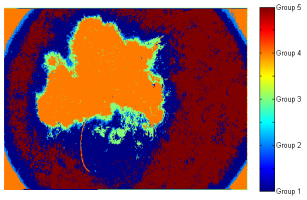
(d)



(e)

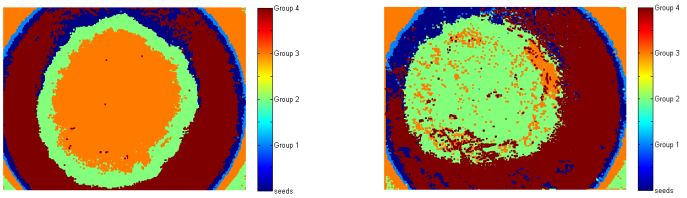


(f)



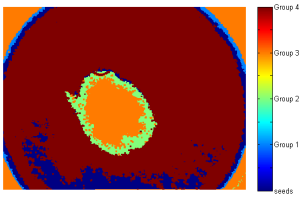
(g)

Fig. G.1: Results for neural networks in competition classification considering 5 class groups: (a) IMD06, (b) IMD14, (c) IMD15, (d) IMD18, (e) IMD19, (f) IMD21 and (g) IMD28.

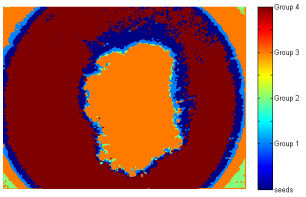


(a)

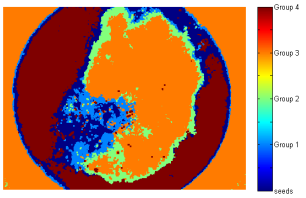
(b)



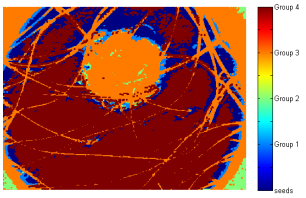
(c)



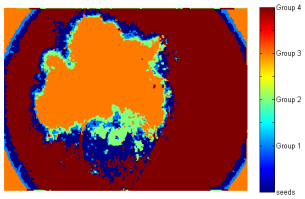
(d)



(e)

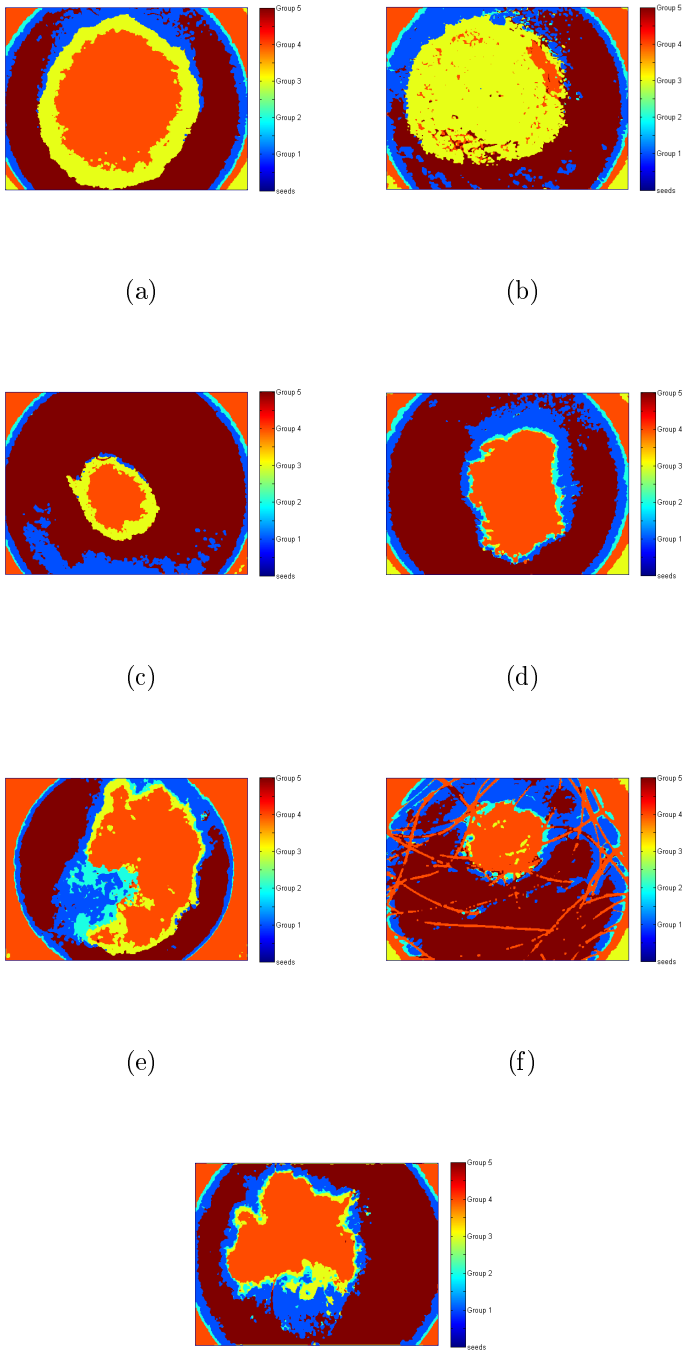


(f)



(g)

Fig. G.2: Dilation - step 1: (a) IMD06, (b) IMD14, (c) IMD15, (d) IMD18, (e) IMD19, (f) IMD21 and (g) IMD28.



(a) (b)

(c) (d)

(e) (f)

(g)

Fig. G.3: Mode filter - step 1: (a) IMD06, (b) IMD14, (c) IMD15, (d) IMD18, (e) IMD19, (f) IMD21 and (g) IMD28.

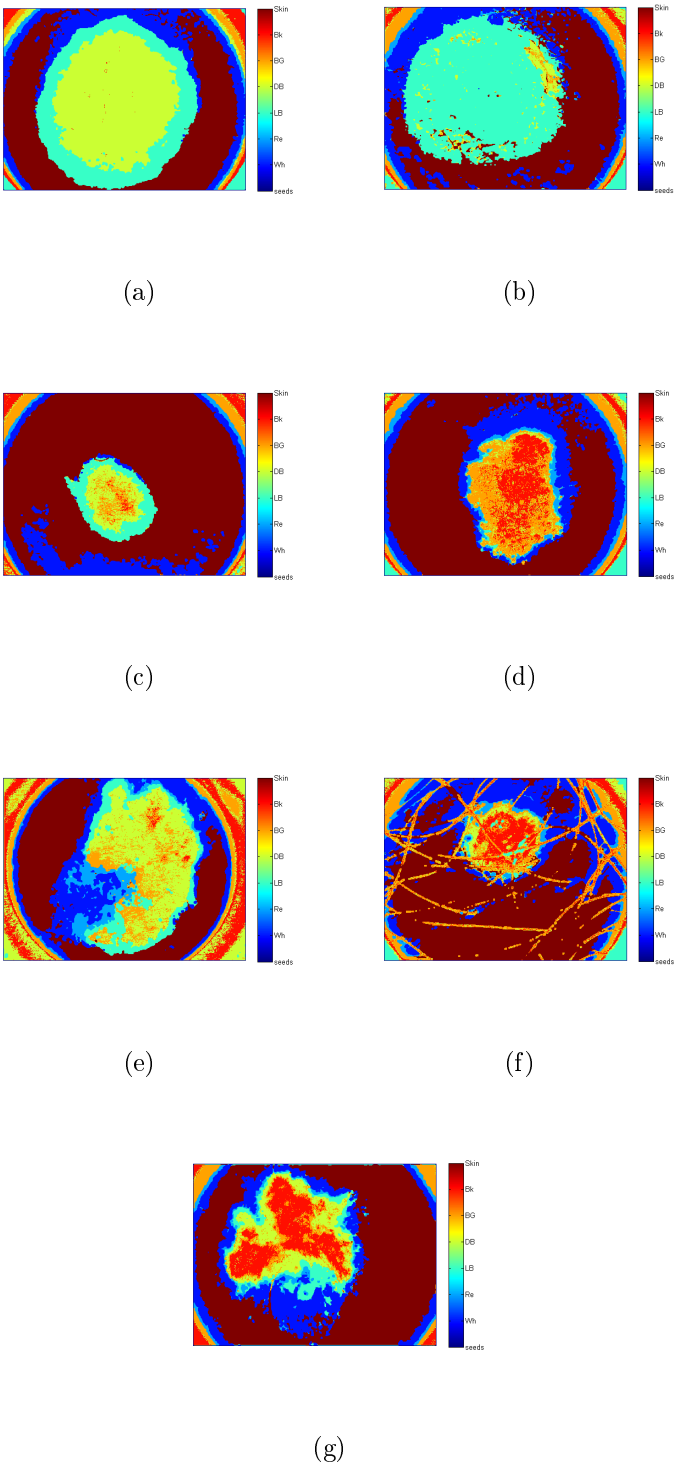
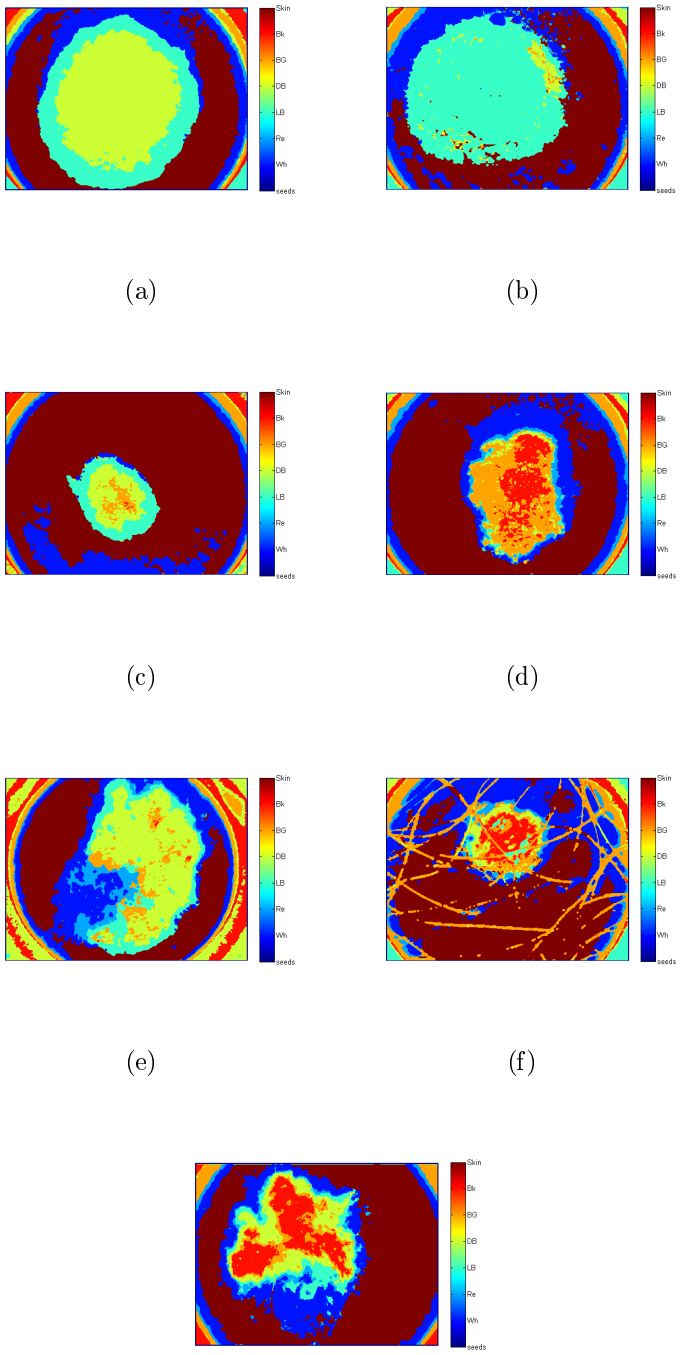


Fig. G.4: Final results obtained with the neural networks in competition in order to divide class group 4 into the three possible color classes (DB, BG and Bk): (a) IMD06, (b) IMD14, (c) IMD15, (d) IMD18, (e) IMD19, (f) IMD21 and (g) IMD28.



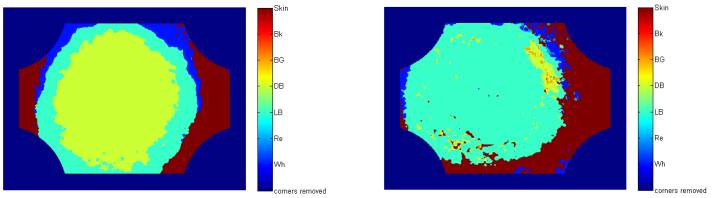
(a) (b)

(c) (d)

(e) (f)

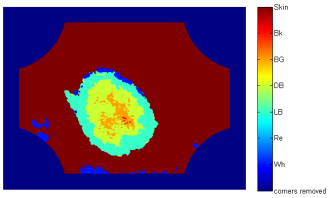
(g)

Fig. G.5: Filtering - Step 2: (a) IMD06, (b) IMD14, (c) IMD15, (d) IMD18, (e) IMD19, (f) IMD21 and (g) IMD28.

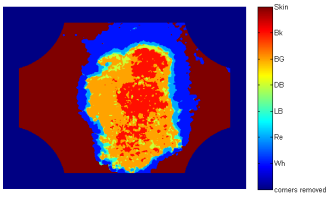


(a)

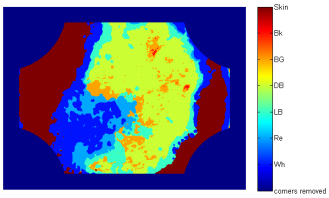
(b)



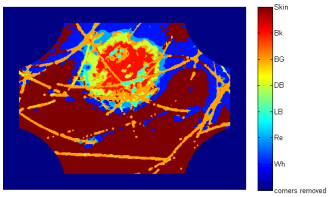
(c)



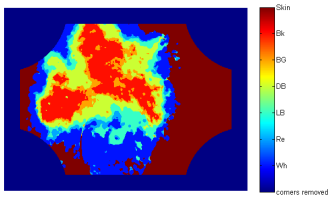
(d)



(e)



(f)



(g)

Fig. G.6: Black corners removal: (a) IMD06, (b) IMD14, (c) IMD15, (d) IMD18, (e) IMD19, (f) IMD21 and (g) IMD28.