



PhD Thesis

**An approach to melanoma classification
exploiting polarization information**

Mojdeh Rastgoo

2016



PhD Thesis

An approach to melanoma classification exploiting polarization information

Mojdeh Rastgoo

2016

DOCTORAL PROGRAM IN TECHNOLOGY

Supervised by: Dr. Rafael García, Dr. Franck Marzani, and Dr. Olivier Morel

Work submitted to the universitat de Girona and Université de Bourgogne in partial
fulfillment of the requirements for the degree of Doctor of Philosophy

Dedicated to Mojgan, Majid, Vahid, Moujan, and Guillaume

Publications

Journal:

M. Rastgoo, O. Morel, F. Marzani and R. Garcia, “Automatic Differentiation of Melanoma from Dysplastic Nevi”, *Computerized Medical Imaging and Graphics*, vol.43, pp 44-52, 2015

International Conferences:

M. Rastgoo, G. Lemaitre, J. Massich, O. Morel, F. Marzani, R. Garcia and F. Meriaudeau, “Tackling the Problem of Data Imbalancing for Melanoma Classification”, *3rd International Conference on BIOIMAGING 2016*. Rome: Italy (February 2016)

M. Rastgoo, G. Lemaitre, O. Morel, J. Massich, F. Marzani, R. Garcia and D. Sidibe, “Classification of melanoma lesions using sparse coded features and random forests”, *SPIE Medical Imaging 2016*. San Diego: USA (February 2016)

M. Rastgoo, O. Morel, F. Marzani and R. Garcia, “Ensemble Approach for Differentiation of Malignant Melanoma”, *International Conference on Quality Control and Artificial Vision (QCAV) 2015*. Le Creusot: France (June 2015)

Abbreviations

ACC accuracy

AdB AdaBoost

ANN Artificial Neural Network

AOLP angle of linear polarization

AUC area under the curve

AWB auto-white-balance

BD barrel deformation

BoW Bag of Words

CAD Computer-Aided Diagnosis

CDS Clinical Diagnosis Support

CLBP Completed Local Binary Pattern

CLSM confocal laser scanning microscopy

CUS Clustering

CWB custom-white-balance

DOCP degree of circular polarization

DoG difference of gaussian

DOLP degree of linear polarization

DOP degree of polarization

DOS data space over-sampling

ECOC Error Correcting Output Coding

ENN Edited Nearest Neighbor

FCM Fuzzy-C-*means*

FP false positive

GB Gradient Boosting

GLCM Grey-Level Co-Occurrence Matrix

GVF Gradient Vector Flow

HOG Histogram of Oriented Gradients

JPEG Joint Photographic Experts Group

k-CV k fold cross-validation

LBP Local Binary Pattern

LC liquid crystal variable retarders

LCTF liquid crystal filter

LDA Linear Discriminant Analysis

LOOCV leave-one-out-cross-validation

MRI Magnetic Resonance Imaging

mRMR Minimum Redundancy Maximum Relevance

MSI multispectral imaging

NB Naive Bayes

NCR Neighborhood Cleaning Rule

NM1 NearMiss-1

NM2 NearMiss-2

NM3 NearMiss-3

NN Nearest Neighbor

NPD nonpolarized dermoscope

OCT optical coherence tomography

OS Over-Sampling

PCA Principal Component Analysis

PD polarized dermoscope

pdf probability density function

PEM photoelastic modulator

PI polarized imaging

PSA polarized state analyzer

PSG polarized state generator

PSLs pigmented skin lesions

RDGM random deformation using gaussian motion

RF Random Forests

ROC Receiver Operating Characteristics

ROS Random Over-Sampling

RUS Random Under-Sampling

SBFS Sequential Backward Feature Selection

SCF Sparse Coded Features

SE sensitivity

SFFS Sequential Forward Feature Selection

SIFT Scale-Invariant Feature Transform

SMOTE Synthetic Minority Over-Sampling

SP specificity

SURF Speeded Up Robust Features

SVD Singular Value Decomposition

SVM support Vector Machine

TIFF Tagged Image File Format

TL Tomek-Link

US Under-Sampling

US ultra-sound

List of Figures

1.1	Skin Anatomy	2
1.2	Clinical and dermoscopy images	6
1.3	Dermoscopes	7
1.4	Clinical, nonpolarized dermoscope (NPD) and polarized dermoscope (PD) images of melanoma	8
2.1	An elliptically polarized wave and the polarization ellipse.	15
2.2	Rotation of the opticalto Polarization field components by rotator	21
2.3	Polarimetry system proposed by [11]	26
2.4	Polarimetry images of ex-vivo cervical dysplasia [11]	27
2.5	Polarimetry system proposed by [209]	28
2.6	Polarimetry images of Chilblain skin sample [209]	29
2.7	Polarimetry system proposed by [186]	30
2.8	Speckle polarimetry images of skin lesions	31
2.9	Polarimetry system proposed by Wood et al. [203]	31
2.10	Polarimetric device proposed by [31]	32
2.11	Snapshot Mueller polarimeter for the configuration, [68].	33
2.12	Full Mueller polarimetry using focused illumination [13]	34
2.13	Mueller matrix obtained by full Mueller imaging with focused illumination system [13]	35

2.14 Fourier space Mueller polarimetry [13]	36
2.15 Polarimetric matrix	37
2.16 Real spape Mueller imaging proposed by [13]	38
3.1 General classification framework	40
3.2 Color variations of dermoscopic images	41
3.3 Color variations between calibrated Joint Photographic Experts Group (JPEG) and RAW images	42
3.4 Artifacts in dermoscopic images	42
3.5 Comparison of inpainting methods	44
3.6 Lesion delineation variations	45
3.7 Kernel Principal Component Analysis (PCA)	49
3.8 Random Forests classifier	59
3.9 Stacking ensemble	60
3.10 Weighted Combination ensemble	61
3.11 Confusion matrix	62
3.12 Sensitivity and Specificity	63
3.13 Receiver Operating Characteristics (ROC) curve	64
4.1 Summary of the literature	69
4.2 Sample images of the Vienna dataset	72
4.3 Image samples of the PH ² dataset	72
4.4 The proposed automated framework	73
4.5 Exemplar inpainting	75
4.6 Results for the proposed hair removal algorithm	77
4.7 probablity density function-based segmentation	78
4.8 Fusion filters	80
4.9 Segmentation results obtained by the pdf-based, level-set and Fuzzy-C-means (FCM) algorithms	83

4.10 Completed Local Binary Pattern (CLBP) algorithm	86
4.11 Six orientation of Gabor filter	86
4.12 Data space over sampling algorithm	92
4.13 Data space over sampling example	100
4.14 Proposed framework for an automated classification using sparse coded features. .	110
4.15 Twelve highest results obtained by Experiment #5	113
5.1 Internal design of polarimeter dermoscope	119
5.2 The first prototype of polarimeter dermoscope	120
5.3 The illumination system and electronic board of the polarimetric dermoscope .	121
5.4 Example of a polarimetric acquisition saved in JPEG and sRGB images converted from Raw-CR2 to TIFF	122
5.5 The first and second Stokes parameters, I, Q, obtained from polarimetric images	123
5.6 The third Stokes parameter, U, obtained from polarimetric images	124
5.7 First and second Stokes parameters in RGB model.	124
5.8 Degree of linear polarization (DOLP) and angle of linear polarization (AOLP) in the R,G, and B channels	126
5.9 DOLP and AOLP in grayscale	127
5.10 The first and second Stokes parameters for melanoma nevi	128
5.11 Third Stokes parameter for melanoma nevi	129
5.12 Degree of linear polarization (DOLP) and AOLP in the R,G,B channels of melanoma nevus	130
5.13 Proposed framework for automatic classification of melanoma lesions based on polarimetric images.	131
5.14 Integral image and approximation of the Gaussian second order derivatives using box filter	134
5.15 Examples of hair removal algorithm on polarimetric images	135
5.16 Segmentation results on polarimetric images	137

5.17 Images of polarimetric features.	139
---	-----

List of Tables

2.1	Stokes vector for polarization states.	17
2.2	Mueller matrix of quarter-wave and half-wave retarders.	21
2.3	Medium Characteristics	24
2.4	Polarized state generator and polarized state analyzer (PSA) units setting proposed by [13]	35
4.1	Summary of the features used in the state of the art	70
4.2	Summary of Vienna and PH ² Datasets. M, D, and B stand for melanoma, dysplastic and benign samples, respectively.	71
4.3	Extracted Features in our proposed framework	82
4.4	Extracted features by Grey-Level Co-Occurrence Matrix (GLCM) descriptor . .	87
4.5	Outline of experiments	94
4.6	Results obtained using globally mapped features, Experiment #1	96
4.7	Results obtained using locally mapped features, Experiment #1	97
4.8	A list of melanoma and dysplastic lesions in the data space over-sampling (DOS)-balanced training, Experiment #2	100
4.9	Results of Experiment #2 - DOS-balance-training case	101
4.10	Results of Experiment #2 - comparison of the three cases	103
4.11	The PH ² benchmark results	104
4.12	Results of Experiment #3	105

4.13 Results of Experiment #4	108
4.14 Classification costs in Experiment #4	109
4.15 Results of Experiment #5	112
5.1 Summary of acquired polarimetric dataset	127
5.2 Features used in the automatic classification of melanoma lesions from polarimetric images	140
5.3 The outline of the experiments	143
5.4 Results of Experiment #1	145
5.5 Results of Experiment #2	146
5.6 Results of Experiment #3 - steps 1 and 2	149
5.7 Results of Experiment #3 - step 3	150
5.8 Results of Experiment #4 - step 1 and 2	152
5.9 Results of Experiment #4 - step 3	153

Acknowledgments

I am absolutely grateful to my supervisors, Rafael García, Franck Marzani and Olivier Morel. Without their supports, guidelines, insights and encouragements, this work would not be simply possible.

I would also like to thank Fabrice Mériadeau for his constant support. Thanks to Désiré Sidibé as well for sharing his brilliant ideas. Thanks to Pablo Iglesias and Beatriz Alejo Galindo for being patient and providing me with the medical insight and data; also to Dr. Josep Malvehy and Susana Puig for sharing their time and valuable clinical insights for the improvement of this project. A particular thanks to Steven Jacques for sharing his valuable knowledge and most importantly being inspiring.

I would like to thank my friends; Habib, Sarah, Amanda, Sonia, Sharad, Shihav, David, Abir, Mohamed, Cansen, Qinglin, Richa, Armine, Priyanka, Joan -for having his unique way of thinking- and Konstantin whom patiently and perfectly corrected the final manuscript and provided me with his thoughtful ideas. Thank you all for the shared happy moments and your help and support during this time.

Also I would like to thank my colleagues, Nuno, Tudor, László, Josep, Ricard P., Ricard C. and Konstantin for all the interesting lab meetings and their help for the final preparation of the thesis; to Joseta, Mireia and Montse, for being always there and helping me with endless amount of paperwork and questions.

My gratitude, as well, to the anonymous reviewers and the members of the defense panel for evaluating my work. I would also like to acknowledge AGAUR FI-DGR 2012 grant, provided by the Autonomous Government of Catalonia which is the sole reason for existence of this thesis. It is an honored to be among its holders.

Foremost, I would like to thank my family that has always been there for me, despite the long distance between us. Their love, support and motivation gave me the strength to finish the thesis. My most loving thanks to my parents, Mojgan and Majid, brother Vahid, sister Moujan, both grandmothers: Maman Alam and Maman joon, aunt Mahboubeh, aunt Manijeh, cousin Sudeh, cousin Somayeh and cousin Maryam. And all the family.

I would like to thank Ginette, Patrice, Cedric, Vero, Julian and Sarah for being my family here.

Last but not least, I want to thank my husband and best colleague Guillaume for his love and support and being my best friend.

Contents

List of Abbreviations	vii
List of Figures	x
List of Tables	xv
1 Introduction	1
1.1 The Human Skin	1
1.2 Pigmented Skin Lesions	3
1.3 Malignant Melanoma	4
1.4 Melanoma Diagnosis and Screening	5
1.4.1 Clinical photography	6
1.4.2 Dermoscopy	6
1.4.3 Polarized imaging	8
1.5 Automated Diagnosis of Melanoma	8
1.6 Research Motivation	10
1.7 Thesis Outline	10
2 Polarization Principles and Image Polarimetry	13
2.1 Polarization Properties of Light	13
2.2 Stokes Parameters	16

2.3	Optical Properties of Medium	18
2.3.1	Mueller matrix parameters	18
2.3.2	The Lu-Chipman decomposition	22
2.4	Image Polarimetry in Biomedical Field	24
2.4.1	Partial Stokes polarimeters	25
2.4.2	Stokes polarimeters	27
2.4.3	Mueller polarimeter	30
2.5	Conclusion	37
3	Machine Learning and Classification Principles	39
3.1	Introduction	39
3.2	Preprocessing	40
3.2.1	Image enhancement	40
3.2.2	Artifacts removal	41
3.2.3	Image segmentation	43
3.3	Feature Extraction	46
3.4	Feature Representation	46
3.5	Balancing Strategies	52
3.5.1	Under-Sampling	52
3.5.2	Over-Sampling	53
3.5.3	Combination of OS and US	54
3.6	Classification	54
3.6.1	Single Learner	54
3.6.2	Ensemble	56
3.7	Validation and Evaluation	61
3.7.1	Validation	61
3.7.2	Evaluation	61
3.8	Conclusion	64

4 Automated Melanoma Classification with Dermoscopic Images	65
4.1 Introduction	65
4.2 Related Work	66
4.3 Materials	71
4.4 Methodology	73
4.4.1 Pre-processing	73
4.4.2 Mapping	82
4.4.3 Feature Extraction	84
4.4.4 Feature representation	91
4.4.5 Data Balancing	91
4.4.6 Classification	93
4.5 Experiments and Results	95
4.5.1 Experiment #1	
Melanoma vs. Dysplastic Nevi. Balanced Subsets of the Vienna Dataset	95
4.5.2 Experiment #2	
Melanoma vs. Dysplastic Nevi, Imbalance subset of Vienna Dataset	99
4.5.3 Experiment #3	
Melanoma .vs Benign and Dysplastic nevi, Ensemble approach on PH ² dataset	102
4.5.4 Experiment #4	
Melanoma vs. Benign and Dysplastic nevi, Study of balancing techniques on PH ² dataset	106
4.5.5 Experiment #5	
Malignant vs. Benign and Dysplastic nevi, Sparse coded features using PH ² dataset	109
4.6 Conclusion	111
4.6.1 Future Work	115
5 Automated Melanoma Classification with Polarimetric Images	117

5.1	Polarimeter Dermoscope	118
5.2	Polarimeter Images	118
5.3	Dataset	127
5.4	Methodology	131
5.4.1	Pre-processing	131
5.4.2	Mapping	136
5.4.3	Feature extraction	136
5.4.4	Feature representation	138
5.4.5	Balancing	141
5.4.6	Classification	141
5.5	Experiments and Results	141
5.5.1	Experiment #1 Melanoma vs. Benign lesions. leave-one-out-cross-validation (LOOCV) validation on a subset of the data.	144
5.5.2	Experiment #2 Melanoma vs. Suspicious and Benign lesions. LOOCV validation on a subset of the data.	147
5.5.3	Experiment #3 Melanoma vs. Benign lesions. A comparison of balancing techniques.	147
5.5.4	Experiment #4 Melanoma vs. Suspicious and Benign lesions. A comparison of balancing techniques.	151
5.6	Conclusion	154
5.6.1	Future work	156
6	Conclusion	157
6.1	Summary of the thesis	158
6.2	Contributions	159

6.3	Limitations	160
6.4	Future work	161
	Bibliography	182

Abstract

Malignant melanoma is the deadliest of skin cancers and causes the majority of deaths in comparison to other skin-related malignancies. Yet it is the most treatable type of cancer, thanks to its early diagnosis. Subsequently, early diagnosis is crucial for patient survival rate and numerous Computer-Aided Diagnosis (CAD) systems have been proposed by the research community to assist dermatologists in early diagnosis. These systems are based on the most common skin imaging modality, cross-polarized dermoscopy. Cross-polarized dermoscopes (PD) allow for the visualization of the subsurface anatomic structure of the epidermis and papillary dermis and eliminate the specular reflection of the surface. Although this modality has been used extensively, the full potential of polarized measurements has not been realized in the field of skin imaging.

This research first extensively analyzes different aspects of the automated classification of pigmented skin lesions (PSLs) and proposes a CAD system for automatic recognition of melanoma lesions based on the PD images. The proposed CAD system is evaluated over extensive experiments on two dermoscopic datasets. Later for further investigation of polarized imaging, a novel partial Stokes polarimeter system is proposed. This system is able to acquire polarized images of in-vivo PSLs and capture the epidermis and superficial dermal layers, where skin lesions are often originated. The polarized and dermoscopy properties of the acquired images are then analyzed to propose a new CAD system based on image polarimetry. The initial tests with the first prototype of Stokes polarimeter revealed the potential and benefits of such systems for providing additional information beyond RGB images acquired with PD devices. In order to acquire a wider clinical dataset and identify the drawbacks of the first prototype, this device is currently being used in the Melanoma Unit at the Clinic Hospital of Barcelona.

Resum

El melanoma maligne és el més mortal dels càncers de pell i provoca la majoria de les morts en comparació amb altres tumors malignes relacionats amb la pell. No obstant això, és el tipus més tractable de càncer, gràcies al seu diagnòstic precoç. Per tant, el diagnòstic precoç és crucial per a la supervivència dels pacients. Nombrosos sistemes de diagnòstic assistit per ordinador (CAD, de l'anglès Computer Aided Diagnosis) han estat proposats per la comunitat investigadora per ajudar als dermatòlegs en el diagnòstic precoç. Aquests sistemes es basen en la modalitat més emprada d'adquisició d'imatges de pell, la dermatoscòpia de polarització creuada (PD, de l'anglès Polarized Dermatoscopy). La dermatoscòpia de polarització creuada permet la visualització de l'estructura anatòmica del subsòl de l'epidermis i la dermis papilar, eliminant les reflexions especulars de la superfície. Tot i que aquesta modalitat ha estat utilitzada àmpliament, no tot el potencial de les mesures polaritzades ha estat aprofitat en el camp de la imatge de pell.

Aquest treball de recerca analitza àmpliament, en primer lloc, diversos aspectes de la classificació automatitzada de les lesions cutànies pigmentades (PSLs, de l'anglès Pigmented Skin Lesions) i proposa un sistema CAD per al reconeixement automàtic de lesions de melanoma en base a les imatges de PD. El sistema CAD proposat es va avaluar en el transcurs d'extensos experiments en dos conjunts de dades dermatoscòpiques. Posteriorment, en una investigació més extensa pel que fa a la formació d'imatges polaritzades, es proposa un nou sistema de partial Stokes polarimeter. Aquest sistema és capaç d'adquirir imatges polaritzades dels PSLs en viu, capturant l'epidermis i les capes dérmiques superficials, on sovint s'originen les lesions de la pell. Les propietats de polarització i dermoscopia de la imatge són analitzades a continuació, proposant un nou sistema CAD basat en la imatge de polarimetria. Les proves inicials, amb el primer prototip d'Stokes polarimeter, han revelat el potencial i els beneficis de tals sistemes per proporcionar informació addicional més enllà de les imatges RGB adquirides amb dispositius PD. Per tal d'adquirir un conjunt de dades clíniques més ampli i identificar els inconvenients del primer prototip, aquest dispositiu s'està utilitzant actualment a la Unitat de Melanoma de l'Hospital Clínic de Barcelona.

Resumen

El melanoma maligno es el más mortal de los cánceres de piel y provoca la mayoría de las muertes en comparación con otros tumores malignos relacionados con la piel. Sin embargo, es el tipo más tratable de cáncer, gracias a su diagnóstico precoz. Por lo tanto, el diagnóstico precoz es crucial para la supervivencia de los pacientes. Numerosos sistemas de diagnóstico asistido por ordenador (CAD, del inglés Computer Aided Diagnosis) han sido propuestos por la comunidad investigadora para ayudar a los dermatólogos en el diagnóstico precoz. Estos sistemas se basan en la modalidad más empleada de adquisición de imágenes de piel, la dermatoscopia de polarización cruzada (PD, del inglés Polarized Dermatoscopy). La dermatoscopia de polarización cruzada permite la visualización de la estructura anatómica del subsuelo de la epidermis y la dermis papilar, eliminando las reflexiones especulares de la superficie. Aunque esta modalidad ha sido utilizada ampliamente, no todo el potencial de las medidas polarizadas ha sido aprovechado en el campo de la imagen de piel.

Este trabajo de investigación analiza ampliamente, en primer lugar, varios aspectos de la clasificación automatizada de las lesiones cutáneas pigmentadas (PSLs, del inglés Pigmented Skin Lesions) y propone un sistema CAD para el reconocimiento automático de lesiones de melanoma en base a las imágenes de PD. El sistema CAD propuesto se evaluó en el transcurso de extensos experimentos en dos conjuntos de datos dermatoscópicos. Posteriormente, en una investigación más extensa en cuanto a la formación de imágenes polarizadas, se propone un nuevo sistema de partial Stokes polarimeter. Este sistema es capaz de adquirir imágenes polarizadas de los PSLs en vivo, capturando la epidermis y las capas dérmicas superficiales, donde a menudo se originan las lesiones de la piel. Las propiedades de polarización y dermoscopia de la imagen son analizadas a continuación, proponiendo un nuevo sistema CAD basado en la imagen de polarimetría. Las pruebas iniciales, con el primer prototipo de Stokes polarimeter, han revelado el potencial y los beneficios de tales sistemas para proporcionar información adicional más allá de las imágenes RGB adquiridas con dispositivos PD. Para adquirir un conjunto de datos clínicos más amplio e identificar los inconvenientes del primer prototipo, este dispositivo se está utilizando actualmente en la Unidad de Melanoma del Hospital Clínic de Barcelona.

Résumé

Le mélanome malin est le plus mortel des cancers de la peau. Il cause la majorité des décès au regard des autres pathologies malignes de la peau. Toutefois, ce type de cancer se soigne dès lors qu'un diagnostic est posé précocement. Ainsi, le taux de survie est fortement corrélé à un diagnostic précoce ; de nombreux systèmes d'aide au diagnostic (CAD) ont été proposés par la communauté pour assister les dermatologues dans leur diagnostic. La modalité d'imagerie de la peau la plus classiquement utilisée est la dermatoscopie avec polarisation croisée. Les dermatoscopes avec polarisation croisée (PD) permettent la visualisation de la structure anatomique inférieure de l'épiderme, le derme papillaire et éliminent la réflexion spéculaire de surface. Bien que cette modalité ait été utilisée très fréquemment, le fort potentiel des mesures de polarisation n'a pas été étudié dans le domaine de l'imagerie de la peau.

Dans un premier temps, notre recherche a porté sur une analyse poussée des différents aspects de la classification automatique des lésions pigmentaires (PSLs) ce qui nous permet de proposer un système CAD pour la reconnaissance automatique des lésions de type mélanome à partir d'images de modalité PD. Ce système CAD est évalué à partir de nombreuses expérimentations effectuées sur deux bases de données d'images. Dans un deuxième temps, afin d'étudier l'imagerie de polarisation, un nouveau système de polarimétrie partiel de type Stokes est proposé. Ce système est capable d'acquérir des images polarisées de PSLs in-vivo de l'épiderme et des couches superficielles du derme, fréquemment à l'origine des lésions de la peau. Les propriétés de polarisation et de dermatoscopie des images acquises sont ensuite analysées afin de proposer un nouveau système CAD basé sur l'imagerie de polarisation. Les tests préliminaires avec ce premier polarimètre de Stokes montrent le potentiel et les bénéfices possibles afin de produire des informations complémentaires à celles issues des images couleur RGB classiquement obtenues avec la modalité PD. Ce prototype est actuellement en cours d'utilisation au Melanoma Unit de la Clinic Hospital de Barcelone (Espagne) afin de constituer une base d'images plus conséquente et ainsi identifier les désavantages d'un tel système.

Chapter 1

Introduction

1.1 The Human Skin

Skin is the largest organ of the human body and consists of three main layers; epidermis, dermis and subcutaneous fat [12] (see Fig.1.1).

Epidermis is the outer layer or surface of the skin. This layer is divided into four sub-layers from top to bottom; stratum corneum, stratum granulosum, stratum spinosum and stratum basal [134].

The aforementioned layers contain four types of cells including keratinocytes, melanocytes, Langerhans' cells and Merkel cells. The majority of the cells in the epidermis are keratinocytes, the main force for continuous renewal of the skin [134]. These cells contain two main attributes of division and differentiation which enable them to renew the outer layer of the epidermis within thirty days. During this journey, the keratinocytes, which are produced by division of the basal cells (keratinocytes in the basal layer are called basal cells), will move to the next layer while they go through a differentiation process. Differentiation refers to this morphology and biochemistry transformation of the cells. At the end of this journey, the keratinocyte cells will lose their nucleus and be transformed to flattened cells filled with keratin. These cells form the outer most layer of the epidermis (stratum corneum). At the end of the differentiation process, the corneocytes lose their cohesion and separate from the surface in the desquamation process, resulting in renewed skin.

Langerhans' cells are responsible for the detection of foreign bodies (antigens) which invade the epidermis, transporting them to the local lymph nodes, while Merkel cells act

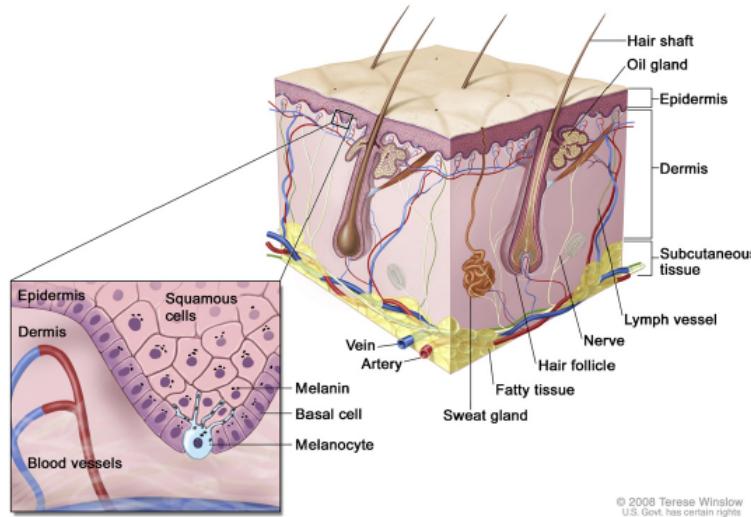


Figure 1.1: Anatomy of the skin, showing the three structure layers of epidermis, dermis and subcutaneous tissue [112].

as mechanosensory receptors in response to touch, forming close connections with sensory nerve endings [134].

Melanocyte cells are found in the basal layer of the epidermis [134]. These cells are responsible for distributing packages of melanin pigment, which lead to individual skin and hair colors, to the surrounding keratinocytes. This chromophore mainly protects the subcutaneous tissue from being damaged by UV radiation. Whenever the level of UV radiation increases, melanocytes start producing more melanin, resulting in our tanning reaction to sun exposure. Melanin is the major chromophore of the epidermis which occupies the top 50-100 μm , with the exception of superior layers of epidermis.

In most light propagation models through skin, the sublayers of epidermis are considered as one layer [121, 108]. The epidermal thickness can vary depending on different body parts, however on average, it is usually around 0.1 mm. The most external sublayer of the epidermis is the stratum corneum. It is composed of dry dead cells without organelles and is filled with keratin fibres. Light is absorbed mainly in this layer, due to the epidermis major component melanin.

Dermis is the middle layer between the epidermis and the hypodermic layer. This layer is thicker than the epidermis, with an average thickness of 0.6 to 3 mm [12]. The thermoregulation, the mechanical resistance and nourishing of the epidermis are the main functions of this layer. The dermis is composed of elastic collagen fibres, blood vessels,

nerves, lymph vessels, hair follicles and sweat glands but its principal molecules, with relevant optical properties, are haemoglobin, carotene and bilirubin. Haemoglobin is a chromophore of red colour found in the microvascular network of the dermis, typically 100-500 µm below the skin's surface. This chromophore carries oxygen through vessels and capillaries and accordingly is called oxy-haemoglobin since it contains oxygen and deoxy-haemoglobin. The dermis is divided into two sublayers, the papillary dermis with the principal function of thermoregulation and the reticular dermis which gives the skin its strength and elasticity.

Subcutaneous fat or hypodermic fat is the deepest layer of the skin, with an average thickness of 4 to 9 mm. This layer is composed of connective tissues, fat cells and blood vessels.

Cancer can develop in almost any cell in the body. However, certain cells are more cancer prone compared to others and the skin is no exception. The three most common malignant skin cancers are called basal cell carcinoma, squamous cell carcinoma and melanoma which develop from basal cells, squamous keratinocytes, and melanocytes, respectively. Melanocyte cells and their transformation, due to their malignant transformation potential, are our main concern in this thesis. Melanoma is less common in comparison to basal cell carcinoma and squamous cell carcinoma. However, it is the deadliest and most aggressive type of skin cancer. The characteristics and treatment of this cancer, as well as some skin lesions very prone to malignancy are described in the following sections.

1.2 Pigmented Skin Lesions

Pigmented skin lesions or melanocytic nevi appear on the surface of the skin [110], where the melanocyte cells grow in clusters beside the normal skin cells. This is a natural transformation of skin cells and creates benign pigmented skin lesions (PSLs) such as:

Freckle or ephelis are pale-brown macular lesions which are usually of less than 3 mm diameter with a poorly defined lateral margin [140].

Common nevi which are typical flat melanocytic nevi or moles.

Congenital nevi which are moles that appear at birth, also known as "birth marks".

Atypical or dysplastic nevi are common nevi with inconsistent coloration, irregular edges, blurry borders, scale-like texture and a diameter greater than 5 mm [110]. Atypical mole syndrome or dysplastic nevus syndrome, describes individuals with large quantities of dysplastic nevi. Such individuals face a higher risk of developing melanoma (6 to 10

times greater than other people with few nevi) [140]. However, only a small number of these dysplastic nevi might develop into melanoma and most dysplastic nevi will never become cancer.

Blue nevi are melanocytic nevi comprised of abnormal collections of benign pigment melanocytes located in the dermis rather than at the dermoepidermal junction [140]. The blue or blue-black appearance of the lesion is caused by light reflection of melanin in the dermis.

Pigmented Spitz nevi are uncommon benign nevi that share a similar physical characteristics to melanoma and are usually seen in children [140].

From the aforementioned lesions, congenital and dysplastic nevi are most likely to develop into malignant melanoma [81].

1.3 Malignant Melanoma

Although malignant melanoma accounts for less than 2% of all skin cancer cases, it is the deadliest type and causes the vast majority of deaths [183]. The incidence of melanoma has increased in the past few decades currently reaching 132,000 melanoma cases per year according to the World Health Organization.

Melanoma cancer is incurable in its advanced stages and the patient should go through surgery, possibly immunotherapy, chemotherapy, and/or radiation therapy. However, if it is diagnosed at an early stage, it is the most treatable kind of cancer [183, 79]. In fact, patient survival rate has increased significantly, over the past few decades, thanks to early diagnosis and treatment of melanoma in its early stages.

The stages of melanoma are measured based on how the lesion has grown, including its invasion depth through the skin to nearby lymph nodes or other organs. This factor is measured through physical exam, biopsies, and different imaging tests such as computed tomography or Magnetic Resonance Imaging (MRI) [183]. Depending on the measurements obtained, melanoma skin cancers are divided into four types.

The first three types begin *in situ*, meaning that they spread along the top layers of the skin and become invasive in the final stages, while the fourth is invasive from the start. Invasive melanoma can be very dangerous since they are in the deeper layers of the skin and can spread faster to other body parts. The four types of cutaneous melanoma are listed in the following:

Superficial spreading melanoma is the most common type and is the leading cause of cancer death in young adults. Approximately 70% of all melanoma diagnosis are counted as superficial spreading melanoma. This type grows along the top layer of the skin and often

occurs in a previously benign mole. The location of this melanoma can be anywhere in the body, however, it is usually found on the trunk and back in male patients and on the legs and back in females patients.

Acral lentiginous melanoma accounts for less than 5% of all melanoma diagnosis and is the most common type in dark skinned individuals. This disease is usually located on the palms, soles of the feet and under the finger nails and often looks like a bruise or injury on the body. For this reason, the melanoma may be discovered later than other forms.

Letingo maligna melanoma accounts for 5-10 % of melanoma diagnosis. Letingo maligna melanoma arises from a pre-existing letingo rather than a mole and mostly occurs on the face of middle-aged and elderly individuals as a result of sun damage. If this cancer goes undiagnosed, being mistaken for a sun spot, it can spread to the deeper layers of the skin and endanger the patient's life. Cancer lesions of this type usually have a very irregular border and vary in shades of brown or black. However, like other types of melanoma, they can be blue, red, gray or white.

Nodular melanoma accounts for 15-30% of all melanoma diagnosis. This is the most aggressive type due to the fact that it spreads more rapidly in depth and it is difficult to visualize the progression of the cancer. Nodular melanoma is more common in males than in females. The lesion is usually darkly pigmented (blue-black) and is often found in pink or red.

1.4 Melanoma Diagnosis and Screening

The clinical prognosis of early stage melanoma is commonly done via visual inspection of the lesions, based on a set of rules or guidelines, such as "ABCDE" [9] or Glasgow 7-point checklist [9]. The "ABCDE" rule characterizes the lesion based on its asymmetry (A), irregular borders (B), variegated colors (C), diameter ≥ 6 mm (D) and evolving stage over time (E). The Glasgow 7-point checklist contains 7 criteria: 3 major (changes in size, shape and color) and 4 minor (diameter ≥ 7 mm, inflammation, crusting or bleeding, and sensory change). The former rule has been extensively used in clinical routine, rather than the latter one, due to its simplicity.

The visual inspection of lesions is carried out using different non-invasive imaging techniques such as: clinical photography, dermoscopy, confocal laser scanning microscopy (CLSM), optical coherence tomography (OCT), multispectral imaging (MSI), high frequency ultra-sound (US), and Magnetic Resonance Imaging (MRI) among other spectroscopic imaging. Concerning the aforementioned techniques, some are well-utilized by clinicians and dermatologists. We will



(a) In situ melanoma (stage 0)



(b) Invasive melanoma (stage I/II)

Figure 1.2: Clinical and Dermoscopy images, right and left column, respectively. Images submitted to www.dermoscopyatlas.com by Dr. Alan Cameron (a), Dr. Jean-Yves Gourhant (b). Used with permission.

refer here and after to these techniques as “conventional” techniques. Clinical photography, dermoscopy, OCT, and CLSM belong to this category. While the rest, such as MRI, US, MSI and polarized imaging (PI) are categorized as “non-conventional” techniques.

We focus our research on conventional techniques such as clinical and dermoscopy as well as non-conventional techniques such as PI.

1.4.1 Clinical photography

Clinical photographs are referred to as digital or non-digital images captured from the surface of the skin, showing one or more lesions. These images reproduce what a clinician sees with the naked eye [60] and are commonly tainted with unnecessary highlights and reflections from the outer layer of the epidermis. The left column of Fig. 1.2 shows such images.

1.4.2 Dermoscopy

Dermoscopy (also known as dermatoscopy or epiluminescence microscopy) is a well-established and effective non-invasive technique for early recognition of melanoma. This technique, introduced in 1971 [124, 123], uses a hand-held lighted magnifier to analyze skin lesions.

Initially the device was used in conjunction with a thin layer of glass and an oil or alcohol interface to reduce light reflection, refraction and diffraction. This material made the epidermis

translucent and allowed *in vivo* visualization of subsurface anatomic structures of the epidermis and papillary dermis which are not visible with naked eye [162, 196]. This first type of dermoscopes is called a nonpolarized dermoscope (NPD) [196].

The second type, the polarized dermoscope (PD), was introduced later and made this process much easier by using cross-polarized light. This device is equipped with two polarized filters, one in front of the light source and one in front of the sensor. The two polarized filters are perpendicular to each other in order to capture the backscattered light from the deeper levels of the skin. The light reflected from the skin has the same angle as the polarized incident light, hence it is eliminated with the cross-polarized filter in front of the sensor. However, the backscattered light from the skin, due to the structural nature of the tissue, will become unpolarized and passes through the filter. This technique eliminates the need for fluid or direct contact with the skin.

A sample of commercially available dermoscopes (Dermlite®) with and without polarized light are shown in Fig 1.3. MoleMax (Derma Medical System, Vienna, Austria) is another commercially available system that includes a polarized dermoscope. MoleMax is a computer based device and by providing its own software, allows live-video dermoscopy and total body photography [132, 162].



Figure 1.3: Commercially available dermoscopes by Dermlite®: (a) immersion fluid dermoscope using non-polarized light; (b) cross-polarized light dermoscope. Both devices can be attached to a digital camera.

The images captured using a PD or a NPD are relatively similar. However, the surface dermoscopic structures (such as the blue white veil) are better observed with a NPD while deep structures (such as vessels) are better seen with a PD [196, 27]. The right column of Fig 1.2 demonstrates the dermoscopic images of the same lesions captured with an NPD. Figure 1.4 shows difference of clinical, NPD, and PD imaging for a melanoma lesion.

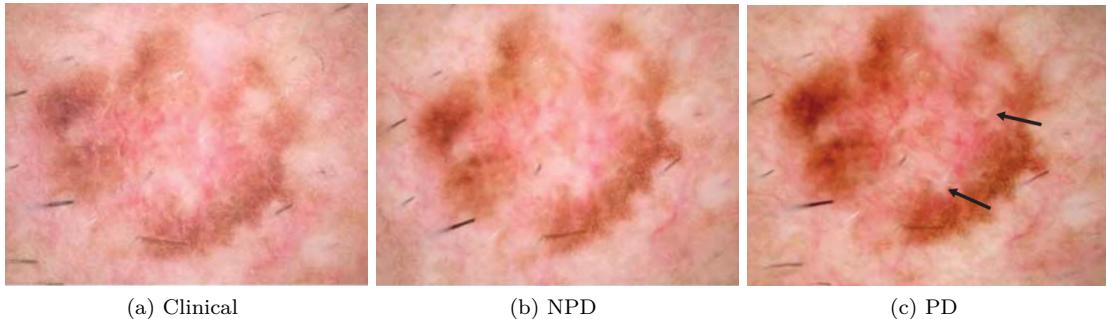


Figure 1.4: Melanoma lesion captured by clinical photography, An NPD, and a PD, respectively. Shiny white streaks (arrow in (c)) within the melanoma are only visible with the NPD. The images are taken from Benvenuto et al. [27]

Besides the NPD and the PD, another dermatoscopy technique, via transillumination, was introduced by Dhawan et.al. [63, 64]. In this technique, light is directed to the skin in such a way that allows the backscattered light to illuminate the lesion from within. Novescopic is a patented device developed for this technique.

1.4.3 Polarized imaging

In polarized imaging (PI) or image polarimetry systems a polarizer state generator and analyzer are used to create a set of polarized images. These images define the polarization state of the light beam and the depolarization property of the tissue. The former is represented by four measurable quantities called Stokes parameters and the later is represented by the Mueller matrix. The advantage and benefits of PI systems besides being cross-polarized are not as evident in the field of skin imaging or tissue imaging in general. Over the last few decades, few studies have been dedicated to finding and exploring the depolarization properties of tissues [11, 127, 127, 106] The Stokes and Mueller measurements along with the basics of polarization and the methods proposed by the research community are further explained in Chap. 2

1.5 Automated Diagnosis of Melanoma

Computer-Aided Diagnosis (CAD) or Clinical Diagnosis Support (CDS) systems are proposed to provide automated diagnosis of melanoma lesions. These systems are intended to reproduce the decision of the dermatologists when observing images of PSLs. Automated diagnosis of melanoma was proposed to assist dermatologists and increase the sensitivity and specificity of melanoma recognition in their early stages as well as to reduce unnecessary excisions. From a

computer vision and pattern recognition point of view, CAD systems for melanoma intend to classify and differentiate melanoma lesions from others. In general, image processing techniques are used to locate and delineate the lesions and extract the image parameters (features) which coincide with the dermatologist's point of view and dermatological features. The extracted parameters are further used with machine learning tools to perform a diagnosis (classification). Such systems are being developed for various imaging modalities [131, 194, 112]. However, dermoscopy being the most conventional imaging technique, most CAD systems are dedicated to this modality.

This research is aimed at analyzing the effects of polarized illumination beyond that of usual dermoscopes, for the detection of melanoma lesions and providing a CAD system for an automated classification of melanoma lesions based on conventional dermpscopes and image polarimetry. The general steps of a CAD system are extensively discussed in Chap. 3.

Clinical impact

Numerous research projects have been dedicated to the development of a CAD or CDS system and studies show that their performance is sufficient under experimental conditions [82]. The proposed systems can assist dermatologists, however, their practical value is still unclear and they cannot be recommended as a sole determinant of malignancy of a lesion. Even though most patients would accept a computerized analysis of a melanoma, the CAD systems proposed cannot function alone due to their tendency to over diagnose benign melanocytic and non-melanocytic lesions [82]. Day and Barbour [59] listed two main shortcomings for general approaches which are adapted to develop a CAD system for recognition of a melanoma:

1. A CAD system is expected to reproduce the decision of pathologists (a binary result like “melanoma/non-melanoma lesion”) with only the input used by dermatologists: clinical or dermoscopic images;
2. Histopathological data are not available for all lesions, only for those considered suspicious by dermatologists.

The items listed refer to the lack of sufficient information and interaction with dermatologists used in the proposed method. These items refer to the fact that the proposed methods lack sufficient information and interaction with dermatologists. This fact was highlighted by Dreiseitl et al. [66] as well. The authors mentioned that current systems are designed to work “in parallel with and not in support of” physicians, thus only a few systems have been used in clinical routines. In this regard, an ideal CAD or CDS system for melanoma recognition should be able to reproduce the dermatologists's decision with extensive information regarding the reason and ground for that decisions [66].

1.6 Research Motivation

Malignant melanoma is the deadliest type of skin cancer and accounts for the vast majority of skin cancer deaths [183]. According to the latest reports, melanoma causes over 20,000 deaths annually in Europe [79]. The American Cancer Society also reported an estimated deaths of melanoma in 2014 as 9710 individuals and new cases as 76,100 individuals [183]. Nevertheless, melanoma is the most treatable kind of cancer if diagnosed early. Therefore, prevention and early diagnosis of melanoma lesions is crucial for the patients survival rate.

Computer-Aided Diagnosis systems have been proposed by the research community to assist dermatologists in early diagnosis of melanoma. These systems are proposed to classify pigmented skin lesions (PSLs). Dermoscopy being the most common source of skin imaging, most of the CAD systems proposed are based on polarized dermoscopy (PD). The PD allows the visualization of the subsurface anatomic structure of the epidermis and papillary dermis. Although PD are used extensively by the dermatologist to document and analyze the PSLs. The advantage and benefits of polarized imaging (PI) beyond the cross-polarized filters, have not been fully explored in the field of skin cancer.

This work attempts to analyze more closely PI in the field of skin cancer and intends to propose an automatic classification framework based on dermoscopy and image polarimetry. For this purpose, a novel image polarimetry system able to provide the first three Stokes parameters is presented. This system is implemented based on Stokes polarimetry since it can provide automatic, fast, accurate and less complex acquisition, in comparison to Mueller polarimetry.

Towards providing a CAD system, different aspects of automated classification of PSLs are extensively investigated and a CAD system based on dermoscopy images is proposed. Further using the proposed image polarimetry system, polarization properties of the PSLs are analyzed and a CAD system based on new polarized features is presented.

1.7 Thesis Outline

This thesis describes the research work that resulted in the development and validation of the Stokes polarimetry device and a classification frameworks for differentiation of melanoma using conventional cross-polarized and non-conventional image polarimetry techniques.

Prior to developing a polarimetry system and a CAD framework, the basics and principles of polarization and classification frameworks are explored. The basics of polarization and a history of polarimetry images is presented in **Chapter 2** while **Chapter 3** discuss the related machine learning and computer vision techniques related to the classification framework.

Chapter 4 is dedicated to dermoscopy modality. State of the art of CAD systems, our proposed framework, experiments and results obtained for dermoscopy modality are depicted

in this chapter.

Chapter 5 presents the framework developed for the PI modality. The PI system and the developed classification framework along with the results obtained are presented in this chapter.

Finally **Chapter 6** concludes the thesis and presents avenues for future research.

Chapter 2

Polarization Principles and Image Polarimetry

Light was thought as being non-scalar for the first time by Christian Huygens when he observed the propagation of light through crystals [94]. Through observations it appeared that light had “sides”, in the words of Newton. This new vectorial nature of light was called polarization. Later, Frensel and Arago discovered that light consisted of only two transverse components and the perpendicular components were assumed to propagate in the z direction.

The concept of polarization arises from the transverse and vector nature of electromagnetic radiation. This concept describes the resultant pattern of an electric field vector ($E(r, t)$) as a function of time (t) at a fixed location in space (r). The classical theory of polarization and the nature of light are evidence that polarization is another fundamental property of light besides coherent, frequency and intensity.

In this chapter, the polarized properties of light and the mathematical representation of these properties are explained first. This section focuses on the mathematical formulations used in dermopolarimetry. Later the polarized properties of the medium and its mathematical formulations through the Stokes representation are presented and, finally, the related work on image polarimetry systems are explained.

2.1 Polarization Properties of Light

The propagation of the optical field in the isotropic medium is described using three independent waves.

$$\nabla^2 u_i(r, t) = \frac{1}{\nu^2} \frac{\partial^2 u_i(r, t)}{\partial t^2}, \quad i = x, y, z \quad (2.1)$$

In Eq. 2.1, ν is the velocity of the oscillation and $r = r(x, y, z)$ defines a point in the Cartesian system. In this system, $u_x(r, t)$ and $u_y(r, t)$ are called transverse components and $u_z(r, t)$ is the longitudinal component of the optical field:

$$u_x(r, t) = u_{0x} \cos(\omega t - k.r + \delta_x) , \quad (2.2a)$$

$$u_y(r, t) = u_{0y} \cos(\omega t - k.r + \delta_y) , \quad (2.2b)$$

$$u_z(r, t) = u_{0z} \cos(\omega t - k.r + \delta_z) . \quad (2.2c)$$

In 1818, Fresnel and Arago carried out several experiments and concluded that the longitudinal component of light does not exist: in another words, light only contains transverse components. By assuming the direction of propagation in the z direction, the transverse components of the Eq. 2.2 are reformulated by Eq. 2.3, where $\tau = \omega t - kz$ is the propagator, E_{0x} and E_{0y} are the maximum amplitude and δ_x and δ_y are the time independent phases of the two transverse components.

$$E_x(z, t) = E_{0x} \cos(\omega t - kz + \delta_x) = E_{0x} \cos(\tau + \delta_x) , \quad (2.3a)$$

$$E_y(z, t) = E_{0y} \cos(\omega t - kz + \delta_y) = E_{0y} \cos(\tau + \delta_y) . \quad (2.3b)$$

Transverse waves of light are said to be “instantaneous” since, at optimal frequencies, the time for a wave to go through one complete cycle is only 10^{-15} s [94, 190]. Due to this feature, the electrical field vector is the result of two waves tracing a single curve almost instantaneously. The theoretical case of this curve is an ellipse (Polarization ellipse) when $\delta = \delta_x - \delta_y$ is constant over time. Equation 2.4 presents the formulation of the curve and the polarization ellipse.

$$\frac{E_x^2}{E_{0x}^2} + \frac{E_y^2}{E_{0y}^2} - 2\frac{E_x E_y}{E_{0x} E_{0y}} \cos \delta = \sin^2 \delta \quad (\delta = \delta_y - \delta_x) . \quad (2.4)$$

Figure. 2.1 represents this ellipse bounded by a rectangle. The sides of this rectangle are parallel to the coordinate axes of the ellipse with lengths of $2E_{0x}$ and $2E_{0y}$ respectively. This ellipse is tangent to the sides of the rectangle at four points (A, B, C, D). The formulations of these points are represented by Eq. 2.5, and Eq. 2.6 shows the calculation of the polarized ellipse area.

$$A : +E_{0x} \cos \delta, +E_{0y} \quad B : +E_{0x}, +E_{0y} \cos \delta \quad (2.5)$$

$$C : -E_{0x} \cos \delta, -E_{0y} \quad D : -E_{0x}, -E_{0y} \cos \delta$$

$$A = \pi E_{0x} E_{0y} \sin \delta . \quad (2.6)$$

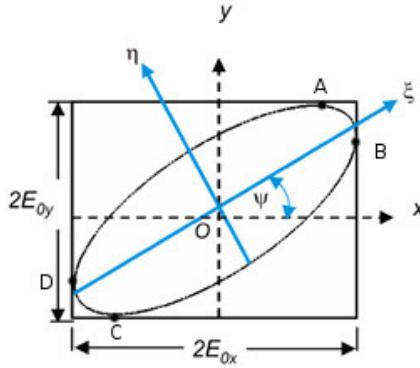


Figure 2.1: An elliptically polarized wave and the polarization ellipse.

Different values of E_{0x} , E_{0y} , and δ form various shapes of polarized ellipse and create different states of polarized light, including;

- **Linearly horizontal / vertical polarized light**

When one of the transverse components does not exist ($E_{0y} = 0$ or $E_{0x} = 0$), oscillation happens in one direction only (x or y respectively). The oscillation in x direction is called linearly horizontal polarized and in the y direction, linearly vertical polarized light.

- **Linear +45° / -45° polarized light**

When $\delta = 0$ or π , Eq. 2.4 is formulated by Eq. 2.7. In this form, light is linearly polarized with a slope of $\pm E_{0y}/E_{0x}$. In this case, if $E_{0x} = E_{0y}$, the light is said to be linearly +45° polarized for parameter $\delta = \pi$ and linearly -45° polarized for $\delta = 0$.

$$E_x = \pm \left(\frac{E_{0y}}{E_{0x}} \right) E_y . \quad (2.7)$$

- **Left / right circular polarized light**

When $\delta = \pi/2$ or $3\pi/2$, Eq. 2.4 indicates the identical standard equation of the ellipse. In this condition, if $E_{0x} = E_{0y} = E_0$, the ellipse transforms into a circle and the light is said to be circularly left polarized ($\delta = \pi/2$) or circularly right polarized ($\delta = 3\pi/2$).

- **Un-polarized light**

When the phase difference δ is unpredictable and rapidly varies in time, the light is said to be un-polarized. In this case, there is no particular polarization direction and the electrical field is equally distributed in all directions.

Besides amplitudes and phase differences, a polarization ellipse contains two other elliptical parameters, the angle of rotation ψ and the ellipticity angle χ . These two parameters are

defined using the angle α . This angle is defined as $\tan(\alpha) = E_{0y}/E_{0x}$, where α is within the limits of $0 \leq \alpha \leq \pi/2$. Using this definition, the rotation and ellipticity angle of an ellipse is formulated as:

$$\tan 2\psi = \frac{2E_{0x}E_{0y} \cos \delta}{E_{0x}^2 - E_{0y}^2} = (\tan 2\alpha) \cos \delta \quad 0 \leq \psi \leq \pi , \quad (2.8a)$$

$$\sin 2\chi = \frac{2E_{0x}E_{0y} \sin \delta}{E_{0x}^2 + E_{0y}^2} = (\sin 2\alpha) \sin \delta \quad -\pi/4 \leq \chi \leq \pi/4 . \quad (2.8b)$$

The handedness of the elliptical polarization state can be defined using the sign χ . If χ is negative, then δ is also negative and the field rotates counter-clockwise (from x to y) which leads to a left handed orientated elliptical state. On the other hand, if χ is positive, the polarization state is orientated to the right.

2.2 Stokes Parameters

In 1852, Sir George Gabriel Stokes discovered that polarization behavior could be represented in terms of observables. Sir. Stokes in an attempt to mathematically characterize un-polarized light, defined un-polarized light as light whose intensity is unaffected when a polarizer is rotated or while a retarder of any retardance value is present. In this characterization Sir Stokes defined four measurable quantities known as the Stokes polarization parameters. (See Eq.2.9).

$$(E_{0x}^2 + E_{0y}^2)^2 - (E_{0x}^2 - E_{0y}^2)^2 - (2E_{0x}E_{0y} \cos \delta)^2 = (2E_{0x}E_{0y} \sin \delta)^2 , \quad (2.9)$$

The first parameter expresses the total intensity of the optical field in terms of the total amount of horizontal and vertical linear polarization. The second and third parameters describe the amount of linearly polarized light and the fourth parameter describes the amount of left or right circularly polarized light. The total degree of polarization (DOP), the degree of linear polarization (DOLP) and the degree of circular polarization (DOCP) are calculated using Stokes parameters. Equation. 2.10 represents the Stokes vector with the parameters represented first in terms of amplitudes of the transverse components (E_{0x}, E_{0y}) and δ angle, then using different states of polarization. Here I_H and I_V stand for linear horizontal and vertical polarized, respectively and I_P, I_M, I_L, I_R represent linear $+45^\circ$, linear -45° , left circular and right circular polarized light, respectively.

$$S = \begin{bmatrix} S_0 \\ S_1 \\ S_2 \\ S_3 \end{bmatrix} = \begin{bmatrix} I \\ Q \\ U \\ V \end{bmatrix} = 1/2 \begin{bmatrix} E_{0x}^2 + E_{0y}^2 \\ E_{0x}^2 - E_{0y}^2 \\ 2E_{0x}E_{0y} \cos \delta \\ 2E_{0x}E_{0y} \sin \delta \end{bmatrix} = \begin{bmatrix} I_H + I_V \\ I_H - I_V \\ I_P - I_M \\ I_L - I_R \end{bmatrix} . \quad (2.10)$$

The four Stokes parameters are real quantities in terms of intensities and, for any state of

Table 2.1: Stokes vector for polarization states.

Polarization states	H	V	+45°	-45°	R	L	Elliptical
Stokes Vector							
$\begin{bmatrix} S_0 \\ S_1 \\ S_2 \\ S_3 \end{bmatrix}$	$\begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 1 \\ -1 \\ 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 1 \\ 0 \\ -1 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix}$	$\begin{bmatrix} 1 \\ 0 \\ 0 \\ -1 \end{bmatrix}$	$\begin{bmatrix} 1 \\ \cos 2\chi \cos 2\psi \\ \cos 2\chi \sin 2\psi \\ \sin 2\chi \end{bmatrix}$

polarized light, they always satisfy the relation shown in Eq. 2.11. The equality sign applies when completely polarized light exists and inequality applies when partially polarized light or un-polarized light exists.

$$S_0^2 \geq S_1^2 + S_2^2 + S_3^2 . \quad (2.11)$$

As mentioned above, the first element of the Stokes vector represents the total intensity of light. The DOP of light is defined as the ratio of polarized intensity per total intensity. In a similar manner, DOLP and DOCP are calculated.

$$DOP = \rho = \frac{I_{pol}}{I_{tot}} = \frac{(Q^2 + U^2 + V^2)^{1/2}}{I} , \quad (2.12a)$$

$$DOLP = \rho_l = \frac{(Q^2 + U^2)^{1/2}}{I} , \quad (2.12b)$$

$$DOCP = \rho_c = \frac{V}{I} . \quad (2.12c)$$

With respect to Eq. 2.8, the relation between rotation and the ellipticity angle of the polarized ellipse and Stokes parameters are defined by:

$$\sin 2\chi = S_3/S_0 \quad \tan 2\psi = S_2/S_1 . \quad (2.13)$$

With regard to Eq. 2.13, the Stokes representation in terms of rotation, the ellipticity angle and α is given by:

$$S = S_0 \begin{bmatrix} 1 \\ \cos 2\chi \cos 2\psi \\ \cos 2\chi \sin 2\psi \\ \sin 2\chi \end{bmatrix} = I_0 \begin{bmatrix} 1 \\ \cos 2\alpha \\ \cos 2\alpha \cos \delta \\ \sin 2\alpha \sin \delta \end{bmatrix} . \quad (2.14)$$

Table 2.1 illustrates polarization states and their relative Stokes vector. There are different ways of measuring Stokes parameters. A classic way of measuring these parameters is by passing an optical beam through two optical elements, known as the retarder and the polarizer. The

retarder is an element that changes the phase between two transverse components of the light, and the polarizer is an element that changes the amplitude of the transverse components. The incident light passes through the retarder which then advances the phase of the x component (E_x) by $\phi/2$ and retards the phase of the y component by $\phi/2$. The beam merging from the retarder (the phase shifting element) passes through the polarizer next. The polarizer has the ability to transmit the optical field through only one axes called the transmission axis. For instance, if the transmission axis of the polarizer is at θ , only the E'_x and E'_y components in this direction pass through.

Using this set-up, the first three Stokes parameters are measured by removing the retarder and rotating the transmission axis of the polarizer to the angles $\theta = 0^\circ$, $\theta = +45^\circ$ and $\theta = +90^\circ$, respectively. For measuring the final parameter, a retarder is added to the set-up with the angle of $\phi = 90^\circ$ (quarter-wave retarder), while the transmission axis of the polarizer is set to $\theta = +45^\circ$. Equation. 2.15 represents how this set-up can be used to measure the Stokes parameters. $I(\theta, \phi)$ shows the intensity produced by the retarder with a ϕ phase shift angle and the polarizer with its transmission axis at θ .

$$S_0 = I(0, 0) + I(90, 0) , \quad (2.15a)$$

$$S_1 = I(0, 0) - I(90, 0) , \quad (2.15b)$$

$$S_2 = 2I(45, 0) - I(0, 0) - I(90, 0) , \quad (2.15c)$$

$$S_3 = 2I(45, 90) - I(0, 0) - I(90, 0) . \quad (2.15d)$$

2.3 Optical Properties of Medium

Besides the nature of the light and polarized states of incident light, we are interested in checking the polarized properties of the optical elements and subjected materials. In this regard, this section is dedicated to describing the optical properties of the medium and their formulation.

2.3.1 Mueller matrix parameters

The Mueller matrix is used to describe the transfer function and interaction of any medium with polarized light. The 4×4 Mueller matrix was invented by Hans Mueller in the 1940s [94]. If the incident polarized beam interacts with a polarizing medium, its elements will change, so the emerging Stokes vector is as shown in Eq. 2.16.

$$\begin{bmatrix} S_0 \\ S_1 \\ S_2 \\ S_3 \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \\ m_{41} & m_{42} & m_{43} & m_{44} \end{bmatrix} \begin{bmatrix} S'_0 \\ S'_1 \\ S'_2 \\ S'_3 \end{bmatrix}. \quad (2.16)$$

After interaction with the medium, the polarization state of the optical beam can change in terms of:

- Amplitude of the polarization state
- Phase of the polarization state
- Direction of the orthogonal field components
- Energy of the polarized and un-polarized states by transferring energy between them.

The optical device that changes the amplitude of the transverse components of an optical beam is called a *polarizer* or *diattenuator* and the optical element that introduces a phase shift between orthogonal components is called a *retarder* (phase shifter, compensator or wave plate). The optical element that rotates the orthogonal components is called a *rotator*. Finally, the element that transfers the energy of a polarized state to an un-polarized state is called a *depolarizer*.

A medium can have similar properties to any optical elements to change the polarization states of the incident light. A medium can contain depolarization, birefringence (retardants) and diattenuation properties.

Depolarizer, as mentioned above, transfers the energy of a polarized state to an un-polarized state. If an initial state of the light is 100% polarized and the polarization degree of the existing state is less than 100%, the medium possesses a depolarization property. Depolarization is usually due to multiple scattering of photons.

Polarizer (diattenuator) changes the amplitude of orthogonal components, thus the emerging beam after using the polarizer is represented by:

$$E'_x = p_x E_x \quad E'_y = p_y E_y \quad ; 0 \leq p_x, p_y \leq 1 \quad (2.17)$$

The p_x and p_y are the amplitude attenuation coefficients along the orthogonal transmission axis. For no attenuation or perfect transmission along an orthogonal axis $p_x p_y$ is set to one, and for complete attenuation, $p_x p_y$ is set to zero. It is obvious that if one of the axis has an absorption coefficient of zero, there is no transmission along this axis and the polarizer is said to have a single transmission axis. The Mueller matrix of the polarizer,

considering the two amplitude attenuation coefficients, is given by Eq. 2.18. In the case of a non-single transmission axis, the existence of m_{44} in the Mueller matrix assures that initially elliptically polarized light remains unchanged.

$$M = 1/2 \begin{bmatrix} p_x^2 + p_y^2 & p_x^2 - p_y^2 & 0 & 0 \\ p_x^2 - p_y^2 & p_x^2 + p_y^2 & 0 & 0 \\ 0 & 0 & 2p_x p_y & 0 \\ 0 & 0 & 0 & 2p_x p_y \end{bmatrix} ; 0 \leq p_x, p_y \leq 1 , \quad (2.18)$$

A medium can have the same property and change the attenuation of orthogonal polarization states due to absorption and scattering effects.

Retarder introduces a phase shift in the transverse components of the optical field. The emerging beam after this element is given by:

$$E'_x(z, x) = e^{+i\phi/2} E_x(z, x) \quad E'_y(z, y) = e^{-i\phi/2} E_y(z, y) . \quad (2.19)$$

Accordingly, the general form of the retarder Mueller matrix is shown as:

$$M = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \cos \phi & \sin \phi \\ 0 & 0 & -\sin \phi & \cos \phi \end{bmatrix} . \quad (2.20)$$

Often, two special cases of retarder are used in optics; a *quarter-wave retarder* ($\phi = 90^\circ$) and a *half-wave retarder* ($\phi = 180^\circ$). In these elements, the phase of one of light's component is delayed with respect to the other component by a one quarter wave and a half wave respectively. The quarter-wave retarder is used to transform linearly polarized light (its axis at $\pm 45^\circ$) to circularly polarized light (right/left respectively) or circularly polarized light to linearly polarized light. The half-wave retarder has the ability to rotate the polarization ellipse. The Mueller matrix for a quarter-wave and a half-wave retarder are shown in Table 2.2.

Similarly, if a medium possesses this property (retardance), it causes a phase shift between the orthogonal components of polarized light. Birefringence is a special case of retardance that occurs due to the anisotropic structure of the medium. Retardance normally occurs due to a refractive index difference between a variety of polarized states.

Rotator rotates the orthogonal components of an optical field with an angle θ . This is an angle between E_x and E'_x . If we consider angle β as the angle between E and E_x (see Fig. 2.2), the orthogonal components of the emerging beam after the rotator are defined by Eq. 2.21. The Mueller matrix of this element is given by Eq. 2.22.

Table 2.2: Mueller matrix of quarter-wave and half-wave retarders.

Quarter-wave	Half- wave
$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}$

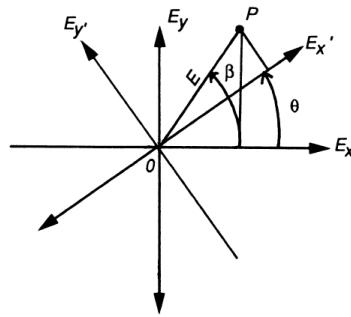


Figure 2.2: Rotation of the opticalo Polarization field components by rotator

$$E'_x = E_x \cos \theta + E_y \sin \theta \quad E'_y = -E_x \sin \theta + E_y \cos \theta . \quad (2.21)$$

$$M(2\theta) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos 2\theta & \sin 2\theta & 0 \\ 0 & -\sin 2\theta & \cos 2\theta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} . \quad (2.22)$$

The polarizer and retarder are often rotated while being used in optics. For this reason, it is useful to mention the Mueller matrix of the rotated polarizer and rotated retarder. First, assume the polarizer is rotated at the same time. The incident beam first goes through the rotator with the Mueller matrix of $M_R(2\theta)$, and then through the polarizer with the Mueller matrix (M). In this case, the emerging Stokes vector in terms of original axis is given by 2.23. Eq. 2.24 represents the Mueller matrix of the rotated polarizer.

$$S' = M_R(-2\theta)MM_R(2\theta) . \quad (2.23)$$

$$M = \frac{p^2}{2} \begin{bmatrix} 1 & \cos 2\gamma \cos 2\theta & \cos 2\gamma \sin 2\theta & 0 \\ \cos 2\gamma \cos 2\theta & \cos^2 2\theta + \sin 2\gamma \sin^2 2\theta & (1 - \sin 2\gamma) \sin 2\theta \cos 2\theta & 0 \\ \cos 2\gamma \sin 2\theta & (1 - \sin 2\gamma) \sin 2\theta \cos 2\theta & \sin^2 2\theta + \sin 2\gamma \cos^2 2\theta & 0 \\ 0 & 0 & 0 & \sin 2\gamma \end{bmatrix}, \quad (2.24)$$

The matrix represented in Eq. 2.24 is the general form. Thus, for different values of $\gamma = 0^\circ, 45^\circ$ and 90° , the matrix represent the linear horizontal polarizer, the neutral density filter and the linear vertical polarizer, respectively. The Mueller matrix of an ideal linear horizontal polarizer is shown in Eq. 2.25.

$$M = \frac{1}{2} \begin{bmatrix} 1 & \cos 2\theta & \sin 2\theta & 0 \\ \cos 2\theta & \cos^2 2\theta & \sin 2\theta \cos 2\theta & 0 \\ \sin 2\theta & \sin 2\theta \cos 2\theta & \sin^2 2\theta & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}. \quad (2.25)$$

The Mueller matrix of a rotated retarder is calculated in the same manner. Equation. 2.26 represents the general form of rotated retarder Mueller matrix.

$$M(\phi, 2\theta) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos^2 2\theta + \cos \phi \sin^2 2\theta & (1 - \cos \phi) \sin 2\theta \cos 2\theta & -\sin \phi \cos 2\theta \\ 0 & (1 - \cos \phi) \sin 2\theta \cos 2\theta & \sin^2 2\theta + \sin \phi \cos^2 2\theta & \sin \phi \cos 2\theta \\ 0 & \sin \phi \sin 2\theta & -\sin \phi \cos 2\theta & \cos \phi \end{bmatrix}. \quad (2.26)$$

A medium (tissue) can possess different optical properties at the same time. In order to obtain these properties, inverse polarimetry based on Mueller matrix decomposition is used. The Lu-Chipman decomposition method [122] is a well known inverse polarimetry analysis explained in the following.

2.3.2 The Lu-Chipman decomposition

The proposed method of Lu-Chipman [122] allows a Mueller matrix to be decomposed into the product of three matrices of diattenuation, depolarization and retardant (See Eq.2.27). In this representation, M_Δ describes the depolarizing effects of the medium, M_R shows the effects of linear birefringence and M_D includes the effect of linear and circular diattenuation.

$$M \Leftarrow M_\Delta \cdot M_R \cdot M_D . \quad (2.27)$$

$$M_\Delta = \begin{bmatrix} 1 & \vec{0} \\ \vec{P}_\Delta & m_\Delta \end{bmatrix} \quad M_R = \begin{bmatrix} 1 & \vec{0} \\ 0 & m_R \end{bmatrix} \quad M_D = \begin{bmatrix} 1 & \vec{D}^T \\ \vec{D} & m_D \end{bmatrix} . \quad (2.28)$$

Each of these matrices can be calculated and further used to extract individual medium properties such as the *diattenuation* factor D , *depolarization* Δ , *linear retardance* δ , and a circular retardance ψ . Calculation of these parameters are explained step by step in the following.

Starting with diattenuation, the 3×3 sub-matrix m_D is defined by:

$$m_D = \sqrt{1 - D^2} I + \frac{1 - \sqrt{1 - D^2}}{D^2} \vec{D} \vec{D}^T , \quad (2.29)$$

where I is the 3×3 unity matrix, \vec{D} is the diattenuation vector and D is the diattenuation value $D = |\vec{D}|$.

$$\vec{D} = \frac{1}{m_{11}} [m_{12} \ m_{13} \ m_{14}]^T , \quad (2.30)$$

$$D = \frac{1}{m_{11}} \sqrt{m_{12}^2 + m_{13}^2 + m_{14}^2} . \quad (2.31)$$

Using the diattenuation matrix, the depolarization matrix can be computed.

$$M_\Delta M_R = M' = MM_D^{-1} = \begin{bmatrix} 1 & \vec{0} \\ \vec{P}_\Delta & m' \end{bmatrix} , \quad (2.32)$$

In this matrix, \vec{P}_Δ is based on the polarizance vector \vec{P} and the sub-matrix m of the Mueller matrix M , (See Eq.2.33)

$$\vec{P} = \frac{1}{m_{11}} [m_{21} \ m_{31} \ m_{41}]^T , \quad (2.33a)$$

$$\vec{P}_\Delta = \frac{(\vec{P} - m\vec{D})}{1 - D^2} , \quad (2.33b)$$

$m' = m_\Delta m_R$. Since m_Δ is a symmetric matrix, its eigenvalues define its depolarization properties, $m_\Delta^T = m_\Delta$ and $m_\Delta^2 = m'(m')^T$. Thus m_Δ can be represented in terms of eigenvalues of $m'(m')^T$.

$$m_\Delta = \pm[m'(m')^T + (\sqrt{\lambda_1\lambda_2} + \sqrt{\lambda_2\lambda_3} + \sqrt{\lambda_3\lambda_1})I]^{-1} \times [(\sqrt{\lambda_1} + \sqrt{\lambda_2} + \sqrt{\lambda_3})m'(m')^T + \sqrt{\lambda_1\lambda_2\lambda_3}I] .$$

Once the depolarization matrix m_Δ is obtained, the sub-matrix of retardance is processed by $m_R = m_\Delta^{-1}m'$. Following the sub-matrices obtained, the depolarization power and total retardance is calculated by:

$$\Delta = 1 - \frac{|tr(m_\Delta)|}{3} , \quad (2.34)$$

$$R = \cos^{-1} \left[\frac{tr(M_R)}{2} - 1 \right] . \quad (2.35)$$

Table 2.3: Medium Characteristics

Parameters	Notations	Equation
Linear retardance	δ	$\cos^{-1}(\sqrt{[M_R(2,2) + M_R(3,3)]^2 + [M_R(3,2) + M_R(2,3)]^2} - 1)$
Optical rotation	ψ	$\tan^{-1}[(M_R(3,2) - M_R(2,3))(M_R(2,2) - M_R(3,3))^{-1}]$
Diattenuation	D	$M_D(1,1)^{-1}\sqrt{M_D(1,2)^2 + M_D(1,3)^2 + M_D(1,4)^2}$
Depolarization	Δ	$1 - \frac{ Tr(M_\Delta - 1) }{3}$

Using this method, the individual parameters of the medium can be computed. Table 2.3 summarizes these parameters, conditioned to having retardance, depolarization and diattenuation Mueller matrices. In this table, $M(i,j)$ represents the element in the i^{th} row and the j^{th} column of the matrix.

2.4 Image Polarimetry in Biomedical Field

Image polarimetry refers to different approaches to describe the propagation of polarized light, its interactions with the optical system, and its medium polarized properties. The Stokes and Mueller systems are well suited for polarimetry applications, since first they can document the full, partial and un-polarized state of light and second they provide an intensity measure and experimental data.

The Stokes and Mueller polarimetries refer to measuring the Stokes vector and the Mueller matrix, respectively. These systems can be used either to highlight the different layers of the tissue by showing diffuse light coming from deeper layers and showing the medium response with different polarized conditions (Stokes polarimetry) or to highlight some medium characteristics (Mueller polarimetry).

Although the polarization filters are employed in a polarized dermoscope (PD) in order to remove any specular reflection of light and capture the backscattered light from deeper levels of the skin, the full potential of image polarimetry in biomedicine has not been explored [87].

Currently, several researchers are pursuing innovative solutions in the field of image polarimetry. A comprehensive summary of current research in polarimetry and its problems is provided by Ghosh et al. [87]. Polarimetry systems can be used either for tissue imaging or analyzing the tissue's optical characteristics. We categorized the applied polarimetry systems into three groups: (i) partial Stokes polarimeters, (ii) Stokes polarimeters, and (iii) Mueller

polarimeters. The techniques in the first two categories are used to define the polarization state of the backscattered light from the object and the last one is used to detect the optical properties of the sample. The aforementioned groups and their related studies are discussed in the following.

2.4.1 Partial Stokes polarimeters

Beside the conventional use of cross-polarized filters in dermoscopes, the polarization and polarimetry systems have been used by several researchers as imaging techniques. This section describes the image polarimeters which were built to create the first two or three Stokes parameters.

Jacques et al [106] were among the first researchers to consider analysing polarized images rather than single image for skin imaging. In their research they proposed the use of two polarized filters. A polarizer generator was placed in the incident path in order to create the linearly polarized light illumination and polarizer analyser in the detection path. In this system the angle of the incident was chosen in a way that only backscattered light was collected with the camera. The authors proposed two image acquisitions: one with the analysing polarizer oriented parallel to the illumination and one with the analyser positioned perpendicular to the illumination. The algebraic combination of these images ($I_{pol} = \frac{I_{\parallel} - I_{\perp}}{I_{\parallel} + I_{\perp}}$) leads to an image (depolarization ratio image) which is sensitive to superficially scattered polarized light and visualizes the disruption of the normal texture of the papillary and upper reticular dermis by skin pathology [106]. The authors used these imaging techniques for the differentiation of clinical images of skin pathologies.

Another project based on the first two Stokes parameters was proposed by Anastasiadou et al. [11]. In this work the authors proposed polarimetric imaging for imaging cervical cancers. Figure. 2.3 illustrates their proposed systems. In this figure, α is the azimuth angle and the two images are acquired, I_{\parallel} when the the two polarizers are positioned parallel to each other and I_{\perp} when they are perpendicular. The proposed system consists of a white light source (150 W halogen bulb), a fiber handle, a filter wheel (three channels 450, 550, and 650 nm), a linear polarizer/ analyzer, and finally a CCD camera [11]. Using this setup, similar to the previous study, the I_{pol} image is calculated as a final image (referred to as *DOP* in the study). A classification framework of ex-vivo cervical lesions (76 patients) using the images provided by the proposed imaging system, achieved a sensitivity (SE) of 52-86% and specificity (SP) of 93%. Figure. 2.4 illustrates a sample of severe dysplasia comparing images seen under colposcopy and the proposed polarimetry system. This example was presented as a case, in which the confirmed dysplasia region was not obvious in ordinary images. However it was clearly noted in the I_{pol} image.

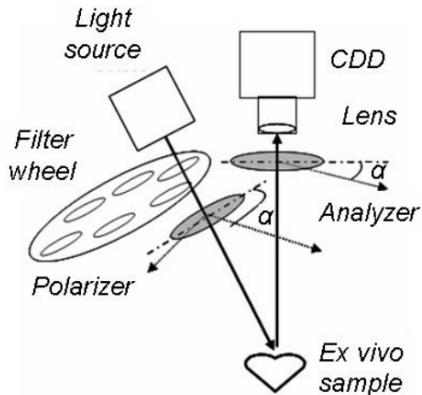


Figure 2.3: Polarimetric image system proposed by [11].

Zhao et al. [209] introduced another partial polarimetry (spectropolarimetric) imaging system to measure pathological changes of tissue birefringence and structure. Figure. 2.5 shows the proposed system by [209]. This system is composed of an incoherent white light source (halogen) illuminated in a collimated beam at an angle of 25° to the normal of the skin surface and a CCD camera along with a liquid crystal filter (LCTF) used to acquire spectropolarimetric images. The LCTF is a bandpass filter that can control the wavelength of the transmitting light and ranges from 400-720nm. The spectropolarimetric images are acquired by manually rotating the LCTF at 0° , 45° , 90° , and 135° angles. Using the acquired images the three first Stokes parameters as well as the DOLP were measured (see Eq. 2.12 and Eq. 2.15). Figure. 2.6 shows an example of the Stokes parameters and DOLP obtained for chilblain skin at 590 nm.

Recently, Tchvialeva et al. [186] proposed using polarization speckle patterns to differentiate and identify skin cancer lesions. In this research the author proposed calculating the depolarization ratio (I_{pol}) from the backscattered speckle patterns using the proposed system shown in Fig. 2.7. This system is composed of dual cameras, two diode lasers: a blue laser ($\lambda = 407$ nm) and a red laser ($\lambda = 663$ nm), a diaphragm and a beam splitter. The speckle patterns are recorded simultaneously by both cameras, one captures the light parallel to the initial polarization and the other captures the light perpendicular to the initial polarization [186]. Figure. 2.8 shows examples of polarization speckle patterns for a malignant melanoma, a squamous cell carcinoma, a basal cell carcinoma, a nevus, and a seborrheic keratosis using the blue laser. The authors evaluated their proposed systems on in-vivo 214 skin lesions containing cancer and benign lesions where they reported that statistical moments of the polarization speckle pattern is able to separate malignant melanoma from sebrroheic keratosis, basal cell carcinoma, and squamous cell carcinoma.

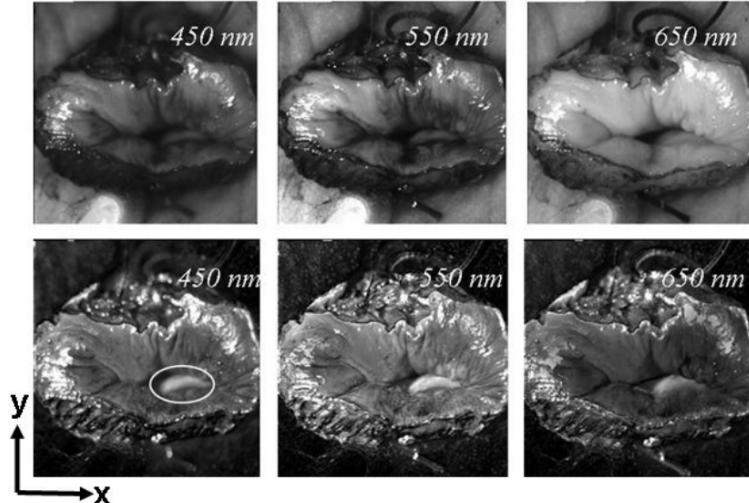


Figure 2.4: I_{\parallel} and I_{pol} , first and second row, respectively. The images were taken with $\alpha = 0$ and at different wavelengths. Images presented by [11].

2.4.2 Stokes polarimeters

A classical approach for measuring the four Stokes parameters was mentioned in Sect. 2.2 (see Eq. 2.15). As discussed in the previous section, this approach was partially used in several studies to measure the two or three first Stokes parameters [186, 209, 106, 11]. The choice of partial measurements using the aforementioned setup is because measuring the fourth parameter (S_3) is relatively more difficult and less accurate than the first three parameters. This is due to reasons such as: (i) the optical energy absorbed by the retarder should be accounted for when measuring S_3 , (ii) the retarder should be removed and replaced during the acquisition, and (iii) the perfect alignment of the transmission axis of the polarizer at 45° from the fast axis of the retarder is required. These factors have made it difficult to accurately measure the S_3 parameter, using this setup. To address this problem, Collett [53] proposed to use one optical element instead of two. Collett proposed using a circular polarizer constructed by a linear polarizer whose transmission axis is rotated at 45° with respect to the horizontal axis followed by a quarter-wave retarder which its fast axis is parallel to the horizontal direction. Using this setup, the three intensities ($I_C(\theta)$) are measured when the circular polarizer is rotated at 0° , 45° , and 90° with respect to the horizontal axis and the fourth intensity ($I_L(\theta)$) is measured by flipping the linear polarizer to the other side (i.e. the linear polarizer follows the quarter retarder) and keeping its axis parallel to the horizontal axis (i.e. $\theta = 0^\circ$) [53, 87]. The Stokes

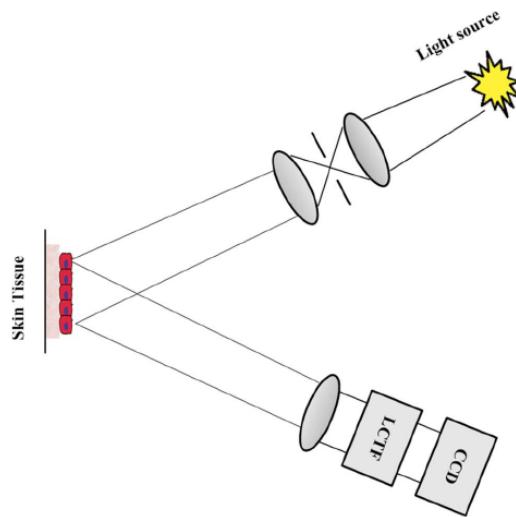


Figure 2.5: The polarimetry system developed by [209].

parameters are calculated from the four measured intensities using the following equation:

$$\begin{bmatrix} S_0 \\ S_1 \\ S_2 \\ S_3 \end{bmatrix} = \begin{bmatrix} I_C(0^\circ) + I_C(90^\circ) \\ S_0 - 2I_C(45^\circ) \\ I_C(0^\circ) - I_C(90^\circ) \\ -S_0 + 2I_L(0^\circ) \end{bmatrix}. \quad (2.36)$$

This method was used in several researches to measure the Stokes parameters of the light or backscattered light from the tissue [169, 88].

Although this method is more accurate than the classical approach, in order to improve the sensitivity of the measurement it was proposed to use polarization modulation with synchronized detection. Although such systems provide signals rather than images, they have been used in several studies to measure sample polarization properties [170, 203]. The system proposed by Wood et al. [203], using polarization modulation and a synchronous lock-in-amplifier is shown in Fig. 2.9. In such a system, usually un-polarized light passes through the mechanical chopper and then the lock-in amplifier in order to establish its overall signal intensity levels. A linear polarizer (P_1), with or without a quarter wave plate (QWP_1) is used as input optics. This set up is required in order to generate any of the four polarization states. The detection path consists of a removable quarter wave plate (QWP_2) with its fast axis oriented at -45° , a photoelastic modulator (PEM) and a linear polarizer, oriented at 45° . The linear polarizer converts the PEM polarization modulation to intensity modulation suited for photodetection. An interested

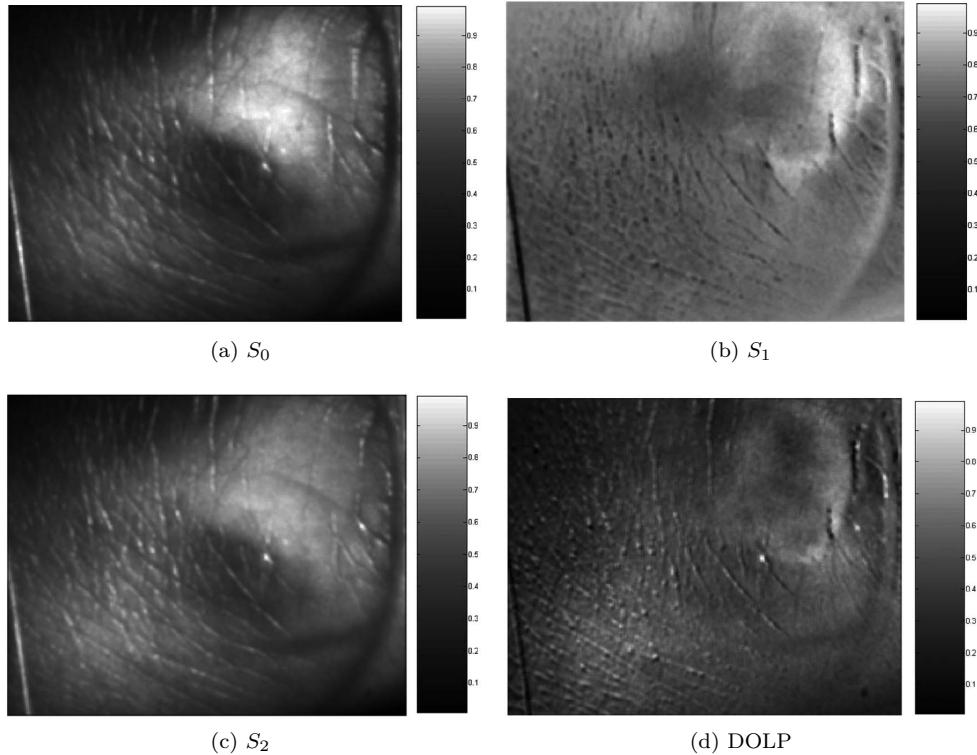


Figure 2.6: Chilblain skin sample described by the first three Stokes parameters and DOLP. Images presented by [209].

reader is guided to the original article [203] and Ghosh et al. [87] for further details regarding these systems.

Besides the above methods for creating a full Stokes polarimetry system, a full Stokes image polarimetry system was proposed by Boulbry et al. [31]. The proposed spectro-polarimetric system is based on hemispherical backscattering for analysis of superficial skin lesions. In order to remove the effect of rough skin backscattering and reflection, the proposed method is constructed with out-of-plane illumination [154]. The proposed system is composed of sixteen polarized light sources positioned over a hemisphere. Each light source generates a collimated incident beam of blue, green or red illumination at the center of the hemisphere [31]. Figure 2.10 shows the position of the illumination source tubes (illuminators) over the hemisphere and the elements within each tube (a tri-color light emission diode, a thin film polarizer and a lens).

A Stokes vector imaging is positioned on the shell at an oblique angle to the sample normal and

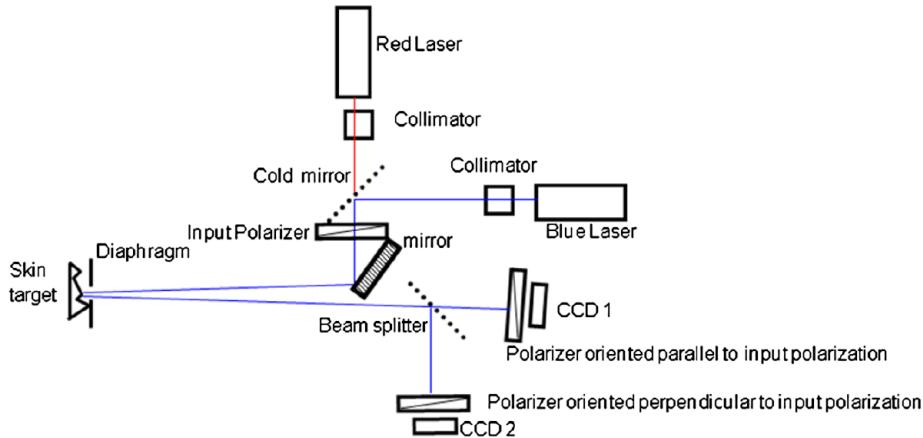


Figure 2.7: The laser speckle device proposed by [186].

consists of a 12 bits camera, two liquid crystal variable retarders (LC) and a fixed polarizer. Using seven retardance combinations of rotated crystal retarders, the four Stokes vectors were obtained [31] via a least square solution, from which the polarized and unpolarized intensity of the light source as well as the parallel and perpendicular part of the polarized intensity were measured. The method was only tested on a actinic keratosis and, unfortunately, was not tested on further skin lesions.

2.4.3 Mueller polarimeter

As explained earlier, Stokes vectors are not enough for measuring medium properties, so additional measurements and analysis is required. The Mueller matrix provides these additional measurements at the cost of a more complicated setup. This matrix and its parameters were explained in Sect. 2.3.1. Sixteen elements of the Mueller matrix are measured via Direct Current (DC) sequential static measurements or Alternative Current (AC) modulation based measurements [87]. In the latter approach, the four states of polarization (linear polarization at 0° , 45° , 90° and right/left circular polarization L and R) are applied individually as the input state, and the Stokes vector corresponding to each input is measured as an output. The 16 elements of the Mueller matrix are then constructed using the measured Stokes vectors (see Eq. 2.37) [89].

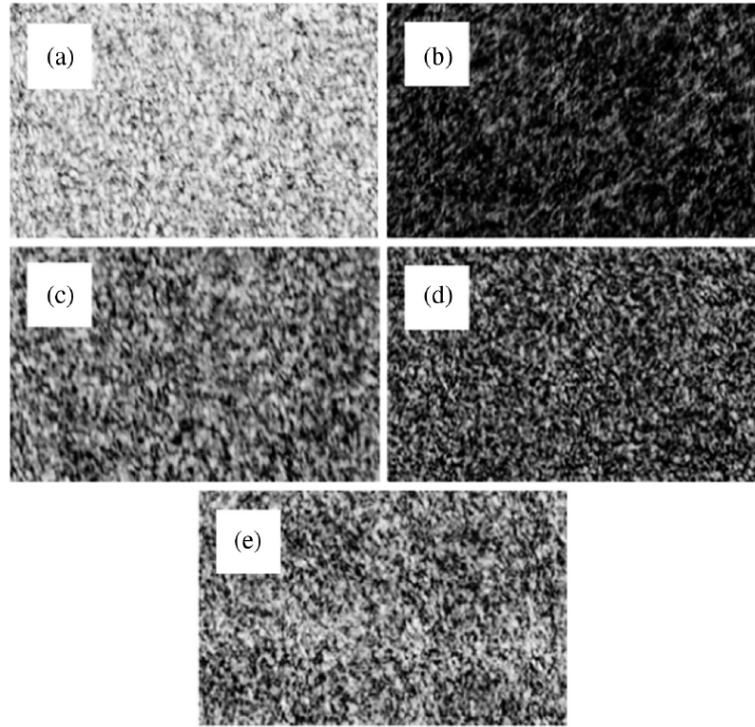


Figure 2.8: Examples of polarization speckle patterns of skin lesions using a blue laser (407 nm); (a) A malignant melanoma, (b) A squamous cell carcinoma, (c) A basal cell carcinoma, (d) A nevus, and (e) seborrheic keratosis. Images are taken from [186].

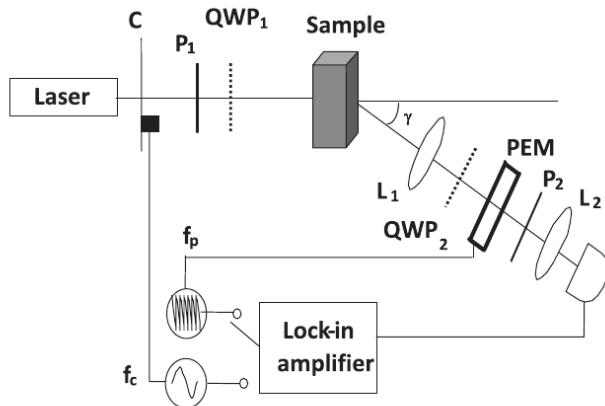


Figure 2.9: Stoke polarimetry system using polarization modulation and a lock-in-amplifier. The image is taken form [203].

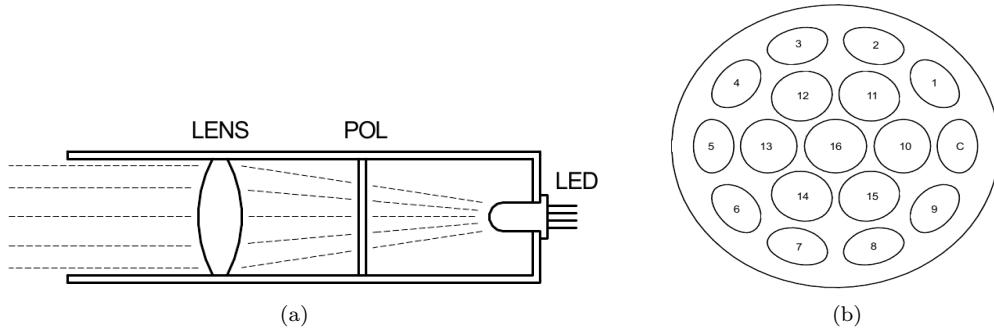


Figure 2.10: (a) The illumination tube and (b) Sixteen illumination tubes and the camera C , on the surface of the hemisphere with respect to the normal of the sample. Images are taken from [31].

$$M = \begin{bmatrix} \frac{1}{2}(I_H + I_V) & \frac{1}{2}(I_H - I_V) & I_P - M(1, 1) & I_R - M(1, 1) \\ \frac{1}{2}(Q_H + Q_V) & \frac{1}{2}(Q_H - Q_V) & Q_P - M(2, 1) & Q_R - M(2, 1) \\ \frac{1}{2}(U_H + U_V) & \frac{1}{2}(U_H - U_V) & U_P - M(3, 1) & U_R - M(3, 1) \\ \frac{1}{2}(V_H + V_V) & \frac{1}{2}(V_H - V_V) & V_P - M(4, 1) & V_R - M(4, 1) \end{bmatrix}, \quad (2.37)$$

where the four input states are denoted with H for 0° , V for 90° , P for 45° , and R for right circular polarized. The Mueller matrix elements are represented by $M(i, j)$ with respect to their row and column. This approach has been used by [89, 90].

A dual rotating retarder approach is another way to generate the modulation-based Mueller matrix polarimeters [87, 93, 16, 182]. In this approach, the incident polarization states are generated by the polarized state generator (PSG) unit which contains a linear polarizer and a rotating linear retarder with retardance and angular speed of σ_1 and ω_1 , respectively. The backscattered light from the sample are then analyzed by a polarized state analyzer (PSA) unit. This unit contains a rotating linear retarder (retardance of σ_2 and angular speed of ω_2) and a fixed linear polarizer in sequence. In this set-up the axis of the polarizer and analyzer are kept parallel and the retardation of the two retarders is chosen to be the same $\sigma_1 = \sigma_2 = \pi/2$, while their angular frequencies differ from each other by five times ($\omega_1 = \omega$, $\omega_2 = 5\omega$). The rotation of the retarders at different rates results in a modulation of a detected signal and encodes all the sixteen Mueller matrix elements into the amplitude and phases of 12 frequencies in the detected intensity signal. The detected signal is Fourier analyzed and the Mueller matrix elements are constructed from the Fourier coefficients [87, 93, 16, 182].

Dubreuil et al. [68] introduced another modulation-based Mueller matrix polarimeter, known

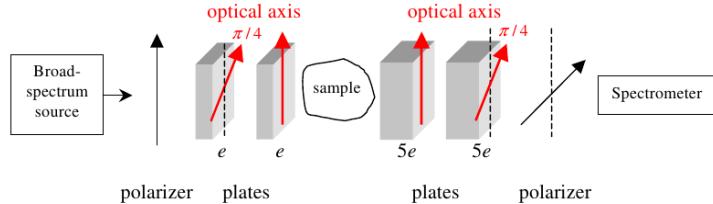


Figure 2.11: Snapshot Mueller polarimeter for the configuration, [68].

as the Snapshot Mueller matrix. The Snapshot system contains two linear birefringence retarders (wave plates) and a linear polarizer in the PSG unit and two birefringence retarders and a fixed linear polarizer (perpendicular to the one in the PSG unit) at the PSA unit (see Fig. 2.11). This technique measures the 16 elements of the Mueller matrix simultaneously by using wavelength polarization coding and decoding. The resultant spectral signal is stored by the spectrometer and is Fourier analyzed in order to achieve the 16 elements of the Mueller matrix. In this system the PSG unit contains a linear polarizer at 0° and two wave plates with their axis rotated at 45° and 0° , respectively. The PSA unit contains two wave plates with their axis at 0° and 45° , respectively, and a linear polarizer at 90° . One crucial parameter in this system is the choice of thickness of the birefringence plates. The author proposed to having the birefringence in the PSA unit five times thicker than those in the PSG.

It should be noted that the Mueller polarimeters mentioned so far are suited for point polarimetry and are not applicable for large area imaging. In this regard, several approaches are proposed and explained in the following.

Antonelli [13] proposed three Mueller image polarimeters in her phd thesis; full Mueller imaging using focused illumination, Fourier space image polarimetry, and real space full Mueller imaging. The latter system was proposed for thick tissue imaging while the two former, ones were designed for thinner tissue. The first proposed instrument is shown in Fig. 2.12. This system was implemented according to the system described by [102]. Hielscher et al. [102] developed the Mueller imaging system for analyzing the intensity patterns of backscattered light from the turbid media.

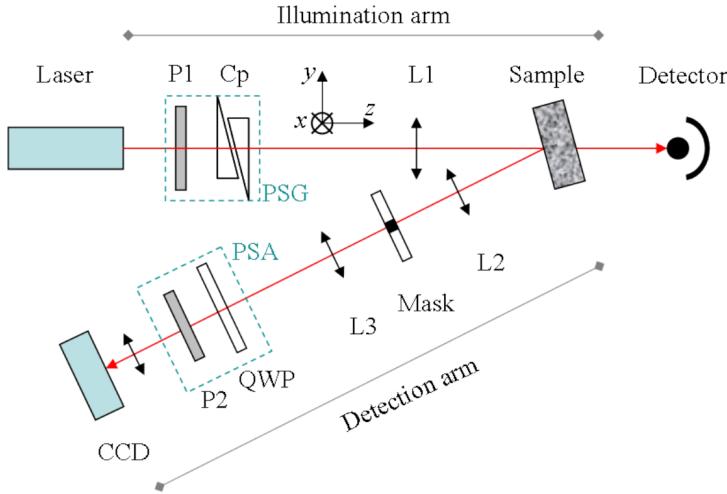


Figure 2.12: Full Mueller polarimetry system with focused illumination. The image is taken from [13].

The PSG unit of this system is composed of a linear polarizer (P_1) and a Babinet-Soleil-Bravais compensator (CP) [13], while the PSA unit is composed of a quarter wave plate (QWP) and a linear polarizer (P_2). The compensator retardance is set to either 180° , or 90° . The former is used to generate the linearly polarized states, and the latter to generate circular polarized states. In the PSA unit the QWP is set parallel to P_2 to measure the linear polarized state and it is rotated by 45° to measure circular polarized states. Using this setup, 36 raw images, corresponding to all the combinations of the polarization states possible in the PSG and PSA units, are acquired to complete the Mueller matrix. The final Mueller matrix, as a combination of the acquired 36 images, is shown in Fig. 2.13. Here O refers to an unpolarized state which is generated by the average of the two orthogonal state (i.e. H and V or P and M or L and R). For each pair, for instance HV , the first element represents the polarization state of the PSG unit, while the polarization state of the PSA unit is represented by the second element.

The second system, Fourier space image polarimetry, is shown in Fig. 2.14. In this system, the PSG unit consists of a half wave plate (HWP), followed by a linear polarizer and QWP and the PSA unit contains one QWP followed by a linear polarizer. Table 2.4 illustrates the existence and rotation of each of the optic elements in the PSG and PSA units for creating different polarization states. It should be noted that the polarization states generated in the PSG unit are independent from those generated in the PSA unit.

Using the 36 raw images with different polarization states, the author also proposed using a so called polarimetric matrix (6×4), whose columns are the components of the normalized Stokes

$M_{11}: (OO)$	$M_{12}: (HO-VO)/2$	$M_{13}: (PO-MO)/2$	$M_{14}: (LO-RO)/2$
$M_{21}: (OH-OV)/2$	$M_{22}: (HH+VV)/4 - (HV+VH)/4$	$M_{23}: (PH+MV)/4 - (PV+MH)/4$	$M_{24}: (LH+RV)/4 - (LV+RH)/4$
$M_{31}: (OP-OM)/2$	$M_{32}: (HP+VM)/4 - (HM+VP)/4$	$M_{33}: (PP+MM)/4 - (PM+MP)/4$	$M_{34}: (LP+RM)/4 - (LM+RP)/4$
$M_{41}: (OL-OR)/2$	$M_{42}: (HL+VR)/4 - (HR+VL)/4$	$M_{43}: (PL+MR)/4 - (PR+ML)/4$	$M_{44}: (LL+RR)/4 - (RL+LR)/4$

$O = \times$ $H = \leftrightarrow$ $P = \nearrow$ $L = \curvearrowleft$
 $V = \downarrow$ $M = \searrow$ $R = \curvearrowright$

Figure 2.13: Mueller matrix as combination of 36 raw images acquired by full Mueller imaging with focused illumination setup [13].

Table 2.4: Setting for the PSG and PSA units for generating 6 polarization states in each unit [13].

Polarization states	PSG			PSA	
	HWP	P	QWP	QWP	P
H	0°	0°	removed	removed	0°
V	45°	90°	removed	removed	90°
P	45°	45°	removed	removed	45°
M	45°	315°	removed	removed	135°
R	45°	90°	45°	45°	0°
L	45°	90°	-45°	-45°	90°

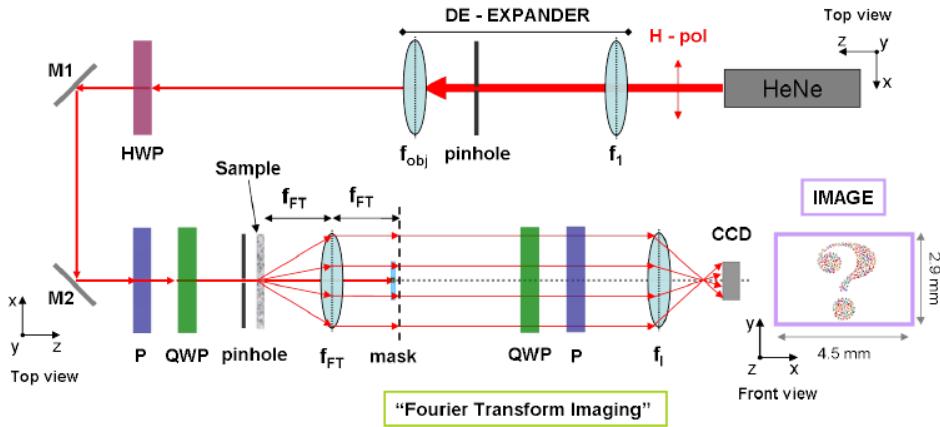


Figure 2.14: Fourier space Mueller polarimetry system proposed by [13].

vector and the lines are the input polarization states (see Fig. 2.15). Similar to the Mueller matrix (see Fig. 2.13), the polarization state of the PSG unit is represented by the first element while the second element shows the polarization state of the PSA unit.

The third system, real space full Mueller imaging, similar to [61] uses two LC to eliminate the need of manual rotation of the optical parts in the PSG and PSA units (see Fig. 2.16). This system was used for imaging ex-vivo cone biopsies.

In this system, the PSG unit consists of a linear polarizer and two LC elements and the PSA unit contains the same elements in reverse order (i.e. two LC and then a linear polarizer). Using different rotation angles for the two LC allows to obtain the four polarization states for the PSG and PSA units (common choice of angles, $\theta_1 = 45^\circ$, and $\theta_2 = 0^\circ$). Consider the four incident polarization states of the PSG unit as 4×4 matrix W (polarization states as columns) and the four basic states of the PSA unit as 4×4 matrix A (polarization states as rows). Then the 16 raw data are the elements of a 4×4 real matrix B (measurement matrix) such as [61, 13]:

$$B = AMW , \quad (2.38)$$

where M is the Mueller matrix of the sample. Having the initial knowledge of the A and W matrices and measuring the B matrix, the Mueller matrix M is calculated. The two matrices of A and W are known through the calibration procedure [54].

Besides the aforementioned systems, Mueller polarimetry was used in several other studies to capture images of human and animal tissues [191, 118].

Li et al. [118] proposed another ex-vivo instrument to characterize muscle depolarization

Output Stokes vector			
I(.)	$Q(.) = \frac{.H - .V}{.H + .V}$	$U(.) = \frac{.P - .M}{.P + .M}$	$V(.) = \frac{.R - .L}{.R + .L}$
H	$\frac{HH - HV}{HH + HV}$	$\frac{HP - HM}{HP + HM}$	$\frac{HR - HL}{HR + HL}$
V	$\frac{VH - VV}{VH + VV}$	$\frac{VP - VM}{VP + VM}$	$\frac{VR - VL}{VR + VL}$
P	$\frac{PH - PV}{PH + PV}$	$\frac{PP - PM}{PP + PM}$	$\frac{PR - ML}{PR + ML}$
M	$\frac{MH - MV}{MH + MV}$	$\frac{MP - MM}{MP + MM}$	$\frac{MR - ML}{MR + ML}$
R	$\frac{RH - RV}{RH + RV}$	$\frac{RP - RM}{RP + RM}$	$\frac{RR - RL}{RR + RL}$
L	$\frac{LH - LV}{LH + LV}$	$\frac{LP - LM}{LP + LM}$	$\frac{LR - LL}{LR + LL}$

Figure 2.15: Polarimetric matrix defined with Fourier space imaging [13].

properties. Comparing relax and stretch muscles and polystyrene solution, the authors concluded that the polarization properties of a muscle are clearly different from the polystyrene solution. However, despite clear changes in raw polarization references, muscle stretching shows minimal changes in terms of its polarization properties.

Twietmeyer et al. [191] proposed a Mueller polarimetry system for in-vivo imaging of the retina and tested their proposed method on 15 normal subjects. The authors proposed adjusting the available scanning laser polarimeter GDx by using two LC in the PSG and PSA units in order to change the polarization states. Although no subjects with retinal diseases were evaluated, the authors concluded that the proposed system is well suited for monitoring the thickness of the nerve fiber and Henle layers and can be used as an indicator of the presence and progression of glaucoma [191].

2.5 Conclusion

The basics of polarization, how to exploit polarization and its parameters in bio-imaging and skin screening and the recent state of the art techniques were discussed. It was shown that

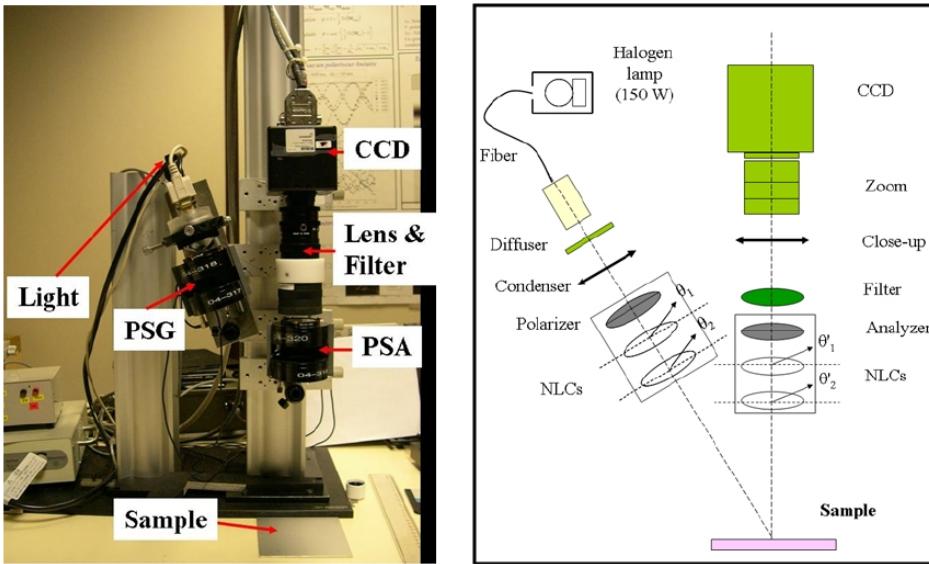


Figure 2.16: The real space Mueller imaging system proposed by [13].

partial or full Stokes and Mueller polarimetry allow screening which is of main interest. However as it was discussed, Mueller and full polarimetry systems are difficult to adjust for in-vivo screening. Therefore as the first step, we consider to exploit partial polarimetry and develop our PD accordingly. This device is further explained in Chap. 5.

Chapter 3

Machine Learning and Classification Principles

3.1 Introduction

Machine learning is refers to a set of methods that can automatically detect patterns in data and use them in decision making and future data prediction [139]. Machine learning techniques are usually divided into two groups: descriptive (or unsupervised) and predictive (or supervised) learning.

Unsupervised learning intends to find “interesting structures” in a given data without any additional information about the expected output. These techniques formulate their problems as unconditional density estimation and multivariate probability models. Clustering approaches and dimension reduction methods, Principal Component Analysis (PCA), and graph structures are two examples of unsupervised learning.

Supervised learning intends to find a mapping $f(\cdot)$ which relates a set of inputs x to a set of outputs y . The learning is comprehended using a set of N samples and their labels $D = \{(x_i, y_i)\}_{i=1}^N$, called the training set. The samples x_i can vary from 2-D points in Cartesian coordinates to more complex forms such as images, sentences, time series, graphs, etc. Similarly, the labels y_i can be represented by any structure, however, in most methods they are considered to be either categorical or numerical. Supervised learning is known as a classification problem when y_i is categorical; and when y_i is a real value the problem is referred to as regression [139].

In this research, we are interested in supervised learning, and the classification problem in particular. Classification aims to map a set of inputs to a set of outputs, where the outputs are divided into different classes, $y \in \{1, \dots, C\}$. If $C = 2$, it is a binary classification problem,

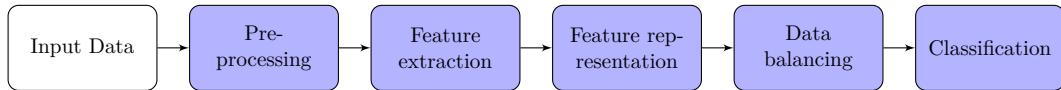


Figure 3.1: General classification framework.

while $C > 2$ makes it a multiclass classification. This project aims to separate melanoma from all other pigmented skin lesions, thus addressing a binary classification problem.

Classification has numerous applications in real life. Some of the most common and challenging areas include email spam filtering, data mining, face detection and recognition, document classification, and the problem of cancer detection among others. In order to solve a classification problem, a framework consisting of several generic steps is required. Figure 3.1 shows such a general framework whose steps are described in the following sections of this chapter. Due to the broad studies of Computer-Aided Diagnosis (CAD) systems of melanoma using dermoscopic images, in comparison to Stokes polarimetry, each step covers the state of the art related to conventional dermoscopic approaches.

3.2 Preprocessing

Data preprocessing is an essential and important step in machine learning. This step ensures the quality of the data before further analysis and serves to compensate their imperfections. Thus, preprocessing acts as a foundation and assures precision in the developed framework. In our research, samples are represented by images that must be preprocessed prior to further analysis to account for artifacts and variations in the images acquisition. Moreover, despite that some studies treated image segmentation as an individual step [19, 35, 38], we consider it as part of the preprocessing step, alongside image enhancement, denoising, and hair removal.

3.2.1 Image enhancement

Image enhancement is a set of adjustment processes on an image that make it more suitable for a specific application. Histogram equalization, gamma transformation, contrast and edge enhancement, illumination correction, white and color balancing, color calibration, as well as denoising belong to image enhancement techniques. In the field of skin imaging, among the aforementioned techniques, median filtering is mostly used to suppress noise, bright spots, reflections and small pores on the skin [49, 28, 168]. While this technique removes certain noise and artifacts, it also smooths the texture and border of the lesion, which is undesirable. From the lesions diagnosis point of view, perhaps the most important enhancement concerns color calibration and correction. This operation aims to recover real colors of an image, allowing

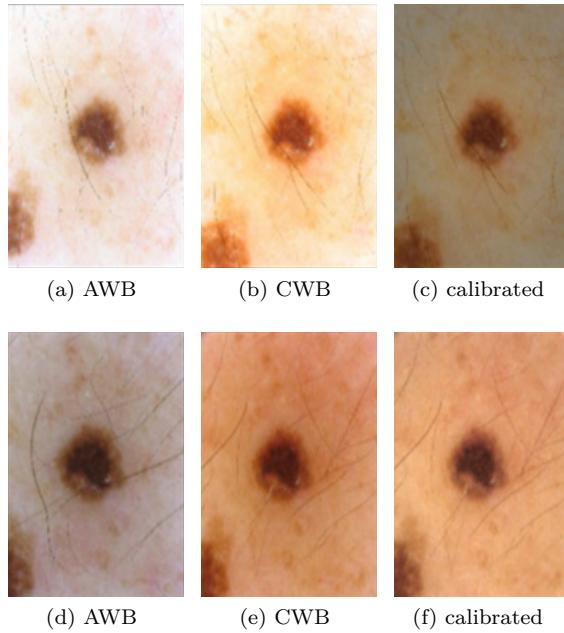


Figure 3.2: Variation in color depending on AWB, CWB and JPEG color calibration between images taken from two different dermoscopes. Images in the first row were taken with a Canon-A640 polarized dermoscope and the images in the second row were acquired with a Canon-G7 polarized dermoscope. Images are taken from [151].

for more reliable use of color information in manual and automatic diagnosis [112]. The need for color correction while using the Joint Photographic Experts Group (JPEG) format as opposed to RAW images was recently highlighted in [151, 200, 45]. Figure 3.2 shows auto-white-balance (AWB), custom-white-balance (CWB) and JPEG color calibrated dermoscopic images acquired with two polarized cameras, while Fig. 3.3 demonstrates the importance of calibrating RAW data in comparison with JPEG dermoscopic images.

3.2.2 Artifacts removal

Artifacts removal refers to noise removal in most image processing applications. However, in the field of skin imaging, it generally refers to elimination of hair, skin pores, ruler marks, air bubbles and specular reflections (see Fig.3.4). Among these operations, hair removal is the most common and necessary step. If a lesion is occluded by hairs, their removal is essential to achieve correct segmentation and texture analysis. To avoid digital hair removal, the patients are usually asked to shave before image acquisition.

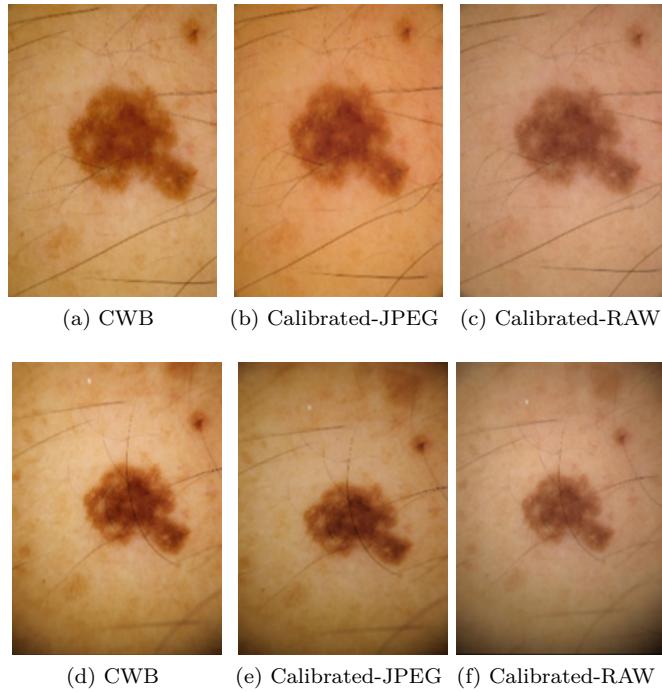


Figure 3.3: Variations in color between CWB, calibrated JPEG and calibrated RAW images obtained with 2 different polarized dermatoscopes. Images in the first row were acquired with Canon-G9 polarized dermatoscope, while those in the second row were acquired with a Canon-5D polarized dermatoscope. Images were taken from [151].

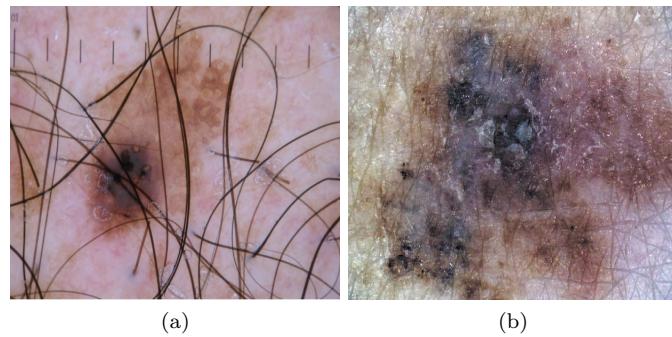


Figure 3.4: Artifacts in dermoscopic images: (a) hair, air bubbles and ruler marks, (b) specular reflection. The images are taken from Zhou et al. [211] and Gutenev et al. [97], respectively.

A hair removal algorithm commonly consists of two steps: hair detection and hair restoration (or “inpainting”) [112]. Inpainting fills the image space previously occupied by hairs with estimated intensity/color values. Considering that original colors and borders of a lesion play an essential role in lesion diagnosis, intensity estimation of missing pixels is crucial and the best inpainting method should be applied. Lee et al. [115] proposed the first hair removal algorithm, DullRazor®, for dermoscopic images. Kiani and Sharafat [111] improved this algorithm to remove light-colored hairs as well as dark.

Considering hair detection as a pixel classification problem, supervised learning was also applied in some studies [62, 197]. A recent survey on the topic was published by Abbas et al. [1], where he reviewed the existing methods and proposed a broad classification based on the inpainting algorithms employed [112]: linear interpolation techniques [115, 78, 175, 141], inpainting by nonlinear PDE-based diffusion algorithms [50, 21, 205, 3] and exemplar-based methods [211, 198, 2]. The authors [1, 7] also proposed their own inpainting method based on fast marching inpainting. Figure 3.5 shows an example demonstrating the results they achieved using different inpainting techniques.

3.2.3 Image segmentation

Image segmentation aims to decompose the image into meaningful parts with respect to a unique application. This technique uses image information such as grey level, texture, and color to divide the image into non-overlapping parts. Based on the information used, segmentation techniques are divided into four main groups: region-based, edge-based, clustering-based, and texture-based methods. Segmentation is also achievable via supervised learning.

In automatic detection of melanoma, border delineation of the lesions is achieved by segmentation. This is a challenging task due to variations in color, size, shape and texture of pigmented skin lesions (PSLs), as well as occlusions and artifacts. In addition to the aforementioned challenges, segmentation algorithms face a problem of ground truth. It is very difficult to have a unique ground truth because dermatologists do not need to delineate lesion borders to make a diagnosis, and moreover, their individual delineations may vary significantly (see Fig. 3.6). So, normally the ground truth is generated as a fusion of different delineations done by experts. Despite these challenges, numerous methods were proposed by the research community to tackle this problem. Good comparisons of segmentation methods for dermoscopy images were published by Silveira et al. [179], Celebi et al. [42], and Ferreira et al. [75]. The proposed methods can be categorized based on different criteria, for instance, technical properties, level of automation or their complexity. A review of all the proposed methods is beyond the scope of this research. However, an overview of available reviews and certain approaches is presented in the following.

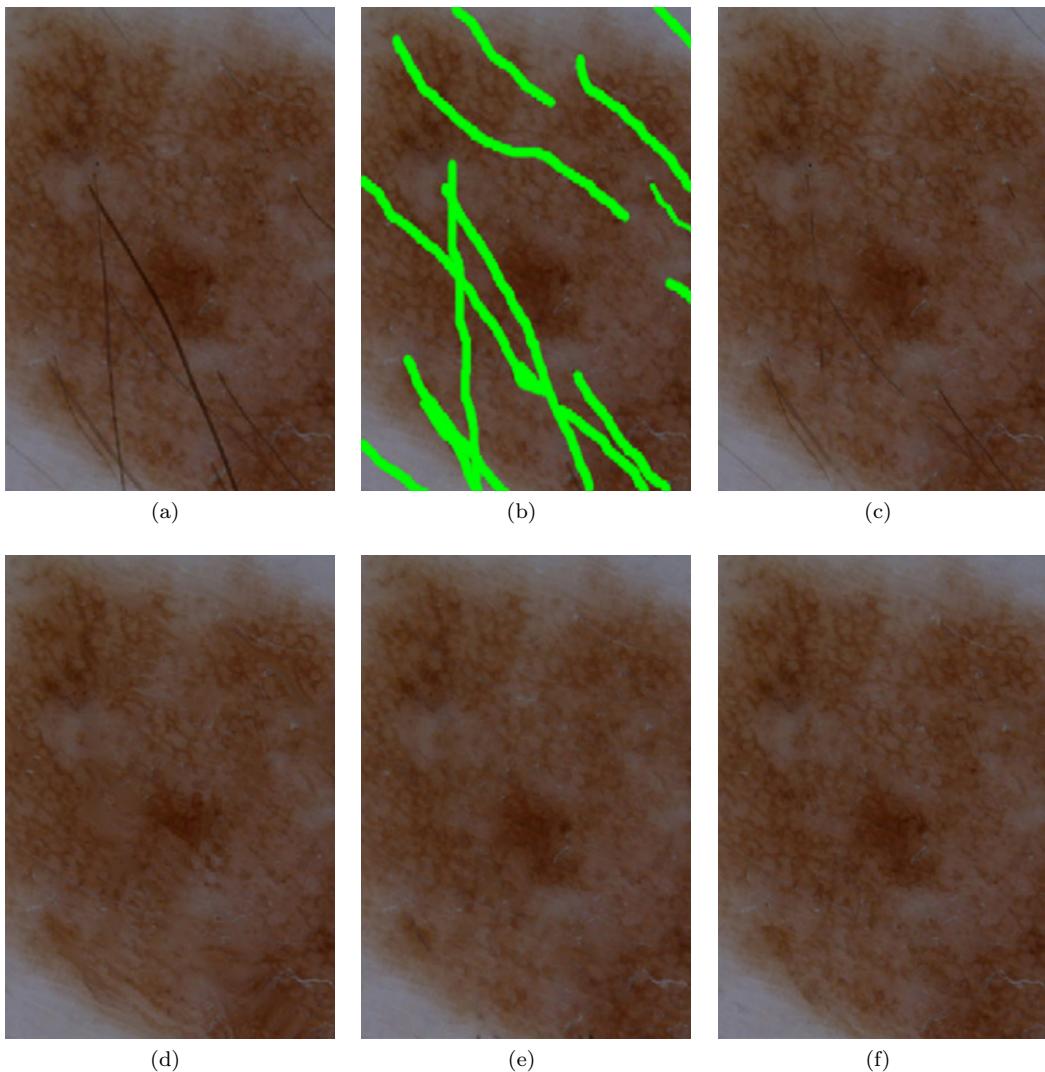


Figure 3.5: Comparison of state of the art inpainting methods for hair removal algorithms. (a) Original image, (b) Highlighted hair mask, (c) DullRazor, (d) PDE non-linear inpainting, (e) exemplar-based inpainting, and (f) fast marching inpainting. The images are taken from [1].

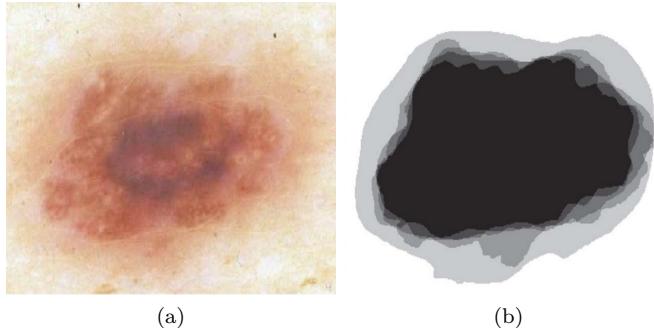


Figure 3.6: Example of manual delineation of clark nevus by 5 dermatologists. (a) Dermoscopy image, (b) Delineation by 5 dermatologists. Images are taken from Iyatomi et al. [103].

Thresholding was one of the first approaches applied to lesion segmentation. This technique was applied in a single color channel at first [98, 78] and then evolved to more sophisticated approaches such as iterative thresholding [153], type-2 fuzzy logic based thresholding [207], fusion of thresholds [37, 41, 36], hybrid [85], local entropy [8] and color histogram thresholding [146].

Among the many segmentation approaches developed, a variety of methods were employed such as clustering [174, 212, 137, 119], active contours [206, 213, 72], supervised learning [62, 197, 199, 217], and dynamic programming [4], just to name a few [112].

It is clear that without a common dataset, a comparison of the proposed methods cannot be made. In this regard, some studies surveyed a comparison of several methods on their own datasets [99, 178, 44, 43, 75].

Six segmentation methods (split-and-merge, center split, multiresolution, fuzzy c-mean, PCT/median cut and adaptive thresholding) were compared in [99], where the two latter approaches outperformed the others [112]. Silveira et al. [178] also made a comparison of 6 different methods including the level set, adaptive thresholding, expectation-maximization level set, fuzzy-based split-and-merge, adaptive snakes and Gradient Vector Flow (GVF) algorithms. In their experiments, the best performance was achieved by the adaptive snakes algorithm.

Celebi et al. [44, 43], introduced and compared statistical regional merging (SRM) with optimized thresholding, orientation-sensitive c-means [174], GVF, a dermatologist-like tumor extraction (DTEA) algorithm [103] and a JSEG algorithm [40]. Their results indicated the superiority of SRM, followed by DTEA and JSEG [112]. In another recent study, Ferreira et al. [75] reported that the GVF snakes outperformed the automatic thresholding, k-means, mean-shift, region growing and watershed algorithms. These comparisons still do not provide unified results, firstly due to the different datasets and ground truth employed, and secondly, to the different evaluation metrics [112].

3.3 Feature Extraction

Feature extraction refers to the process of gathering a set of characteristics from samples that meaningfully and efficiently describe the most important information needed for further data analysis and classification. The need for feature generation in the image processing field originated from our inability to use raw data. Even for a small 64×64 image, having all the pixels as features results in 4096 feature dimensions which is too much for most classifiers [187]. Thus, new features need to be generated from sample images.

Numerous approaches have been proposed for feature extraction in the literature. Based on the characteristics of the desired features, these approaches can be divided into four main categories; shape, color, texture, and edge. Furthermore, they can be categorized into two general categories of pixel-wise and region-wise features. The former means that features are extracted at each pixel, and the later refers to a descriptor describing a region (i.e. histogram, percentile and moment are region-based). The edge features created by convolving the image with an operator such as sobel and prewitt, belong to pixel-wise features [116]. Some texture features such as the Gabor filter [83, 57], wavelet, the Grey-Level Co-Occurrence Matrix (GLCM) [100], fractal analysis [26] and the Local Binary Pattern (LBP)-map extract features for each pixel. However, others, such as Histogram of Oriented Gradients (HOG) [56], histogram of Local Binary Pattern (LBP)-maps [142], the Scale-Invariant Feature Transform (SIFT) [120], and Speeded Up Robust Features (SURF) [24], generate descriptors which are the most representative of a region. Color features, such as statistics and color histograms, and shape features, such as asymmetry and thinness ratio, belong to this category as well.

Among the aforementioned features, shape and color features have been widely adapted for classification of melanoma using dermoscopy images. These features simulate the most, characteristics of diagnostic rules such as “ABCD”. An extensive review of the feature extraction methods applied, using dermoscopic images is discussed in the following chapter (see Chapt. 4, Sect. 4.4.3).

3.4 Feature Representation

The previous section described why there is a need for feature extraction. This section discusses the reasons behind feature representation.

The “Curse of dimensionality”(correlated and irrelevant features) [187] has always been a challenge in the field of classification and pattern recognition. Besides, high computational cost and over-fitting, learning with irrelevant features reduces the precision and performance of the system, hence, it is essential to learn the most representative and uncorrelated features. The goal of feature representation is to create a new feature space either by reducing the original

dimensionality to more representative and uncorrelated dimensions, or by modelling it into a new feature space.

Sequential Forward Feature Selection (SFFS), Sequential Backward Feature Selection (SBFS) [76], and Minimum Redundancy Maximum Relevance (mRMR) [145] are feature selection methods that pick only the most relevant feature dimensions from the original space. Linear Discriminant Analysis (LDA) [133] and PCA are linear dimension reduction approaches, that linearly project data into a new subspace with lower dimensionality. Linear mapping of the data is not always achievable and in some cases, nonlinear projection of the feature space to lower subspaces is required. Kernel PCA (non-linear PCA) [138] is one of the methods used for non-linear mapping, among others.

The aforementioned methods reduce the dimensionality in one way or another, however, there are other approaches that can map feature space to a new separable space, e.g. Bag of Words (BoW) and Sparse Coded Features (SCF). The former tries to find similar patterns in the feature space and perform the mapping based on their clusters (“visual words”). The latter maps the feature space to a sparse space in which each sample is defined by a set of few elements. A more detailed explanation of the above methods is presented in the following.

Sequential Forward/Backward Feature Selection (SFFS/SBFS) are greedy algorithms that start with an empty or full set and sequentially add or remove the best or worst features to the set, respectively. These suboptimal methods obtain a chain of nested subsets of features [76]. Nested subsets are the main drawback of these approaches, since once a feature set is added or subtracted, it can not be discarded or retained.

Minimum Redundancy Maximum Relevance (mRMR) is another feature selection approach that chooses attributes maximizing their relevance to or dependency on the distribution to a target class c , while keeping their redundancy to a minimum [145]. This technique is based on the idea that “the m best features are not the best m features”, meaning that, besides selecting the best relevant features, we have to make sure that they are not highly dependent on each other.

The relevance of a feature set is a measure of its efficacy with respect to the target class. In simple terms, it determines how well a variable discriminates between classes [15], which makes it a measure between the feature and the class. Several criteria can be used to measure the relevance of a feature set such as correlation and mutual information, among others [15]. Redundancy criteria measures the similarity between the attributes distribution in order to find mutually exclusive features. Similar to relevance, different criteria can be used to measure redundancy (e.g., redundancy fit criterion and sign test) [15].

Linear Discriminant Analysis (LDA) (also called Fisher Discriminant Analysis) applies

a supervised linear transformation to map the original space to a new one with a low dimensionality. This approach searches for the subspaces in the underlying space that best discriminates between classes rather than those that best describe the data [133]. Considering a feature space, LDA creates a linear combination of independent features while preserving the largest mean difference between the desired classes and minimum mean distance within each classes. Formulating the problem, two measures are defined:

- (1) within-class scatter matrix, given by

$$S_w = \sum_{j=1}^c \sum_{i=1}^{N_j} (x_i^j - \mu_j)(x_i^j - \mu_j)^T,$$

where x_i^j is the i^{th} sample of class j , μ_j is the mean of class j , N_j the number of samples in class j and C the number of classes.

(2) between-class scatter matrix, give by

$$S_b = \sum_{j=1}^c (\mu_j - \mu)(\mu_j - \mu)^T,$$

where μ represents the mean of all the classes. The final goal is to maximize the between-class measure (S_b) while minimizing the within-class distance (S_w). This is possible by maximizing the ratio $\det|S_b|/\det|S_w|$ [133].

Principal Component Analysis (PCA) is an unsupervised orthogonal linear transformation approach intended to search out subspaces in which the variation of the data is maximized. The transformed subspace by PCA is a new orthogonal uncorrelated coordinate system. PCA transforms the data into a new coordinate system where the greatest data variance is along the first coordinate, the second greatest variance is along the second coordinate and so on. The data variation in the new coordinate system is modeled by eigenvalues λ_i along their associated eigenvectors e_i . In simpler terms, the coordinates of the new systems are defined by the eigenvectors (also called principal components). Considering W as an orthogonal linear transformation model that maps the original D -dimensional space to a d -dimensional space ($d \leq D$), the new space Y is defined by:

$$Y = W^T X . \quad (3.1)$$

The transformation matrix W is constructed by eigenvectors e_i (each column of W cor-

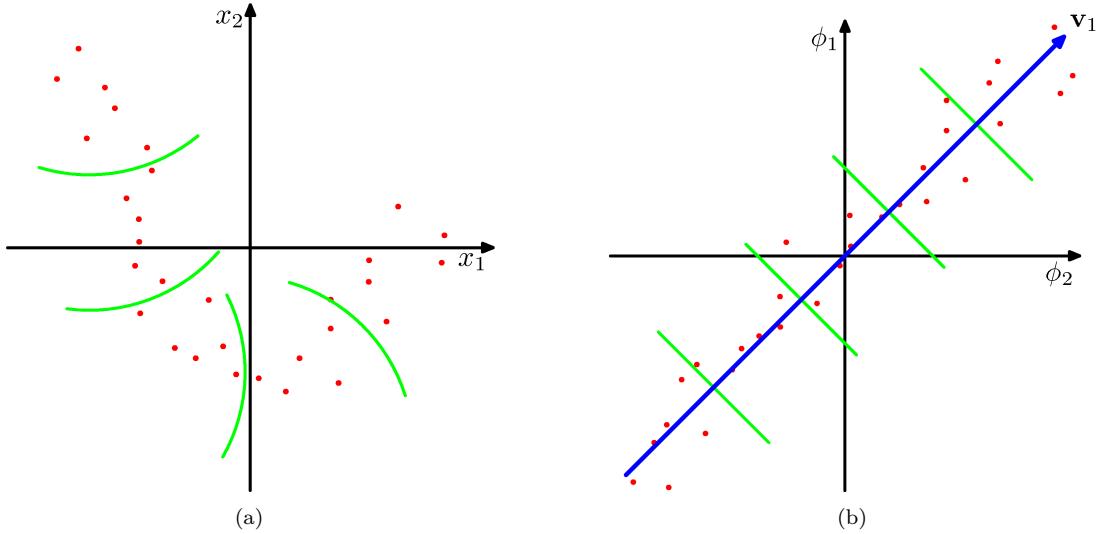


Figure 3.7: Illustration of kernel PCA. (a) Nonlinear principal component in the original data space, (b) Nonlinear projection of the original data space into features space ($\phi(x)$). In this space it is possible to apply linear PCA, v_1 is the first principal component in the feature space and the green lines are linear projections of the data to the principal component. This figure was taken from Bishop et al. [30].

responds to one eigenvector) of the covariance or scatter matrix of the data ($S = XX^T$).

$$\lambda e = Se . \quad (3.2)$$

Using PCA, the dimensionality of the data can be reduced by considering only the first principal components. In doing so, the dimensionality of the data is reduced while maximal variance of the original data is preserved. PCA is scale sensitive, thus data normalization is crucial. The data should also be mean-subtracted to assure that the eigenvector with the highest eigenvalues is along the maximum variance, and not the mean of the data.

Kernel PCA is a nonlinear, unsupervised transformation that assumes that the features are mapped with a nonlinear transformation $\phi(x)$ to a new M -dimensions feature space (M could be much higher than the original data space D). By using this transformation, each point x_n is projected onto a point $\phi(x_n)$. The idea of the kernel PCA is to perform a standard PCA in the feature space rather than the data space, which will create nonlinear principal components in the original data space (see Fig. 3.7). Assuming that the projected data in the feature space is zero-mean, the covariance matrix in the feature space is given

by:

$$C_{M \times M} = \frac{1}{N} \sum_{n=1}^N \phi(x_n) \phi(x_n)^T . \quad (3.3)$$

Similar to the standard PCA, our goal is to find the eigenvectors of this matrix where the eigendecomposition is given by:

$$\lambda_i v_i = Cv_i. \quad (3.4)$$

Substituting Eq. 3.3 in Eq. 3.4, the eigenvectors are given by:

$$v_i = \sum_{n=1}^N a_{in} \phi(x_n). \quad (3.5)$$

Using the previous formulation in the eigendecomposition, and based on the kernel substitution, the mapping in the feature space can be represented in terms of a kernel function $k(x_n, x_m)$:

$$\frac{1}{N} \sum_{n=1}^N k(x_l, x_n) \sum_{m=1}^N a_{im} k(x_n, x_m) = \lambda_i \sum_{n=1}^N a_{in} k(x_l, x_n) , \quad (3.6a)$$

$$Ka_i = \lambda_i Na_i, \quad (3.6b)$$

Here a_i is an N -dimensional vector and K is an $N \times N$ matrix solved by eigenvalue decomposition. Based on the aforementioned equations, we can now formulate the principal component projections in terms of the kernel function as illustrated below:

$$y_i(x) = \phi(x)^T v_i = \sum_{n=1}^N a_{in} \phi(x)^T \phi(x_n) = \sum_{n=1}^N a_{in} k(x, x_n) . \quad (3.7)$$

Bag of Words (BoW) is a modeling or mapping of the extracted features to a new space based on a set of main clusters in the feature space. This method looks for similar or strong patterns among the extracted features and presents all the features based on the number of occurrence of these patterns. BoW clusters a set of low-level features using a k -means algorithm to create a “codebook” of “visual words”. Each visual word is defined by a centroid of the corresponding cluster. After creating the codebook, each sample is represented as a histogram of size k obtained by calculating the occurrence frequency of the k words among the features extracted from the sample.

K-means is an iterative algorithm that finds k centroids by alternating assignment and update steps. The assignment steps are based on L_2 norm (Euclidean) distance. Different

initialization methods can be used in order to assign the initial k clusters [39]. In this research, these clusters are selected based on the greedy k -means++ method [14]. The choice of the visual words (number of k clusters) depends on the application, and different choices can be made. A suitable number is usually found via an exhaustive search or primary tests.

Sparse Coded Features (SCF), or sparse signal representation, has become very popular over the past few decades and has led to state-of-the-art results in various applications such as face recognition [204], image denoising, image inpainting [71], and image classification [177]. The main goal of sparse modeling is to efficiently represent the samples/images as a linear combination of a few typical patterns, called atoms, selected from the dictionary. Sparse coding consists of three main steps: (i) dictionary learning, (ii) low-level feature projection, and (iii) feature pooling [165]. Considering our dictionary $D \in R^{n \times K}$ with K atoms, where each column of D represents one atom, the sparse coding problem of a signal $y \in R^n$ is defined as finding the sparsest vector x so that $y \approx Dx$. This is an optimization problem that can be formulated as:

$$\min_x \|y - Dx\|_2 \quad \text{s.t. } \|x\|_0 \leq \lambda, \quad (3.8)$$

where λ is the sparsity level and l^0 -norm accounts for the minimum number of non-zero elements in the sparse vector x . This optimization problem is NP hard [70], subsequently approximation solutions are proposed either by using greedy algorithms such as matching pursuit [126] and orthogonal matching pursuit [58], or by replacing the l^0 -norm with l^1 -norm such as in the basis pursuit algorithm [48]. The dictionary is learned using K-SVD, a generalized version of k -means clustering that uses Singular Value Decomposition (SVD) [10]. The dictionary is built such that:

$$\min_{Dx} \|y - Dx\|_2 \quad \text{s.t. } \forall i \|x_i\|_1 \leq \lambda, \quad (3.9)$$

where y is a low-level descriptor, x is the sparse coded descriptor (i.e., high-level descriptor) with a sparsity level λ , and D is the dictionary with K atoms. The K-SVD algorithm solves the optimization problem iteratively by alternating between x and D . With D , the sparse code matrix x is computed by any of the pursuit algorithms, and with x , D is updated one atom at a time using SVD.

Once the dictionary is learned, each $y_i \in R^n$ signal can be projected using D to form a set of sparse codes $x_i \in R^K$. In the case of image samples, the sparse representation can be generated for patches in the image. In this case, the final mapping is based on

a combination of sparse codes, for instance by taking the maximum code from all the patches:

$$f_i = \max_j(|X_l(i, j)|) \quad \forall i = 1, 2, \dots, K, \quad (3.10)$$

where $X_l \in R^{K \times P}$ is the sparse code matrix [177].

3.5 Balancing Strategies

While performing classification for real world applications, we usually face a problem in which the number of samples of one class is far less than the samples of another class. This problem is frequently referred to as the “class imbalance” problem [149] and has been encountered in many diverse areas such as telecommunication management [73], bioinformatics [152], fraud detection [147], and medical diagnosis [38]. Imbalanced data substantially compromises the learning process since most of the standard machine learning algorithms expect a balanced class distribution or an equal misclassification cost [101]. Medical data are prone to such drawbacks due to the fact that the portion of diseased samples or patients is far lower than healthy cases. Furthermore, the detection and classification of minority malignant cases is essential, so the sensitivity (SE) of the algorithms developed needs to be maximized. Consequently, the problem of imbalanced data is usually addressed by employing different techniques that do not impair the topology of the data. This section discusses some of the most used balancing techniques.

Considering a binary classification problem, the class with the smallest number of samples is defined as the *minority* class and its counterpart is defined as the *majority* class. Data balancing corresponds to a sample number equalization in both the minority and majority classes. This task can be achieved either in the data or feature space. Balancing strategies in the data space include elimination of some of the majority samples or the generation of synthetic minority samples. An example of synthetic sample generation in handwritten character recognition is the alteration of the mean character shapes using a PCA. In the feature space, three strategies can be employed to overcome the problem of imbalanced dataset: (i) Under-Sampling (US), (ii) Over-Sampling (OS), and (iii) a combination of both. The following sections give an overview of the techniques used to tackle this issue.

3.5.1 Under-Sampling

The goal of US is to reduce the number of samples from the majority class so that it is equal to the number of samples from the minority class. The following methods are considered to perform the data balancing.

Random Under-Sampling (RUS) is performed by randomly removing a subset of samples

from the majority class (without replacement) so that the number of samples is then equal in both classes.

Tomek-Link (TL) can be used to under-sample the majority class of the original dataset [188].

Let (x_i, x_j) define a pair of Nearest Neighbor (NN) samples so their associated class labels are different $y_i \neq y_j$. The pair (x_i, x_j) is defined as a TL if, by relaxing the class label differentiation constraint, there is no other x_k sample defined as the NN of either x_i or x_j . US is performed by removing the samples belonging to the majority class and forming a TL. It must be noted that this US strategy does not enforce a strict balance between the majority and the minority classes.

Clustering (CUS) refers to the use of k -means to cluster the feature space so that k is set to be equal to the number of samples composing the minority class. Hence, the centroids of these clusters define the new samples from the majority class.

NearMiss offers three different methods to under-sample the majority class [128]. In NearMiss-1 (NM1), for each selected sample in the majority class, the average distance to the k NN samples in the minority class is minimum. NearMiss-2 (NM2) diverges from NM1 by considering the k farthest neighbour samples in the minority class. NearMiss-3 (NM3) generates a subset M of the majority class samples by finding the m NNs of each minority class sample. The elements in M whose average distance to the k NN samples in the minority class is maximum are retained.

Editted Nearest Neighbor (ENN) is another US approach which removes any sample whose class label differs from the labels of at least two of its three NNs [201]. This method removes samples from both the majority and minority classes.

Neighborhood Cleaning Rule (NCR) consists of applying two rules depending on the class of each sample [114]. This is a modification of the (ENN) method that performs a better data cleaning. Let us define x_i as a dataset sample with its associated class label y_i , and y_m as the class of the majority vote of the k NNs of x_i . If y_i corresponds to the majority class and $y_i \neq y_m$, x_i is rejected from the final subset. If y_i corresponds to the minority class and $y_i \neq y_m$, then the k NNs are rejected from the final subset.

3.5.2 Over-Sampling

Data balancing can also be performed by OS: new minority class samples are generated to equalize the number of samples in both classes. Two different methods are considered.

Random Over-Sampling (ROS) is performed by randomly replicating samples in the minority class so that the number of samples is equal in both the minority and majority classes.

Synthetic Minority Over-Sampling (SMOTE) is a method to generate synthetic samples in the feature space [47]. Let us define x_i as a sample belonging to the minority class, and x_{nn} as a randomly selected sample from the k NNs of x_i . Then SMOTE can generate a new $x_j = x_i + \sigma(x_{nn} - x_i)$ sample, where σ is a random number in the interval $[0, 1]$.

3.5.3 Combination of OS and US

OS methods can be combined with US to clean the newly generated over-sampled set of minority class samples. In that regard, two different combinations are tested.

SMOTE + TL are combined to clean the samples created using SMOTE [22, 149]. SMOTE over-sampling may lead to overfitting, which can be avoided by removing the TL from both the majority and minority classes.

SMOTE + ENN [23]. The Edited Nearest Neighbor removes all the misclassified samples with respect to its 3 NNs, without any concern for their class, and generally eliminates more samples in comparison with TL. It is expected to have a more in-depth data cleaning when using ENN [149].

Here, to deal with the imbalance problem, the fastest, simplest and least parametric heuristic OS and US methods are discussed. However, there are other approaches, such as ensemble and cost-sensitive learning, that are beyond our interest in this thesis. A curious reader can refer to the following references [101, 46] for more information on this topic.

3.6 Classification

The classification problem was discussed previously in Sect. 3.1. Numerous approaches have been introduced by the research community to solve the classification problem. Here we discuss some of these approaches divided into two categories: “single learner” and “ensemble”.

3.6.1 Single Learner

This group contains a large number of classifiers, or base learners, that have a unique approach to learn from the training set. In the following, we discuss some of these classifiers: support Vector Machine (SVM), K-NN, LDA and Naive Bayes (NB).

Naive Bayes (NB) is one of the simplest probabilistic classifiers, based on Bayes' theorem. This classifier has a “naive” or independence assumption that states that features are conditionally independent given the class variable [139]. Given a sample to be classified represented by n features, $\mathbf{x} = (x_1, \dots, x_n)$, the conditional probability of this sample belonging to class C_k based on Bayes' theorem is given by:

$$p(C_k|\mathbf{x}) = \frac{p(C_k)p(\mathbf{x}|C_k)}{p(\mathbf{x})}. \quad (3.11)$$

This equation simply states that the posterior is equal to the prior times the likelihood while normalized. Considering that the numerator of Eq. 3.11 can be written as a joint probability $p(C_k, x_1, \dots, x_n)$ and taking into account the conditional independence assumption, the NB model is represented by:

$$p(C_k|x_1, \dots, x_n) = \frac{1}{Z} p(C_k) \prod_{i=1}^n p(x_i|C_k), \quad (3.12)$$

where $Z = p(x)$ is the normalization or scaling factor depending on x_1, \dots, x_n . For classification purposes, the NB is combined with decision rules, such as Maximum a Posteriori (MAP) or Maximum Likelihood (ML):

$$\hat{y} = \arg \max_{k \in \{1, \dots, K\}} p(C_k) \prod_{i=1}^n p(x_i|C_k). \quad (3.13)$$

K-Nearest Neighbor (k-NN) is another simple and non-parametric classification method.

This classifier assigns each sample's label based on the majority vote of its K nearest neighbors. If K is equal to 1, the output is assigned to the label of the nearest neighbor.

Linear Discriminant Analysis (LDA) or the Fisher linear discriminant analysis, was introduced in Sect. 3.4 as a dimension reduction approach that takes the class label into account. Since this method contains the class information, it can be used for classification as well. However, it has a drawback, that regardless of the number of feature dimensions, it always maps the data to $L \leq C - 1$. This indicates that for a two class classification, in an attempt to maximize the margin between the two class, it maps the data to one vector [139].

support Vector Machine (SVM) is created based on the combination of the kernel trick and modified loss function [139]. SVM [193] is a well known machine learning approach that aims to separate two classes by finding the best hyperplane maximizing the margin

between the two classes:

$$\min_{\mathbf{w}, \omega_0, \xi} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^N \xi_i \quad \text{s.t.} \quad \xi_i \geq 0, \quad y_i(\mathbf{x}_i^T \mathbf{w} + \omega_0) \geq 1 - \xi_i, i = 1 : N. \quad (3.14)$$

Maximizing the margin is equivalent to minimizing the norm of the normal vector of the hyperplane:

$$\min_{\mathbf{w}, \omega_0} \frac{1}{2} \|\mathbf{w}\|^2 \quad \text{s.t.} \quad y_i(\mathbf{w}^T \mathbf{x}_i + \omega_0) \geq 1, i = 1 : N. \quad (3.15)$$

This constraint intends to force all the points to be on the correct side of the decision boundary (hyperplane) with a minimum distance of 1. This assumption is only valid if the data is linearly separable. Thus, for general cases, a slack variable $\xi_i \geq 0$ is introduced. If the points are on/or inside the correct margin boundary, then $\xi_i = 0$; if the points are inside the margin and still on the correct side of the decision boundary, then $0 < \xi_i \leq 1$; otherwise, if the points lay on the wrong side of the decision boundary, $\xi_i > 1$. This assumption introduces *soft margin constraints*. Considering the rule of ξ_i , the $\sum_i \xi_i$ term in Eq. 3.14 describes the upper bound on the number of misclassified points and C is the regularization parameter that controls the tolerance of the classifiers on the number of errors [139].

3.6.2 Ensemble

Ensemble learners are defined by a combination of several base learners $f_m(\cdot)$:

$$f(y|y, \pi) = \sum_{m \in M} \omega_m f_m(y|x), \quad (3.16)$$

where ω_m is the tunable parameter [139]. Some ensembles are created based on the combination of base learners of the same type, such as Random Forests (RF) and AdaBoost (AdB), whereas others combine different base learners in a specific way, e.g. stacking and Error Correcting Output Coding (ECOC). Some well-known ensembles are described in the following.

AdaBoost (AdB) (short for Adaptive Boosting) is an ensemble learning algorithm that provides a high performance classifier via a linear combination of weak learners [80]. Considering a training set (x_i, y_i) , where x is an M dimensional feature vector and y_i is its associated class label, AdB iteratively adds weighted weak learners ($\alpha_t h_t$) to assemble the final strong classifier. Here, α_t is the weight and h_t is the weak learner:

$$H(x) = \text{sign} \left(\sum_{t=1}^T \alpha_t h_t(x) \right). \quad (3.17)$$

A weak learner is usually selected as a classifier that performs slightly better than randoms ($> 50\%$). Weak learners are selected with their associated weight so that the sum of the training error E_t of the resulting t stage classifier is minimized.

$$E_t = \sum_{i=1}^N e^{-y_i H_t(x_i)}, \quad (3.18a)$$

$$H_t(x) = \alpha_1 h_1(x) + \dots + \alpha_t h_t(x). \quad (3.18b)$$

This ensemble is constructed with the aim of classifying harder patterns by re-weighting the distribution after each iteration [163]. Starting with equally distributed weights for our distribution $D_1(i) = 1/N, i = 1 : N$. The distribution is updated after each iteration in a manner that the weights of correctly classified samples decrease and the weights of misclassified samples increase.

$$D_{t+1}(i) = \frac{D_t(i) \cdot e^{-\alpha_t y_t h_t(x_i)}}{Z_t}. \quad (3.19)$$

Re-weighting the distribution automatically forces the next weak learner to focus on the difficult samples. As mentioned previously, each weak learner is assigned a weight as well and the best weak learners have higher weights in comparison to the weakest learners.

A variety of learners can be adapted in the algorithm as weak learners such as neural networks or decision trees. One common weak learner is decision stump, which is a one-level decision tree, equivalent to the threshold that best splits the data. Each stump learner is characterized by three parameters: (i) the m^{th} dimension of the features set where the classifier is applied, (ii) the decision level, i.e., the threshold splitting the data in the m^{th} given dimension and (iii) the decision sign (-1 or $+1$) determining the inequality direction for the thresholding. For a given batch of data with a set of features of size M , at each iteration of AdB, the decision stump that minimizes the error t in an m^{th} dimension of the training distribution is selected. The information provided by the final set of decision stumps selected by AdB can also be used for determining the significant features of the data-stream and, more importantly the best split in the data.

Random Forests (RF) is an ensemble of decision trees [32] that generalizes the classification process by applying two types of randomization: at the tree level, each tree is fed by a bootstrap made of S' samples build from the original data of size S so that $S = S'$, and at the node level, a subset of feature dimensions m is randomly selected from the original dimension M so that $m = \sqrt{M}$. The trees in RF are grown to their maximum length without any pruning. Figure 3.8a shows the training stage of an RF ensemble. In this

figure rows and columns represent the samples and the feature dimensions, respectively. As it mentioned each tree is trained on the bootstrap of the original samples and in each node of a tree, subset of feature dimensions are used for prediction. In the testing stage, each tree in the ensemble casts a unit vote, and the final prediction is based on the combination of all the votes. This process is shown in Fig 3.8b.

Gradient Boosting (GB) is a generalization form of AdB able to use real-value weak learners and minimize different loss functions [210]. GB builds the ensemble in a greedy manner. It iteratively selects the best pair of real-valued weak learners and adjusts their weights so they minimize a given differentiable loss function.

$$\mathcal{L} = \sum_{i=1}^N L(y_i, \varphi(x_i)), \quad (3.20a)$$

$$\varphi(x) = \sum_{j=1}^M \alpha_j h_j(x), \quad (3.20b)$$

here $h_j(\cdot)$ is the weak learner and α_j is the real value weight and the aim is to minimize \mathcal{L} . A common choice for the weak learner is a decision stump or regression tree while the loss function is generally an exponential or a logarithmic loss [25].

$$\text{Exponential loss} \quad L = e^{-y_i \varphi(x_i)}, \quad (3.21a)$$

$$\text{Log loss} \quad L = \log(1 + e^{-2y_i \varphi(x_i)}). \quad (3.21b)$$

This minimization can be carried out via a gradient descent or a quadratic approximation.

Stacking [202] is another way to create an ensemble from multiple base learners. The idea is to have a meta learner which considers the predictions of the previous base learners as input for the training stage and combines them into a final decision. Using this technique, the training data is partitioned into two sets, which we call training and validation sets. Each base learner is trained on the training set and its prediction on the validation set is fed to the meta learner. Similarly, the test sample is first classified by the base learners and their prediction is passed through the meta learner in order to arrive at the final decision. Stacking seems to avoid overfitting and improves performance when it is used with cross-validation [69]. This technique was used for problems such as segmentation and labeling [139]. Figure 3.9 shows the principal of the staking approach.

Error Correcting Output Coding (ECOC) is an interesting way of ensemble learning, especially suitable for multiclass classification [65]. This technique allows the application of

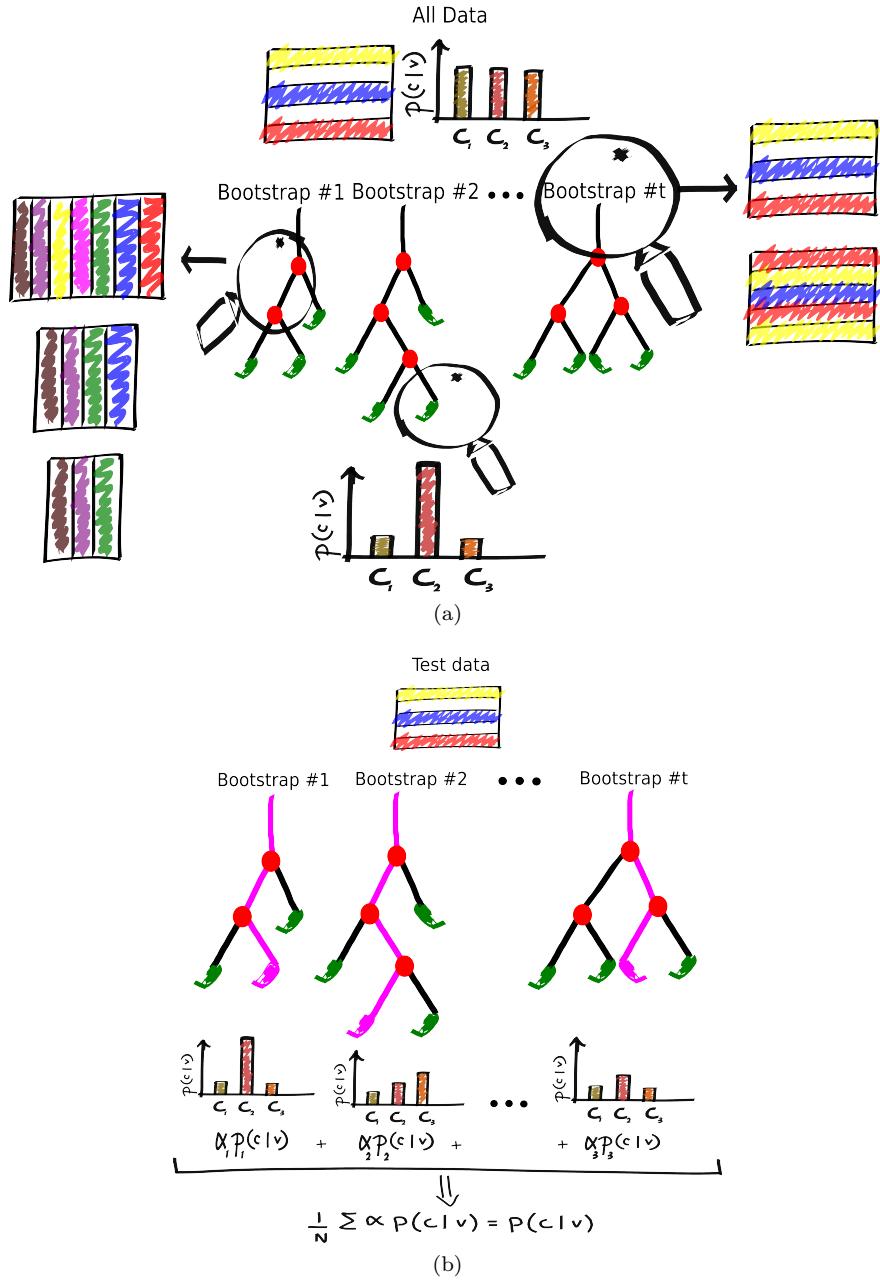


Figure 3.8: RF ensemble: (a) training stage of RF ensemble. This ensemble randomizes the training by first creating different bootstraps for training each tree in the forests. Second, at each node of the tree, it randomly considers a smaller set of feature dimensions. Each node of a tree predicts one class. (b) testing stage of RF ensemble. A test sample is passed through all the trees and its prediction is assigned by majority voting of predicted labels of the trees.

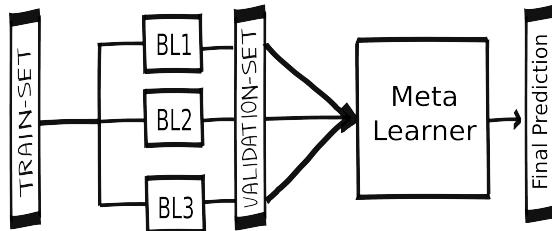


Figure 3.9: Stacking ensemble approach for three base learners (BL1, BL2, and BL3). Different base learners can be combined using this method.

binary classifiers for multiclass classification by randomly assigning different classes to a “super-class” (0 or 1). Let us consider a 5 class classification $C \in \{A, B, D, E, F\}$, each time a set of classes is randomly assigned to one of the super-classes. For instance, let's consider a binary classifier, where A and F are categorized in super-class 1 and B, D , and E belong to super-class 0. Repeating the binary classification several times, each class will be represented by a binary code. This binary code states that in each binary classification, our considered class was categorized in one or the other super-class. Going back to our example and assuming a 5 binary classification, having a binary code 11010 for class F shows that for binary classification problems of 1, 2, and 4, F was assigned to super-class 1 and the rest to 0. Subsequently, an ensemble is created as a combination of binary classifiers. Similarly, a test sample is classified with all the binary classifiers and receives a predicted binary code. The predicted class is assigned based on the closest class vector to the predicted vector. Hamming distance is used to find a closest class vector. Although the classes can be assigned in a pre-designed manner to maximize their distance to each other, James et. al [107] showed that the random code works just as well as the optimal code [139].

Weighed combination is a very simple way of ensemble learning, and, as its name suggests, it creates a weighted combination of base learners. Each base learner will have a strength proportional to its assigned weight. The assigned weight can be fixed or dynamically determined. There are different approaches to determine the base learners' weights: majority voting (plurality vote), Bayesian combination, entropy weighting and performance weighting among others [163]. Each base learner's weight is proportional to the aforementioned measures on the validation set. Figure 3.10 shows an ensemble based on performance weighting of three base learners (BL).

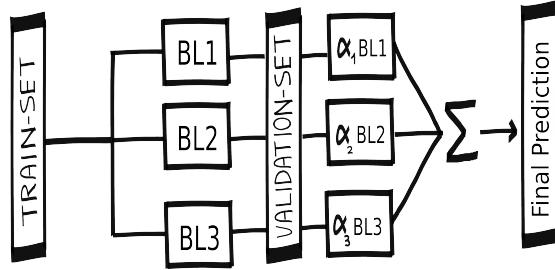


Figure 3.10: An ensemble of three base learners (BL1, BL2, and BL3, respectively) weighted by their performance on the validation.

3.7 Validation and Evaluation

This section first discusses different methods of validation and then presents different measurements of validation.

3.7.1 Validation

Validation, or cross-validation, is an important part of the classification framework that ensures its generalization. Validation also avoids having a bias classifier or overfitting. k fold cross-validation (k -CV) and leave-one-out-cross-validation (LOOCV) are two different techniques for performing validation. In this section, we first discuss different methods for validation of a data classification framework and then present various measurements of its evaluation.

k fold cross-validation randomly divides the dataset into k partitions of equal size. Among the k partitions, one is used as the test set, and the classifier is trained on the remaining $k - 1$ subsets. Using k -CV, the classification is performed k times to ensure that each of the subsets is used as the test set once.

Leave-one-out-cross-validation can be seen as a form of k -CV when $k = n$, n is the total number of samples. In this validation, the classification is repeated n rounds. At each round, one sample is left out for testing and the classifier is trained on the remaining samples. This method is suitable when dealing with a small or unbalanced dataset.

3.7.2 Evaluation

Evaluation is another important part of the framework that measures the quality and performance of the framework. While solving classification problems, a simple way to quantify the performance of the classifier is to look at the confusion matrix. The confusion matrix is an error table that shows the performance of the algorithm. Each row of the matrix represents instances

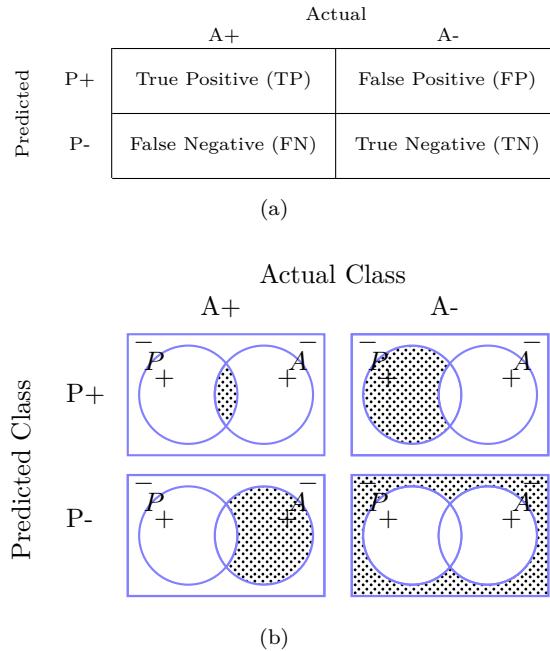


Figure 3.11: Tabular (a) and graphical (b) representations of a confusion matrix with true and false positive samples (TP, FP) in the first row and false and true negative samples (FN, TN) in the second row (from left to right).

in the predicted class and each column shows instances in their actual class (or vice-versa). Figure 3.11 shows the table and graphical representation of a confusion matrix.

True positive (TP) and true negative (TN) instances are simply the samples that are correctly predicted to belong to the positive or negative class, respectively. False positive (FP) samples are predicted as the positive class while in reality they belong to the negative class; and false negative (FN) instances are predicted as the negative class while belonging to the positive class. Various statistic measures are derived from the confusion matrix, among which the most popular ones are accuracy (ACC), sensitivity (SE), specificity (SP), and precision.

Accuracy (see Eq. 3.22) measures the overall performance of the algorithm by considering the number of instances classified correctly without considering their classes. Although this measure is important, it can be misleading in real-world classification problems because an overfitted or biased classifier can have high accuracy while not performing as desired.

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} . \quad (3.22)$$

The sensitivity, recall or true positive rate is another statistic that measures an performance of

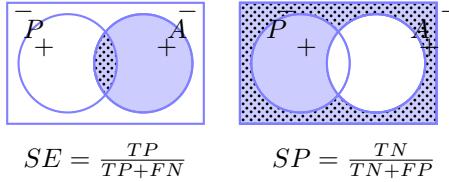


Figure 3.12: Sensitivity and specificity evaluation, corresponding to the ratio of the doted area over the blue area.

the algorithm with respect to the positive class only.

$$SE = \frac{TP}{TP + FN} . \quad (3.23)$$

Opposite to SE, specificity or true negative rate measures the algorithm's performance with respect to the negative class:

$$SP = \frac{TN}{TN + FP} . \quad (3.24)$$

Figure 3.12 shows a graphical illustration of SE and SP in the confusion matrix.

Precision or positive predictive value (PPV), as its name suggests, is a measure of the algorithms precision. Considering an algorithm that is more biased towards the positive class, this algorithm has high SE, however, it is less precise, since it considers most of the instances as more or less positive.

$$precision = \frac{TP}{TP + FP} . \quad (3.25)$$

All the aforementioned measures can be used to quantify the performance of an algorithm. They can also be combined to generate a graphical plot such as Receiver Operating Characteristics (ROC) or precision-recall curve. The ROC [218] is a graphical curve of SE against a false positive rate (FPR) or (1 - SP). The curve is constructed by varying the discriminant threshold of the classifier. Adjusting the threshold will lead to classifying more TP samples, usually at the cost of increasing FPs. Due to this property, the ROC is a valuable tool, specially in the medical field when having high SE is important. This tool allows us to analyze the cost of having the highest SE against the number of FP. This curve is shown in Fig. 3.13. Here each point represents an SE/SP pair corresponding to the discriminant threshold. The ROC curve statistics are often represented by the area under the curve (AUC), which is a single number corresponding to the area under the ROC curve. The AUC varies between 0.5 to 1 for algorithms with an average to perfect performance.

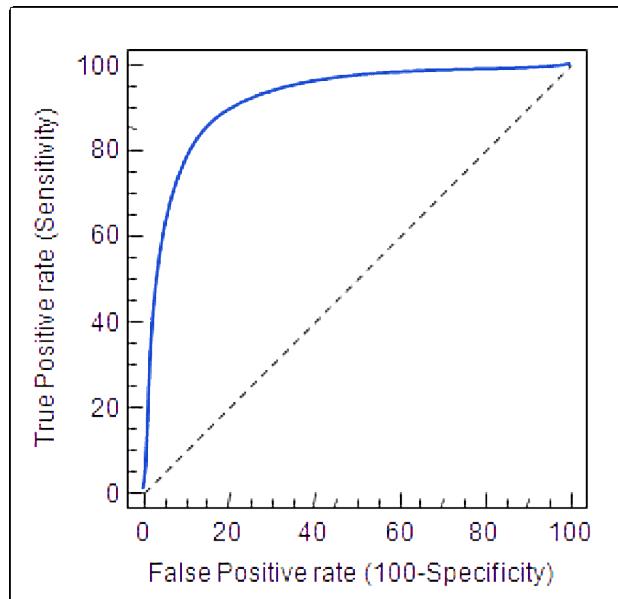


Figure 3.13: ROC curve.

3.8 Conclusion

This Chapter presented a broad-view of different aspects of machine learning and classification. It also discusses, the most used and well-known techniques of each steps and the flow of the classification framework. It is important to notice that the choice of different techniques in each step is application dependent and the best configuration is mostly found empirically. Next chapters presents our proposed frameworks, where different techniques are employed in each step.

Chapter 4

Automated Melanoma Classification with Dermoscopic Images

4.1 Introduction

Dermoscopy is a well-known conventional skin screening technique. It is used by most dermatologists for screening and analyzing skin lesions. The main principle of the dermoscope, and more specifically, the polarized dermoscope, was discussed previously in Chapt. 1, Sect. 1.4.2. Due to the importance of an early diagnosis of melanoma, numerous research studies have been dedicated to vision-based automated systems that assist dermatologists in dermoscopy image analysis. These studies have focused either on specific areas such as lesion delineation, feature extraction and lesion classification, or complete Computer-Aided Diagnosis (CAD) systems, which include all the aforementioned areas.

In this chapter, we describe our framework for the classification of dermoscopic images in the following order. First, a summary of work related to the classification of dermoscopy images is presented in Sect. 4.2. Then in Sect. 4.3 and 4.4, the material, datasets and the progress of our research are discussed. The experiments developed throughout the thesis, their results and conclusions are explained in Sect. 4.5. Finally, we conclude our findings regarding the developed classification framework in Sect. 4.6.

4.2 Related Work

This section presents a summary of the algorithms for dermoscopic lesion classification proposed over the past decade. Graphically, this summary is described in Table 4.1 and Fig. 4.1.

In 2001, Merler et al. [136] proposed a cost-sensitive boosting approach for the classification of melanoma (M) against benign (B) lesions, where they report sensitivity (SE) and specificity (SP) of 97% and 54%, respectively. They used a dataset called “MEDS”, which contained 152 lesions, including 42 melanoma.

In another study, Kreutz et al. [113] presented a classification framework of M vs. B lesions, in a dataset of 423 dermoscopic lesions. In this study, the authors proposed a segmentation approach based on color clustering and a region growing technique. Using Multi Layer Perceptron (MLP) and features such as color statistics, shape and Gabor features, they reported SE and SP of 98.7% and 76.5%, respectively.

In the same year, Dreiseitl et al. [67] reported a comparison of different classifiers (kernel support Vector Machine (SVM), k-Nearest Neighbor (NN), Artificial Neural Network (ANN), and Decision Trees (DT)) with a dataset of 1619 lesions, among which 105 lesions were melanoma. This comparison of M vs. B+D (Dysplastic) lesion classification methods reported that the best SE (92.05%) and SP (94.97%) was achieved with the kernel SVM.

Ganster et al. [84] proposed an automated melanoma recognition framework using a very large dataset of 5363 lesions, including 96 melanoma. The authors proposed segmenting lesions based on a fusion method of several thresholding stages. Using a k-NN classifier and a multiclass classification, the results for SE and SP were 87% and 92%, respectively.

Rubegni et al. [164] also proposed an automated diagnosis framework for pigmented lesions, where they used ANN with shape and color statistics to differentiate M vs. B lesions. Classifying a dataset of 588 lesions, including 217 melanomas, the authors reported SE and SP of 99% and 92.6%, respectively.

Later, Sboner et al. [171] proposed an ensemble approach for classification of M vs. B lesions. Their algorithm on a dataset of 152 lesions, containing 42 melanomas, achieved SE and SP of 81% and 74%, respectively.

Burroni et al. [33] proposed an algorithm for differentiation of M vs. D lesions. In their method, the authors segmented 174 lesions using a Laplacian filter and a zero-crossing algorithm. Using the extracted shape, color statistics, and Grey-Level Co-Occurrence Matrix (GLCM) features and the stepwise discriminant analysis as the classifier [33], they obtained SE and SP of 71% and 75%, respectively.

Similar features were used later in [38]. In their framework, the authors proposed segmenting lesions using a region growing and merging algorithm with a color similarity constraint. Their experiments for classification of M vs. B+D lesions using kernel SVM, with a dataset of 564

lesions containing 88 melanomas, achieved SE and SP of 92.34% and 93.33%, respectively.

Iyatomi et al. [105] proposed a CAD system for recognition of M vs. B+D lesions.

The authors segmented lesions using a region growing algorithm and classified some shape, color and texture (GLCM) features, which resemble the “ABCD” characteristics, with an ANN classifier. They reported that with a large dataset of 1258 lesions including 198 melanomas their algorithm achieved SE and SP of 85.9% and 86%, respectively.

Situ et al. [181] proposed representing image features using the Bag of Words (BoW) approach. Classifying wavelet and Gabor texture features using the BoW representation and an SVM classifier, the authors reported SE and SP of 88.8% and 69.26%, respectively. This study was dedicated to the differentiation of M vs. B+D lesions with a dataset of 100 lesions, containing 30 melanomas. In a later study [180], the authors proposed a different segmentation method based on the graph cut algorithm, and, using the same classification framework, they extracted color, Scale-Invariant Feature Transform (SIFT), and wavelet features instead. As a result, with the same dataset, they achieved SE of 83% and SP of 80.93%.

Another computerized framework was proposed by Capdehourat et al. [35] for the classification of M vs. B+D lesions with a dataset of 433 lesions, among which 80 lesions were melanomas. The authors used the Dullrazor® [115] algorithm for hair removal and proposed segmenting lesions using color thresholding based on Otsu’s approach. Using color statistics and texture features from GLCM and Gabor filters, and the AdaBoost (AdB) classifier, they achieved SE and SP of 95% and 91.25%, respectively. They also proposed using the Synthetic Minority Over-Sampling (SMOTE) balancing algorithm to deal with imbalanced data distributions. Using the same pre-processing and classification framework on a larger dataset (544 lesions containing 111 melanomas), they later achieved SE and SP of 90% and 77%, respectively [34].

A combination of SVM and BoW representation was also used in [216] to differentiate M vs. B lesions in a dataset of 164 lesions (including 80 melanomas). Using the BoW representation of Local Binary Pattern (LBP) features, the authors eliminated the segmentation stage and achieved SE and SP of 73% and 73%, respectively. Later, using a larger dataset (206 lesions containing 37 melanomas), the authors proposed a framework for classification of M vs. B+D based on shape, color and texture features mimicking the “ABCD” characteristics [215]. In this work, using feature selection techniques and the Quadratic Discriminant Analysis (QDA) classifier, their best results reached SE of 86% and SP of 52%.

The support Vector Machine classifier was also used in [92] for automatic recognition of M vs. D lesions. In this work, the authors proposed extracting color statistics and shape features from the entire image without considering the segmentation step. They later used a Principal Component Analysis (PCA) to reduce the feature dimensions and achieved SE and SP of 89% and 64%.

In other work, Garnavi et al. [86] proposed a CAD system for classification of M vs. B

lesions. In this study, the author compared manual and adaptive thresholding approaches for the segmentation of lesions. As for feature extraction, they combined wavelet package features with shape statistics. Their paper also provided a comparison between SVM, Random Forests (RF), Logistic Model Trees, and hidden Naive Bayes (NB) classifiers, where the best accuracy (ACC) of 88.30% was achieved with the RF classifier.

Abbas et al. [5] proposed another CAD system for the classification of M vs. B+D lesions. In this framework, the authors proposed their own hair removal [1] and lesion segmentation [4] algorithms. Using the SVM classifier and LBP, wavelet, and color statistics as features, with a relatively small dataset of 120 lesions (containing 20 melanomas) they reported SE and SP of 91.46% and 94.14%, respectively. Using the same framework, they later reported their results with a different dataset of 120 lesions, which contained 60 melanomas [6]. With this dataset, the SE and SP were at 88.2% and 91.3%, respectively.

Barata et al. [17, 20] and Ruela et al. [167, 166] recently presented their research on different subsets of the PH² dataset [135]. In this work, the authors compared the role of shape and color descriptors [167, 166] for the detection of M vs. B+D lesions using the AdB classifier. In [17, 19], they proposed using the BoW representation of color and gradient features. Their results comparing the AdB, kernel SVM, and k-NN classifiers indicated that the combination of the BoW representation and k-NN classifier achieved the best performance with SE and SP of 100% and 75%, respectively.

Recent studies on automated melanoma recognition were presented in [172, 176]. Schaefer et al. [172] proposed a M vs. B+D classification framework, using a dataset of 564 lesions, among which 88 lesions were melanoma. The authors chose the SMOTE sampling to avoid the imbalance problem and used ensemble approaches to perform the classification. Using shape, color and GLCM features, they achieved SE and SP of 93.76% and 93.84%, respectively. Shimizu et al. [176] tackled a multiclass classification problem between melanoma, basal cell carcinoma, seborrheic keratosis, and dysplastic nevus. They used an ensemble approach for classifying each group with a dataset of 964 lesions with 109 melanomas. In their study, using color and GLCM features SE of 90.48% for melanoma lesions was achieved.

All the studies described above are summarized in Fig. 4.1 and Table 4.1. Figure. 4.1 categorizes the methods in terms of the dataset size, classification scope (M vs. B, M vs. B + D, and M vs. D) and results achieved (SE and SP). In this figure, a categorization of different datasets is presented together with the ratio of melanoma samples over the total size of the dataset. The datasets with less than 100 lesions are categorized as group 1, group 2 contains datasets with more than 100 and less than 200 lesions; datasets with more than 200 and less than 500 lesions are in group 3, while groups 4 and 5 contain datasets from 500 up to 1000 and over 1000 lesions, respectively. Table 4.1 categorizes the research studies in terms of their choice of features. In this table, the references that chose a BoW representation of the extracted

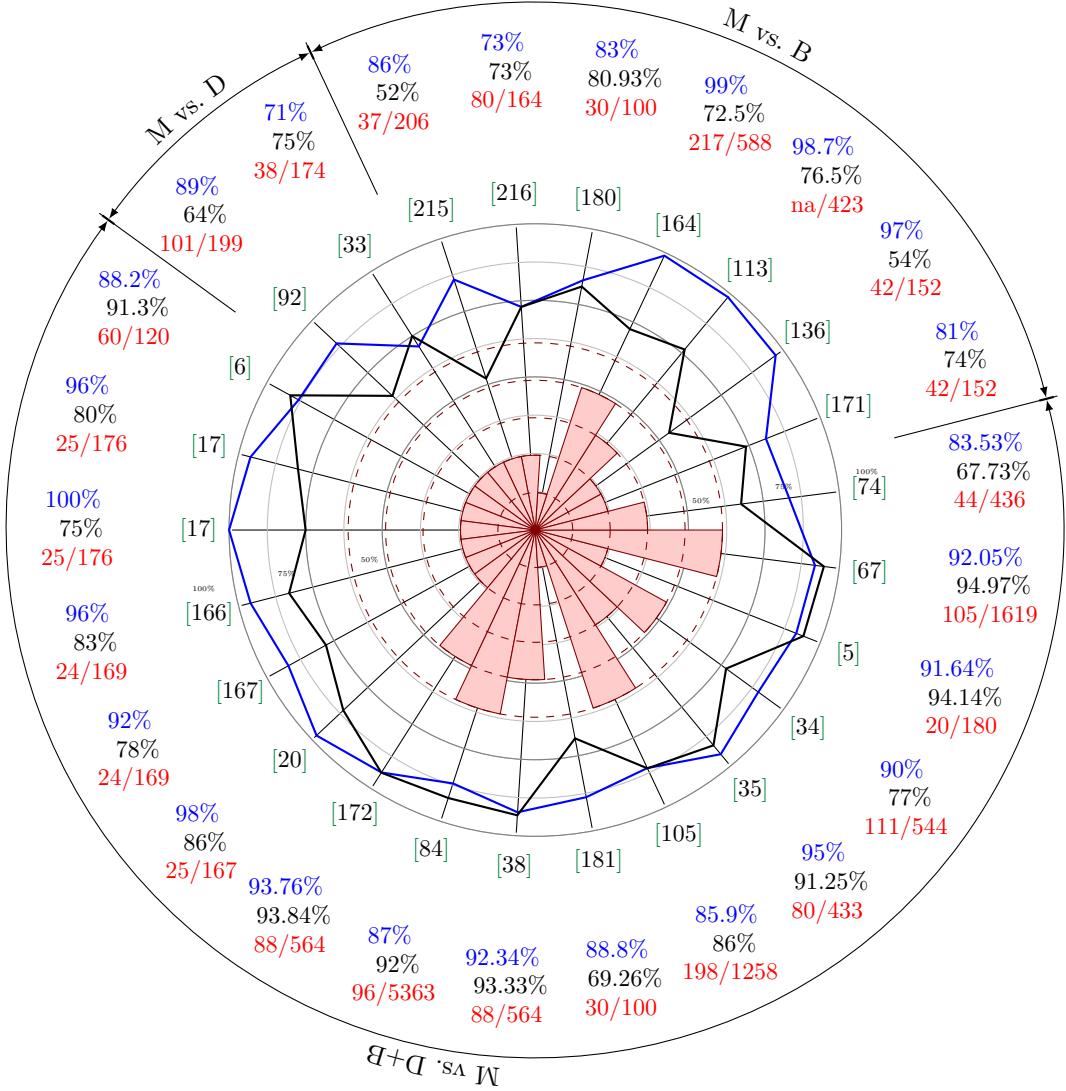


Figure 4.1: Summary of the classification performances of the methods reviewed from the dermoscopic imaging literature. The main results of the authors (references are presented in green) are illustrated in blue and black as sensitivity and specificity, respectively, while the size of the dataset is represented in red (cf., number of melanoma lesions over the total number of lesions). A comparison of the size of the datasets is also presented in the middle of the graph, which contains five groups. We categorized datasets with less than 100 lesions as group 1, datasets with more than 100 and less than 200 lesions as group 2, dataset with less than 500 lesions as group 3, any datasets with over 500 lesions and less than 1000 lesions as group 4 and finally datasets with over 1000 lesions as group 5. In this graph, M vs. D and M vs. B indicate that the dataset contains only malignant/dysplastic or melanoma/benign lesions, respectively, while M vs. $D + B$ indicates that melanoma lesions were classified against benign and dysplastic lesions.

Table 4.1: Summary of the features used in the literature. The references highlighted in **bold** use BoW to represent features in a high-dimensional space.

	Feature	Reference
<i>Shape</i>	Irregularity and compactness ratio	[113],[67],[84],[172],[167]
	Fractal geometry	[171],[113],[67],[38]
	Fourier features	[167]
	Area and perimeter	[171],[164],[172],[33],[38],[185],[84],[34],[74],[167]
<i>Color</i>	Alternative color space descriptors	[35],[84],[17]
	Color opponent angle	[17]
	Color variance and ratio	[171],[176],[113],[172],[67],[164],[33],[38],[105],[185],[84],[92],[34],[74],[180]
	Color quantization	[171],[45],[18],[166],[20]
<i>Texture</i>	Wavelet based descriptor	[86],[181],[180]
	Local binary pattern	[216]
	Gabor filter	[113],[67],[35]
	Gray-level co-occurrence matrices	[38],[105],[176],[185],[34],[74],[172],[181]
	Contrast and Entropy	[164]
	Histogram of gradients	[17]
	SIFT and Color SIFT	[180]
	Difference of Gaussian	[20]
	Harris / Hessian -Laplace	[20]

features are highlighted in **bold**.

Based on our study of the state-of-the-art lesion classification methods, we made several important findings listed in the following:

- i **Color, shape, and often GLCM features are used more extensively compared to others. Although these features are applied to represent characteristics similar to the “ABCD” rule, the lack of attention to other well-known pattern recognition features is evident.**
- ii **Different machine learning techniques are used to perform the classification. However, the application of ensemble classifiers was limited to only few studies.**
- iii **The majority of the datasets employed are imbalanced, in some cases to a great degree [84]. However, balancing techniques are rarely used to reduce the bias of the classifier.**
- iv **As it is highlighted in Fig. 4.1, few methods were proposed for the classification of M .vs D lesions. Classification of M against D lesions is more challenging due to the characteristic similarities between these types of lesions. This is a challenging task for both dermatologists and automated algorithms in comparison**

with differentiation between M .vs B lesions, which is more straight forward for dermatologists.

- v Finally, although some methods achieved very good results, it is impossible to offer a fair comparison of the above frameworks because their performance is reported using different datasets. Consequently, it is not possible to draw definitive conclusions about their performance in classification of melanoma lesions.

Keeping these conclusions in mind, we propose our CAD system for the classification of melanoma lesions, which is discussed in great detail in the rest of this chapter.

4.3 Materials

This section describes the dermoscopy datasets, used in our experiments. The primary tests were conducted based on the Vienna dataset [84], which we had access to thanks to Dr. H. Ganster. However, later, due to confidentiality problems, our access to this dataset was limited and we chose to work with PH² [135], the only publicly available dermoscopy dataset. The remainder of this section details these datasets while a summary is listed in Table 4.2.

Table 4.2: Summary of Vienna and PH² Datasets. M, D, and B stand for melanoma, dysplastic and benign samples, respectively.

Dataset	# samples	# M	# D	# B	Resolution px	Released	Public	Segmentation
Vienna	5380	101	1002	4277	632 × 387	2001	-	-
PH ²	200	40	80	80	768 × 560	2013	✓	✓

The **Vienna dermoscopy dataset** was acquired by the *Department of Dermatology at the Vienna General Hospital, Austria*. The images were captured using a hand-held CCD camera equipped with an epiluminescence microscope [84]. The acquired true-color 8-bit RGB images have a resolution of 632×3871 pixels (1 pixel ≈ 22 μm²) and were recorded without any color normalization. This is a large-scale dataset (group 5) with over 5000 lesions including 4277 benign, 1002 dysplastic and 101 malignant melanoma lesions. In this dataset, the dysplastic and melanoma lesions were surgically removed and their ground truth diagnosis were provided by histological analysis [84]. This dataset was also used in [189, 125]. Figure. 4.2 shows three sample images from this dataset, one for each class. The images in this dataset were acquired under different conditions and are illumination imbalanced.

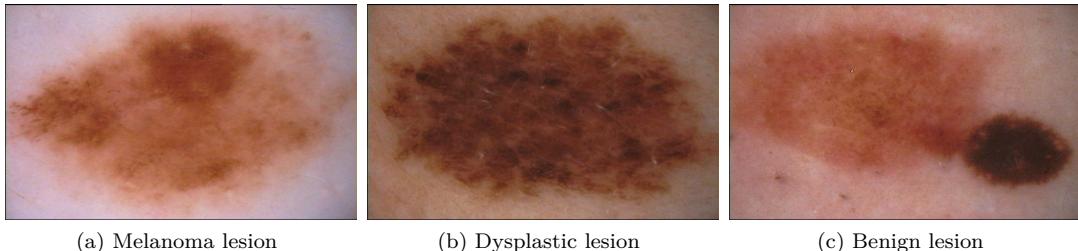


Figure 4.2: Sample images for the Vienna dataset depicting (a) melanoma, (b) dysplastic and (c) benign lesions.

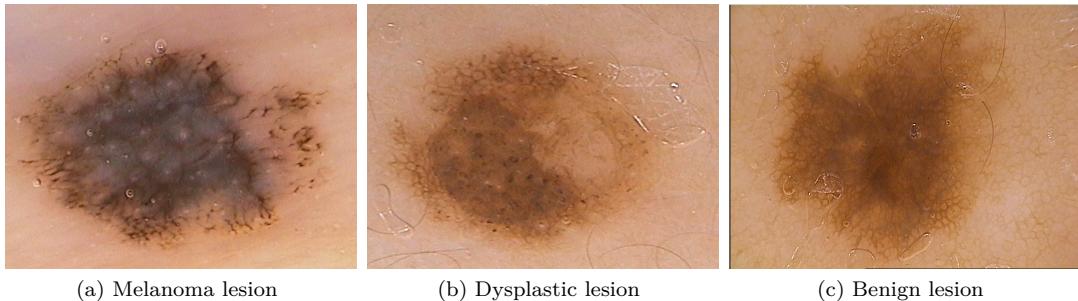


Figure 4.3: Image samples for the PH² dataset depicting (a) melanoma, (b) dysplastic and (c) benign lesions.

The **PH² dataset** was acquired at the *Dermatology Service of the Hospital Pedro Hispano, Matosinhos, Portugal* [135] with the Tuebinger Mole Analyzer system using a magnification of 20×. The 8-bits RGB color dermoscopic images were obtained under the same conditions with a resolution of 768×560 pixels. This dataset contains 200 dermoscopic images divided into 160 benign/dysplastic and 40 melanoma lesions. The lesions are segmented and their histological diagnosis are provided as the ground-truth. Figure 4.3 shows three image samples from this dataset, representing melanoma, dysplastic, and benign lesions.

Both figures 4.2 and 4.3 show the variation of colors between each lesion and similar color characteristics between melanoma, dysplastic and benign lesions.

4.4 Methodology

Figure 4.4 shows our proposed framework for automatic detection of malignant melanoma. As shown in the figure, the framework consists of 6 main stages according to our concept of the general classification framework (see Fig. 3.1). These stages are explained in the following.

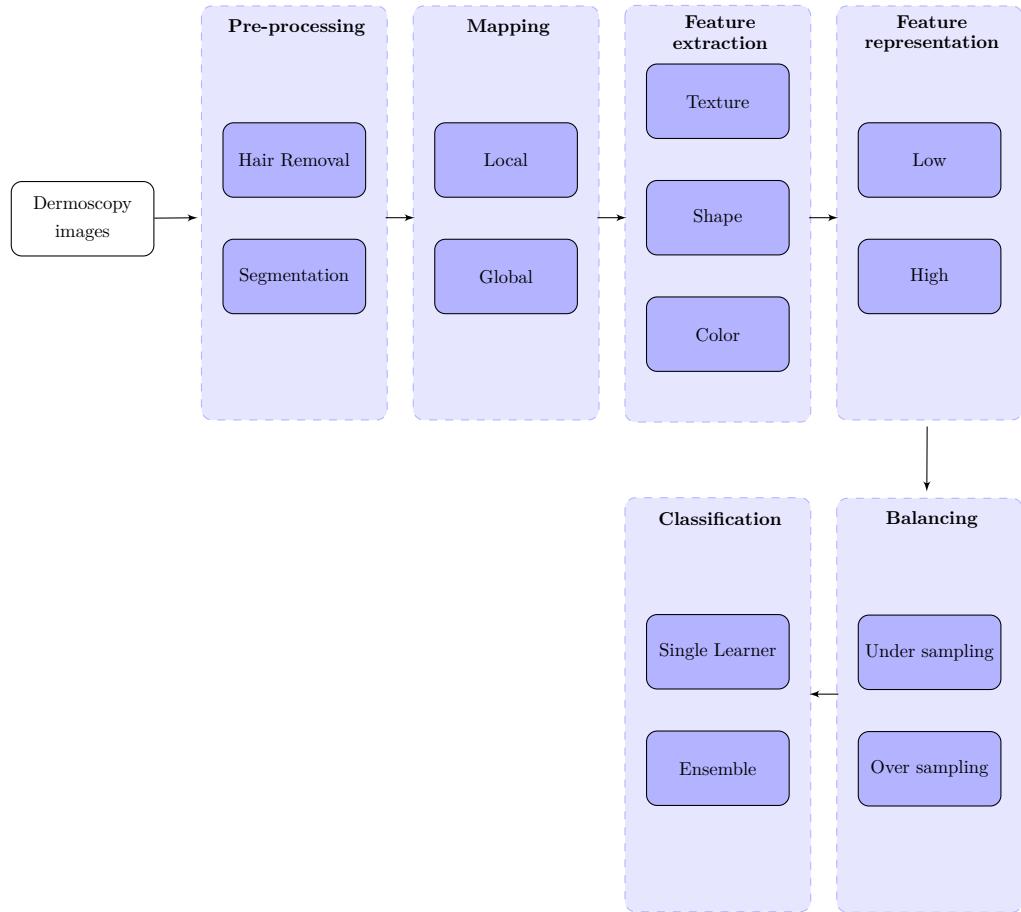


Figure 4.4: Outline of the proposed algorithm for automatic recognition of melanoma.

4.4.1 Pre-processing

In this stage, our proposed unique hair removal and segmentation algorithm is described.

Hair removal

The state of the art methods for hair detection in dermoscopy images were mentioned in Sect. 3.2. The previously mentioned methods [115, 111, 1, 7, 78, 175, 141, 62, 197, 50, 21, 205], were all originally proposed for hair detection and removal. In this research, considering the similarities between hair detection and vessel segmentation in fundus images, a fast, efficient and less parametric hair detection algorithm is proposed inspired by filters proposed for vessel segmentation. From different approaches and filters, the mathematical morphological method proposed by Zana et al. [208, 91] was chosen. This algorithm is based on morphological operations, such as:

Erosion: $\epsilon_B(I)$;

Dilation: $\delta_B(I)$;

Opening: $\gamma_B(I) = \delta_B(\epsilon_B(I))$;

Closing: $\phi_B(I) = \epsilon_B(\delta_B(I))$;

Geodesic opening (reconstruction): $\gamma_{I_{marker}}^{rec}(I_{mask})$;

Geodesic closing: $\phi_{I_{marker}}^{rec}(I_{mask}) = N_{max} - \gamma_{N_{max}-I_{marker}}^{rec}(N_{max} - I_{mask})$.

In these equations, B is the structuring element, I is the image, I_{marker} and I_{mask} are the marker and mask images respectively and N_{max} is the max image where each pixel is represented by its maximum value [195]. The initial stage of this algorithm removes noise while preserving the main structure of the image. As the algorithm intends to find bright linear structures in the image, inverted grayscale images are used as the input:

$$I_{op} = \gamma_{I_0}^{rec}\left(\max_{i=1,2,\dots,n_a}\{\gamma_{L_i}(I_0)\}\right), \quad (4.1)$$

where L_i is a linear structuring element with a constant pixel length (p_l) and different orientations defined by $i \times a_0$, and a_0 is the angle step ($n_a = 360^\circ/a_0$). In this research, we set the default values: $p_l = 5$ and $a_0 = 10^\circ$. However, they can be adjusted depending on the hair density in each image.

In the next step, all the linear bright shapes in the I_{op} image are obtained using:

$$I_{sum} = \sum_i^{n_a} (I_{op} - \gamma_{L_i}(I_0)). \quad (4.2)$$

In I_{sum} , the contrast of the hairs and background as well as some unwanted structures are improved. Therefore, the author proposed filtering the image first using a Gaussian filter

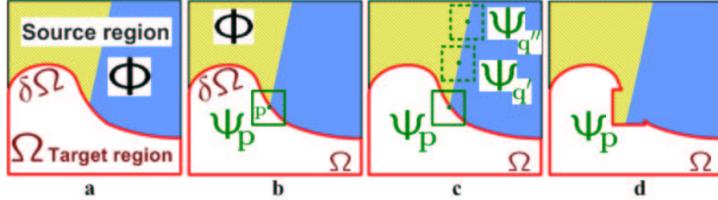


Figure 4.5: (a) Original image with target region Ω , source region Φ and the contour or “fill front” $\delta\Omega$. (b-c) The goal is to synthesize the patch area defined by Ψ_p , using the most likely candidate ($\Psi_{q'}$, $\Psi_{q''}$). (d) Filling region Ψ_p using the best matching candidate and evolving the fill front inward. This figure was taken from Criminisi et al. [55]

$(N(\mu, \sigma) = N(5, 1.25))$, I_G , and then applying the Laplacian operator ($W = 3 \times 3$), I_{lap} . The default values can be adapted based on trial and error to suit different datasets.

Next, in the image obtained, I_{lap} , we remove the noise again by using geodesic opening and closing:

$$I_1 = \gamma_{I_{lap}}^{rec} \left(\max_{i=1:n_a} \{ \gamma_{L_i}(I_{lap}) \} \right), \quad (4.3a)$$

$$I_2 = \phi_{I_1}^{rec} \left(\min_{i=1:n_a} \{ \phi_{L_i}(I_1) \} \right), \quad (4.3b)$$

$$I_{res} = \left(\max_{i=1:n_a} \{ \gamma_{L_i}^2(I_2) \} \right). \quad (4.3c)$$

The final result is achieved by thresholding the last stage $I_{res} \geq 1$. For our purposes in the default algorithm only, the connected components with eccentricity higher than 0.87 are considered. This value is selected since hairs are normally represented by connected lines than circles. The components selected are dilated with a disk structure of size 3 px to create our final hair mask. The final step is optional and can be eliminated in some cases.

In order to restore the original images without hairs, we use the previously created mask and the exemplar-based inpainting method by Criminisi et al. [55]. The exemplar-based method combines the advantage of the “texture synthesis” and “inpainting” methods. Similar to inpainting, it pays special attention to linear structures that influence the fill order for an exemplar-based texture synthesis algorithm [55]. Fig. 4.5 from the original paper demonstrates the main idea of this algorithm.

Considering the source region as the entire image minus the target region ($\Phi = I - \Omega$) and choosing a specific size of the template, all the border points are given a priority order, so that their patch region can be filled with the most similar patch from the source region. The priority

order is based on the so-called confidence term $C(p)$ and the data term $D(p)$:

$$P(p) = C(p)D(p), \quad (4.4a)$$

$$D(p) = |\nabla I_p^\perp \cdot n_p|/\alpha, \quad (4.4b)$$

$$C(p) = \frac{\sum_{q \in \Psi_p \cap \Omega} C(p)}{|\Psi_p|}, \quad (4.4c)$$

where $|\Psi_p|$ is the area of Ψ_p , α is the normalization factor, n_p is the unit vector orthogonal to the fill front ($\delta\Omega$) at point p , and ∇I_p^\perp illustrates the direction and intensity at this point. After prioritizing the pixels on the border, each patch is replaced by a patch from the source region most similar to it. The distance between two patches is based on the sum of square differences of pixels already filled in the two patches.

Fig. 4.6 shows some results of applying our hair removal algorithm with the default parameters. As shown in Fig. 4.6c and 4.6g, the algorithm may not be optimal if the mask detected in the first stage is not optimal. In such cases, the optional morphological operations at the last stage of the hair detection algorithm could be removed or the default parameters adjusted.

Segmentation

A summary of segmentation and border delineation methods of pigmented skin lesions was discussed in Sect. 3.2. In this research, a fusion of a region-based and intensity-based segmentation methods is proposed. Our segmentation fusion is based on a region-based level-set method proposed by [117], Fuzzy-C-means (FCM) [29], and a probability density function (pdf)-based method.

This part of the framework was developed using the Vienna dataset, which is very challenging for border delineation due to image illumination variations. Without applying any illumination corrections, our first approach used a pdf-based method. We were able to correctly segment 95.428% of the dataset (corresponding to 5093 out of 5337 images). The 43 unsegmented images were excluded due to some severe cases of saturation or low contrast. However, after combining the level-set, pdf-based, and FCM algorithms, we were able to segment 5310 images out of 5337 cases (corresponding to 99.49%).

In the rest of this section, first each of the aforementioned methods and finally our fusion algorithm are explained.

pdf-based method is based on analyzing the pdf of a gray-scale, or single-channel, image. It uses the assumption that the lesions appear in darker colors in comparison with the skin, thus they have a separate intensity profile towards the darker side of the histogram. Based on this assumption, a Gaussian mixture with two components is fitted to the pdf of the

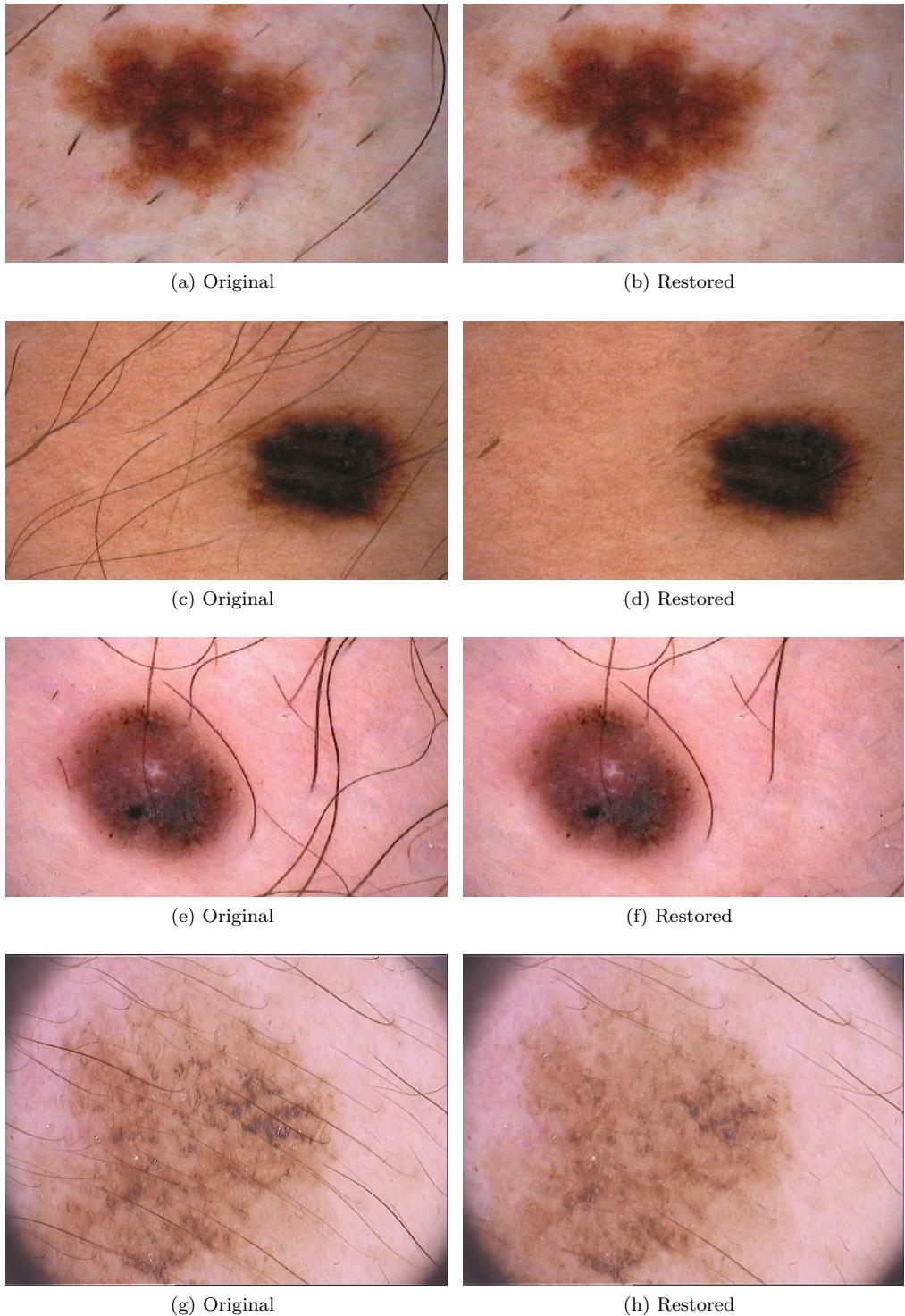
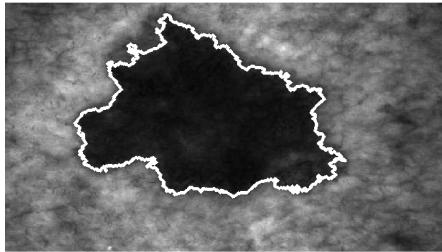
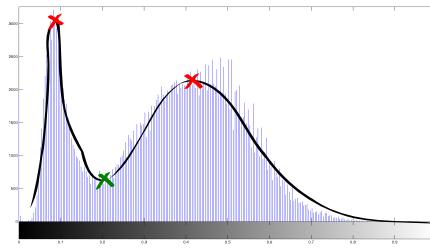


Figure 4.6: Sample results of applying the proposed hair removal algorithm. The algorithm was used with default values for all the images. Images (a, c, e) are from the Vienna dataset and the last image (g) is from the PH² dataset.



(a) Lesion appearance of the Z channel. The white delineation corresponds to our segmentation results.



(b) The pdf corresponding to Fig 4.7a.

Figure 4.7: Illustration of the Z component (*CIE XYZ* color space) of a dermoscopic image (a) with its corresponding pdf (b). The distribution is characterized by a Gaussian mixture with two components. The best threshold is located at the valley between the two Gaussian bells (red cross between the peaks). The threshold in this case was found through the peak detection algorithm and finding the local minimum value.

image, and the local minimum separating the two Gaussians is selected as the optimal threshold. The local minimum is determined by finding the peaks of the smoothed pdf¹. Due to illumination variations in the images, the pdf often does not consist of a Gaussian distribution, and hence, does not have a local minimum. In such cases, the algorithm considers a default threshold, which can be determined empirically.

In this study, it was observed via empirical validation that the *CIE XYZ* color space was the most suitable channel for lesion segmentation. Figure. 4.7b shows the pdf of the Z -channel where the first Gaussian corresponds to the pixel intensities of the lesion and the second Gaussian corresponds to the pixel intensities belonging to the skin.

Thus, as mentioned before, finding the valley between these two Gaussian bells allows us to separate the lesion from the skin (see Fig. 4.7a).

Here, the default threshold, considering the utilization of the Z -channel, was set to 0.2. Some results obtained by this algorithm are shown in the second column of Fig. 4.9. The lesions in Figs. 4.9b and 4.9f are under-segmented, while the lesion in Fig. 4.9n is over-segmented. However, the lesions in Figs. 4.9j and 4.9r are perfectly delineated.

Level-set methods are based on the general principle of the active contour, or snakes, approaches [109]. These methods looks for a contour of an object by minimizing an energy function, which either uses edge information or region descriptor information. The level-set method proposed in [117] is region-based because it uses region descriptor as a criterion

¹The pdf is smoothed using cubic interpolation.

to guide the motion of the active contour. If the image domain is illustrated by Ω , the goal of the proposed method is to find a contour C , which divides the image into disjoined regions $\{ \Omega_1, \Omega_2, \dots, \Omega_N \}$. The chosen level-set method is suited to our application because it deals with intensity inhomogeneities. This model defines a local clustering criterion function for the intensities in a circular neighborhood (\mathcal{O}_y) of each point based on two assumptions: (i) the bias image which accounts for intensity inhomogeneities (b) is changing slowly and can be approximated by a constant in a neighborhood of each point; (ii) the pure image without inhomogeneities and noise takes approximately N distinct constant values, c_1, \dots, c_N for N disjoined regions $\Omega_1 : \Omega_N$ [117]. Based on these assumptions the local clustering criterion function is defined as:

$$\mathcal{E}_y = \sum_{i=1}^N \int_{\Omega_i} K(y-x) |I(x) - b(y)c_i|^2 dx , \quad (4.5)$$

Where $K(y-x)$ is a non-negative window or kernel function equal to 0 for any point outside the neighborhood ($x \notin \mathcal{O}_y$). Considering this function, the optimal partitions (Ω_i) of the entire image are defined so that \mathcal{E}_y is minimized for all y in the image. Subsequently, the energy function is taken as the integral of \mathcal{E}_y with respect to y in the image domain. For the two-phase case (binary) segmentation, the level set energy function is defined as:

$$\mathcal{E} = \int \sum_{i=1}^N \left(\int K(y-x) |I(x) - b(y)c_i|^2 dy \right) M_i(\phi(x)) dx, \quad (4.6a)$$

$$\mathcal{E}(\phi, b, c) = \int \sum_{i=1}^N e_i(x) M_i(\phi(x)) dx, \quad (4.6b)$$

$$e_i(x) = \int K(y-x) |I(x) - b(y)c_i|^2 dy, \quad (4.6c)$$

$$e_i(x) = I^2 1_K - 2c_i I(b * K) + c_i^2 (b^2 * K). \quad (4.6d)$$

The equation above illustrates how to rewrite the energy function in a simplified manner. The last rows indicate how e_i can be computed: $*$ is a convolution operation and $1_K = \int K(y-x) dy$, which is equal to 1 in the entire image, except near the boundary of the image domain [117] and M is linked to the Heaviside function. Using the above energy function, the final level-set formulation is represented as:

$$\mathcal{F}(\phi, b, c) = \mathcal{E}(\phi, b, c) + v\mathcal{L}(\phi) + \mu\mathcal{R}(\phi), \quad (4.7)$$

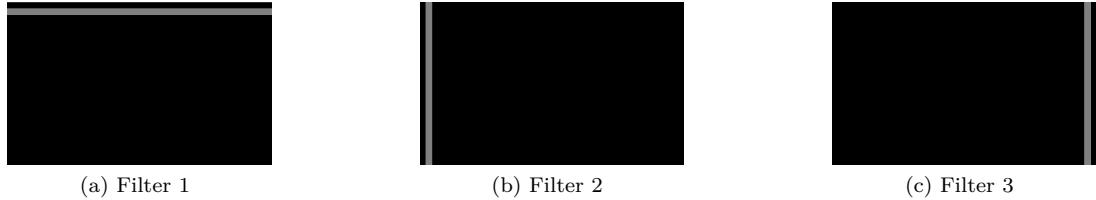


Figure 4.8: The three main filters for discarding over-segmented and wrongly segmented results prior to fusion.

where $\mathcal{L}(\phi)$ and $\mathcal{R}(\phi)$ are the regularization terms [117] (see Eq. 4.8).

$$\mathcal{L}(\phi) = \int |\nabla H(\phi)| dx, \quad (4.8a)$$

$$\mathcal{R}(\phi) = \int p(|\nabla H(\phi)|) dx, \quad (4.8b)$$

$$p(s) = (s - 1)^2 / 2. \quad (4.8c)$$

where, H is the unit step (Heaviside) function.

In this research, an initialization region with a bounding box (200×300) is considered positioned at center of Z -channel. The v is set to 0.001×255^2 , σ is set to 4 and μ is equal to 1. These values were set empirically. The second column of Fig. 4.9 shows some segmentation results achieved by the Level-set method. As shown, this method segments the lesions perfectly in Fig. 4.9a, 4.9e and 4.9m while it fails to segment the lesions in Fig. 4.9i and 4.9q. The difficult and challenging illumination variation and the existing shadows along the left side of the images (Fig. 4.9i and 4.9q) are the main reasons for the failing in the level-set algorithms.

Fuzzy-C-means algorithm [29] is a so called soft data clustering approach. In this technique, contrary to k-means algorithms, each data element is described by a set of membership values, which indicate the strength of its association with each cluster. This membership value is referred to as an association probability between data elements and clusters. Fuzzy clustering uses these memberships to assign the elements to their closest cluster. Considering a set of data elements, $X = \{x_1, \dots, x_n\}$, and a list of clusters, $C = \{c_1, \dots, c_K\}$, a partition matrix $W_{n \times k} = \{w_{ij} \in [0, 1]\}$ is defined so that each of its elements w_{ij} show the degree to which an element x_i belongs to cluster c_j . Thus, the FCM intends to minimize

the following cost:

$$\min_c \sum_{i=1}^n \sum_{j=1}^K w_{ij}^m \|x_i - c_j\|^2, \quad (4.9a)$$

$$w_{ij}^m = \frac{1}{\sum_{k=1}^K \left(\frac{\|x_i - c_k\|}{\|x_i - c_j\|} \right)^{2/m-1}}, \quad (4.9b)$$

The m is called the fuzzifier ($m \geq 1$). Larger values of m lead to a smaller membership and thus a fuzzier clustering [29].

In this study, we performed FCM clustering with $K = 2$, and $m = 2.0$ on the Z -channel of each image. The third column of Fig. 4.9 shows the segmentation results obtained with the FCM algorithm. This algorithm segments the lesions perfectly in Fig. 4.9a, 4.9e, 4.9i, and 4.9m. However, it fails to segment the last lesion in Fig. 4.9q.

Segmentation fusion is considered as combining the segmented results from the three algorithms into one final result. As shown in Fig. 4.9, each algorithm often under-segments (Fig. 4.9f), over-segments (Fig. 4.9n, 4.9t) or wrongly identifies some lesions (Fig. 4.9k).

Although under-segmentation is not a problem and can be corrected by fusion methods, it is preferable to avoid over-segmented or wrongly-segmented results. Also, the initial experiments with the Vienna dataset indicated that often only one algorithm performs correctly while the other two fail. Thus, the majority voting or any other fusion method, such as STAPLE, will not perform efficiently without some preliminary filtering steps.

In this regard, some basic filtering steps are proposed to identify the wrong and over segmented results. Three main filters are used, shown in Fig. 4.8. Prior to applying the filters, the images are first divided into two groups: the first group, the majority group, contains images which the lesions do not touch the borders of the images. The second group, the minority group, contains images of larger lesions. These lesions usually cover the entire image and are connected to the border. Through primary observation it was concluded that the level-set algorithm segments these lesions perfectly. Subsequently, for this group, only this algorithm is used.

While for the former group, the segmented result is multiplied to each of the filters and the number of common pixels between them are measured. This measurement defines a connectivity index (CI) between each filter and a segmented result. Thus, if the segmented results have a high CI with respect to filter 1 and 2 (higher than 60%), the result is discarded as over-segmented and if it only has a high CI with respect to filter 3 (over 40%), it is considered as a wrong segmentation. After discarding the undesired segmentations

Table 4.3: The feature set used in this research.

Feature type	Index
<i>Shape</i>	
Thinness ratio	
Border asymmetry	<i>S</i>
Distance variance of border points to the center [185]	
Statistics of gradient operator along the lesion border [185]	
<i>Color</i>	
Color variance and responses along RGB, HSI and CIELAB	<i>C</i> ₁
Color histogram in RGB ($b = 42$) ¹	
Opponent color space angle and hue histogram ($b = 42$)	<i>C</i> ₂
Color intensities	<i>C</i> ₃
<i>Texture</i>	
Completed Local Binary Pattern ²	<i>T</i> ₁
Gray-Level Co-occurrence Matrix ($\theta = \{0, \pi/4, \pi/2, 3\pi/4\}$, $D = 9$ pxls, $G = 32$) ³	<i>T</i> ₂
Gabor Filter ($s = 4$, $\theta = \{\pi/6, \pi/3, \pi/2, 2\pi/3, 5\pi/6, \pi\}$)	<i>T</i> ₃
Histogram of Oriented Gradients	<i>T</i> ₄
SIFT	<i>T</i> ₅

¹ b stands for number of bins.

² 24 neighbourhood, rotation invariant, uniform and normalized histogram.

³ D stands for distance in pixels and G is the quantized number of grey levels.

for each image, the remaining segmentations are added together. If one or none of the segmentation results are discarded, the final result is based on majority voting of two among all three and if two of them are discarded, the final result is equal to the remaining segmentation.

4.4.2 Mapping

The mapping stage is used to determine a discrete set of elements (or structures) used for the representation of each image/sample. Therefore, two mapping strategies are defined: (i) *global* and (ii) *local* mapping. In the global mapping approach, a single structure is computed for each image while in the local mapping, a set of structures is defined by sliding a window across the image. The final descriptor is then computed based on a single element or a concatenation of elements resulting from the sliding window. The features used to create single or multiple elements are described in the following section.

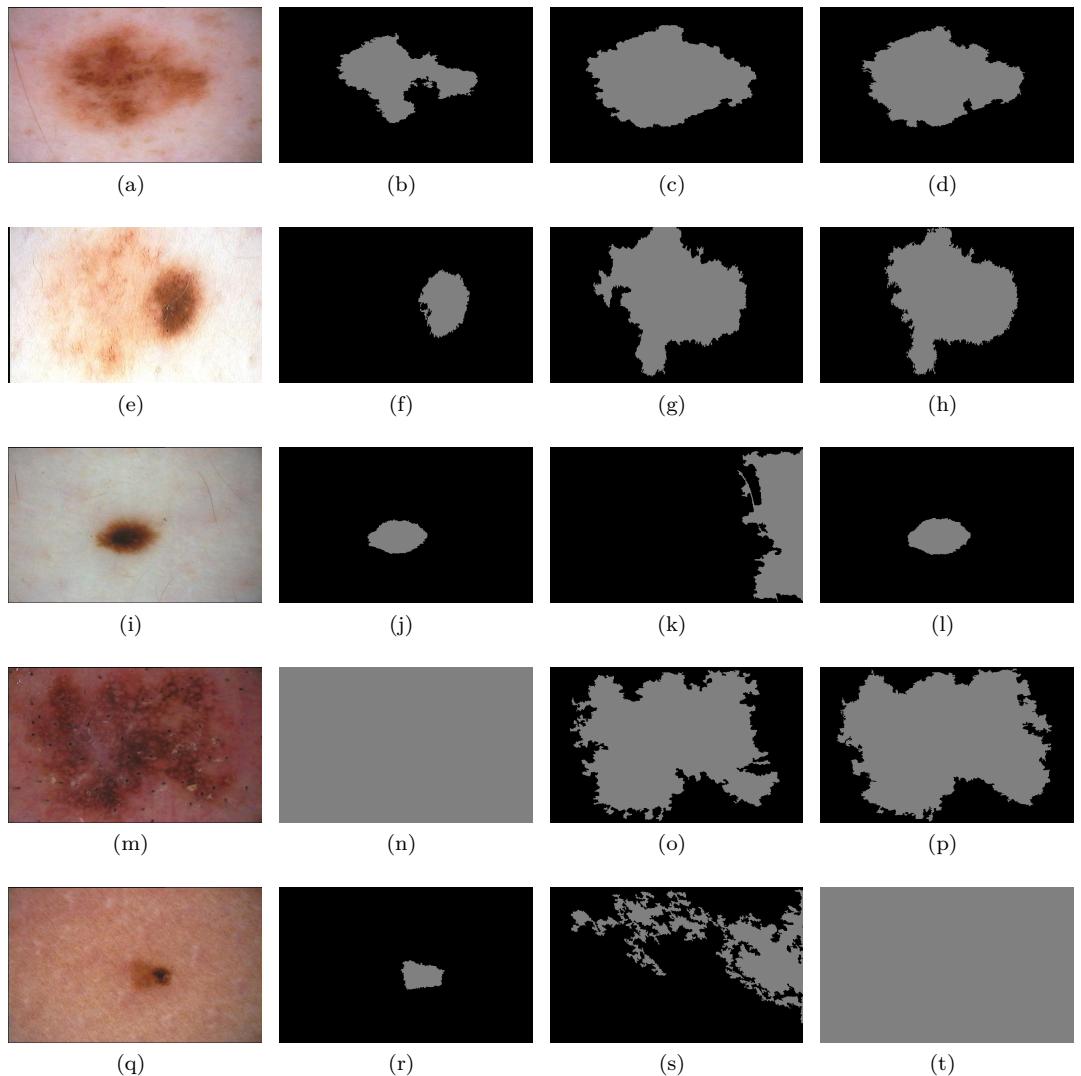


Figure 4.9: Segmentation results obtained by the pdf-based, level-set and FCM algorithms (second, third and forth columns from the left, respectively). The original images are shown in the leftmost column. In order to avoid the influence of the black border in the original images, the segmentation methods ignore 15 pixels along the borders of each image. The results in rows 3–5 demonstrate the methods’ failures.

4.4.3 Feature Extraction

Reviewing the literature, one can observe that some features are used more widely than others, such as color and shape characteristics. This is due to the fact that computerized systems are developed to mimic dermatologists' assessment, by using the discriminative characteristics defined in the clinical "ABCD" rule for instance. Thus, we also incorporate some common shape and color descriptors along with texture features. Shape features (S), color statistics and histograms (C_1), and GLCM features (T_2) are chosen because they were used widely in the past. However, the rest of the descriptors, such as the Opponent Color Space (C_2), color intensities (C_3), LBP (T_1), Gabor Filter (T_3), Histogram of Oriented Gradients (HOG) (T_4), and *SIFT* (T_5) are chosen because they are efficient and well established color and texture descriptors and have barely been used for the detection of melanoma. These features are summarized in Table 4.3 and explained in the following.

Shape (S) features were created with reference to [125]. This group of descriptors measures the thinness ratio, border asymmetry, distance variance of border points to the center and statistics (minimum, maximum, average and variance) of the gradient operator along the lesion's border. The thinness ratio measures the circularity of a lesion: $TR = 4\pi Area/Perimeter^2$. And border asymmetry computes the percent of non-overlapping areas after a hypothetical folding of the lesion around its greatest diameter.

Color Variance and Color Histogram (C_1) is a feature descriptor that has been widely used in the past for the detection of melanoma. This descriptor contains the mean and variance of nine color channels ($R, G, B, H, S, V, L^*, a^*, b^*$) and histograms for R, G and B channels. Each histogram is constructed with 42 bins, leading to a final descriptor of size $(9 \times 2) + (42 \times 3) = 144$.

Opponent Color Space Angle and Hue Histogram (C_2) were first proposed in [192] as local color features. These descriptors were chosen due to their robustness to photometric (shadow, shading, specularities and changes of the light source) and geometrical (viewpoint, zoom and object orientation) variation. These rotation invariant and robust

descriptors are derived from *RGB* channels using the following equations:

$$\begin{pmatrix} \mathcal{O}_1 \\ \mathcal{O}_2 \\ \mathcal{O}_3 \end{pmatrix} = \begin{pmatrix} (R - G)/\sqrt{2} \\ (R + G - 2B)/\sqrt{6} \\ (R + G + B)/\sqrt{3} \end{pmatrix}, \quad (4.10a)$$

$$H^{\mathcal{O}} = \arctan \left(\frac{\sqrt{3}(R - G)}{R + G - 2B} \right), \quad (4.10b)$$

$$\theta_d^{\mathcal{O}} = \arctan \left(\frac{\sqrt{3}(R'_d - G'_d)}{R'_d + G'_d - 2B'_d} \right), \quad (4.10c)$$

Here d denotes the spatial coordinates of (x,y) and R'_d , G'_d , B'_d denote the first order derivatives of *RGB* with respect to the coordinates. This color descriptor is built by taking a 42 bins histogram for the opponent angle $\theta_d^{\mathcal{O}}$ and the hue channel $H^{\mathcal{O}}$, for a final descriptor size of 84 dimensions.

Color intensities (C_3) represent the color information in its simplest form—their intensities. This descriptor concatenates color intensities in R , G and B to create a feature descriptor.

Completed Local Binary Pattern (CLBP) (T_1) is a discriminative rotation invariant feature descriptor proposed by Guo et al. [96]. CLBP is a completed version of LBP [143], especially designed for texture classification. In both descriptors, a central pixel (g_c) in a neighborhood defined by radius R is compared to its neighborhood pixels (g_p , at distance R from the central pixel) and their differences are encoded in terms of binary patterns. The binary patterns are calculated for each pixel in a given image and their histogram defines the final descriptor.

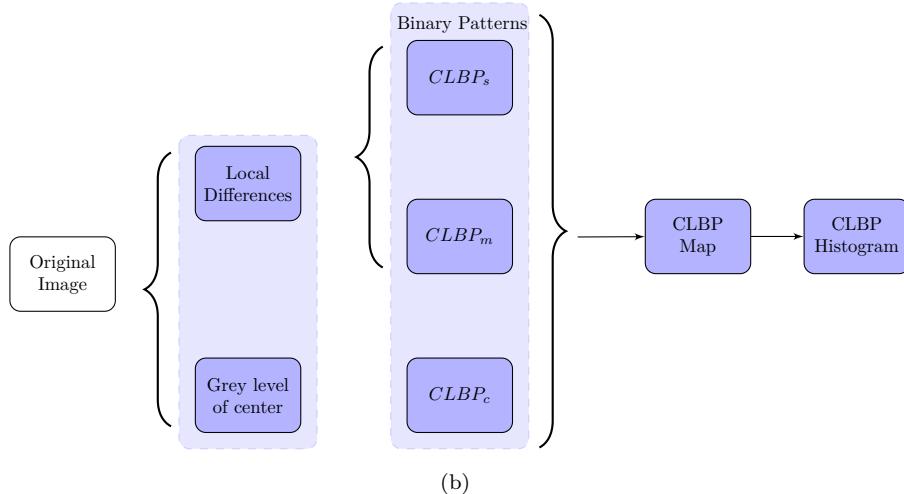
The local patterns of the CLBP descriptor encode the magnitude and sign differences between the central pixel and its neighbors as well as the grey level of the central pixel rather than only the sign differences (general LBP). Figure 4.10 represents this process. The sign $CLBP_S$, magnitude $CLBP_M$, and central grey level $CLBP_C$ binary patterns are created by encoding the local distance components and the central grey levels.

In the proposed framework, the rotation invariant, uniform, and normalized CLBP features with a radius of 3 and 24 samples were calculated. The radius and number of samples were chosen after testing the algorithm with different neighboring samples such as {8, 16, 24} with a radius of {1, 2, 3}, respectively.

Grey-Level Co-Occurrence Matrix (GLCM) (T_2) is one of the earliest texture descriptor methods, proposed by Haralick. et al. [100], widely used for texture analysis applications,

<table border="1"> <tr><td>9</td><td>12</td><td>34</td></tr> <tr><td>10</td><td>25</td><td>28</td></tr> <tr><td>99</td><td>64</td><td>56</td></tr> </table>	9	12	34	10	25	28	99	64	56	<table border="1"> <tr><td>-16</td><td>-13</td><td>9</td></tr> <tr><td>-15</td><td>0</td><td>3</td></tr> <tr><td>74</td><td>39</td><td>31</td></tr> </table>	-16	-13	9	-15	0	3	74	39	31
9	12	34																	
10	25	28																	
99	64	56																	
-16	-13	9																	
-15	0	3																	
74	39	31																	
Original Image	Local Difference																		
<table border="1"> <tr><td>-1</td><td>-1</td><td>1</td></tr> <tr><td>-1</td><td></td><td>1</td></tr> <tr><td>1</td><td>1</td><td>1</td></tr> </table>	-1	-1	1	-1		1	1	1	1	<table border="1"> <tr><td>16</td><td>13</td><td>9</td></tr> <tr><td>15</td><td></td><td>3</td></tr> <tr><td>74</td><td>39</td><td>31</td></tr> </table>	16	13	9	15		3	74	39	31
-1	-1	1																	
-1		1																	
1	1	1																	
16	13	9																	
15		3																	
74	39	31																	
Sign Component	Magnitude Component																		

(a)



(b)

Figure 4.10: CLBP descriptor process. An example of how local distances, sign and magnitude components are calculated is given in (a), while (b) shows an overall view of the CLBP process.

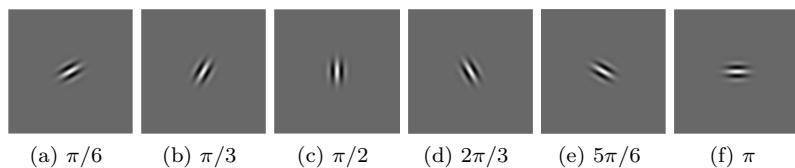


Figure 4.11: The 6 orientations of the Gabor filters.

including melanoma detection. In this approach, texture features are extracted based on the statistical measurement of co-occurrence probabilities. The co-occurrence distribution represents the occurrence probabilities of all the pairwise combinations of the grey levels in a defined window [51]. In other words, it counts how often a pixel with the grey intensity i occurs adjacent to a pixel with the grey intensity j . The spatial distance and orientation of interests between the pixels are defined by the distance D and the angle θ . The co-occurrence probability between grey levels i and j is defined in Eq. 4.11, where P_{ij} represents the conditional probability of occurrences of the grey value i adjacent to the grey value j given the distance D and orientation θ . Here, G is the quantized number of grey levels:

$$C_{ij} = \frac{P_{ij}}{\sum_{i,j=1}^G P_{ij}}. \quad (4.11)$$

The images are quantized to 32 grey levels and the co-occurrence probabilities are calculated given the distance (D) of 9 pixels and four different orientations of $\{\theta = 0^\circ, 45^\circ, 90^\circ, 135^\circ\}$. The final texture descriptor is an average of these four measurements allowing us to obtain a rotation invariant descriptor. For each orientation, a set of 22 texture features (proposed in [100, 51, 184]) are calculated. These features are listed in Table 4.4. The proposed distance and quantization values were chosen after testing a variety of distances $\{1, 3, 7, 9\}$ and quantized grey-levels $\{16, 32, 64\}$.

Table 4.4: The list of GLCM statistics (f_1-f_{22}) [100, 51, 184] used in our experiments. The first section of the table represents the primary measures derived from the co-occurrence matrix C_{ij} for calculating the features. Here, μ_x , μ_y , σ_x , and σ_y are the mean and standard deviation of C_{ij} with respect to i and j , respectively. H_x and H_y are entropy of C_x and C_y

Co-occurrence matrix	$C_{ij} = P_{ij} / \sum_{i,j=1}^G P_{ij}$
$C_x(i)$	$\sum_{j=1}^G C_{ij}$
$C_y(j)$	$\sum_{i=1}^G C_{ij}$
$C_{x+y}(k)$	$\sum_{ij_{i+j=k}} C_{ij}, k = 2, 3, 2G$
$C_{x-y}(k)$	$\sum_{ij_{ i-j =k}} C_{ij}, k = 0, 1, G - 1$
H_{xy}	$-\sum_{ij} C_{ij} \log C_{ij}$
H_{xy1}	$-\sum_{ij} C_{ij} \log(C_x(i)C_y(j))$

H_{xy2}	$-\sum_{ij} C_x(i)C_y(j) \log(C_x(i)C_y(j))$
f_1 : Maximum probability	$\max\{C_{ij}, \forall(i,j)\}$
f_2 : Uniformity	$\sum_{ij} C_{ij}^2$
f_3 : Entropy	$\sum_{ij} C_{ij} \log C_{ij}$
f_4 : Dissimilarity	$\sum_{ij} C_{ij} i - j $
f_5 : Contrast	$\sum_{ij} C_{ij} i - j ^2$
f_6 : Inverse difference	$\sum_{ij} C_{ij} / (1 + i - j)$
f_7 : Inverse difference moment	$\sum_{ij} C_{ij} / (1 + i - j ^2)$
f_8 : Correlation 1	$\sum_{ij} (i - \mu_x)(j - \mu_y) C_{ij} / (\sigma_x \sigma_y)$
f_9 : Inverse difference normalized	$\sum_{ij} C_{ij} / (1 + i - j /N)$
f_{10} : Sum of squares	$\sum_{ij} (i - \mu)^2 C_{ij}$
f_{11} : Inverse difference moment normalized	$\sum_{ij} C_{ij} / (1 + i - j ^2/N^2)$
f_{12} : Sum of average	$\sum_{i=1}^{2N-1} (i+1) C_{x+y}(i)$
f_{13} : Sum entropy	$S_e = -\sum_{i=1}^{2N-1} C_{x+y}(i) \log C_{x+y}(i)$
f_{14} : Sum variance	$\sum_{i=1}^{2N-1} (i+1 - S_e)^2 C_{x+y}(i)$
f_{15} : Difference variance	$\sum_{i=1}^{2N-1} i^2 C_{x-y}(i+1)$
f_{16} : Difference entropy	$-\sum_{i=1}^{2N-1} C_{x-y}(i+1) \log C_{x-y}(i)$
f_{17} : Information measure of correlation 1	$H_{xy} - H_{xy1} / \max(H_x, H_y)$
f_{18} : Information measure of correlation 2	$\sqrt{1 - e^{-2(H_{xy2} - H_{xy})}}$
f_{19} : Auto correlation	$C = \sum_{ij} ij C_{ij}$

f_{20} : Correlation 2	$(AC - \mu_x\mu_y)/\sigma_x\sigma_y$
f_{21} : Cluster shade	$\sum_{ij}(i + j - \mu_x - \mu_y)^3 C_{ij}$
f_{22} : Cluster prominence	$\sum_{ij}(i + j - \mu_x - \mu_y)^4 C_{ij}$

Gabor Filter (T_3) is a linear filter which extracts edges and texture information from the image and was found to be similar to human visual perception [130]. The Gabor filter is defined as a modulation of a Gaussian kernel with a sinusoidal wave. The Gabor expression is shown in Eq. 4.12. This function is basically a Gaussian with standard deviations of σ_x and σ_y that vary along the x and y axes while it is modulated by a complex sinusoidal with a wavelength of λ . In this equation, θ represents the orientation of the Gabor filter, ψ is the phase offset, and s is the scale factor.

$$g(x, y) = \exp\left(-\left(\frac{x'^2}{2\sigma_x^2} + \frac{y'^2}{2\sigma_y^2}\right)\right) \cos\left(2\pi\frac{x'}{\lambda} + \psi\right), \quad (4.12a)$$

$$x' = s(x \cos \theta + y \sin \theta), \quad (4.12b)$$

$$y' = s(-x \sin \theta + y \cos \theta). \quad (4.12c)$$

In this work as described by [129], the images are convolved with a set of Gabor filters characterized by different orientations and scales. The features are created by considering the mean and variance of the resulting filtering. We used 6 different orientations ($\{\pi/6, \pi/3, \pi/2, 2\pi/3, 5\pi/6, \pi\}$) along 4 scales, downsizing by a factor of two between each scale. The first 6 filters are shown in Fig. 4.11.

Histogram of Oriented Gradients (HOG) (T_4) was proposed in [56]. In simple terms, HOG counts the occurrences of gradient orientations in localized patches of an image. This descriptor is invariant to geometric and photometric transformations. HOG calculates the gradient values of the image by applying derivative filters ($[-1, 0, 1]$, $[-1, 0, 1]^T$) or the Sobel mask to the image and measuring the magnitude and orientations of the gradients. Then, it creates the weighted histograms of the gradient orientations in the image patches. The orientations are weighted either by the magnitude of gradients or by a function of gradient magnitudes. In order to account for changes in illumination and contrast, the patches are grouped into larger spatially connected blocks and the final descriptor is formed by a vector of the normalized cell histograms from all the blocks.

Scale-Invariant Feature Transform (SIFT) (T_5) proposed by Lowe et al.[120] is a rotation, translation and scale invariant feature descriptor, used successfully by the computer vision society. To identify the key-points, this descriptor first convolves the original image ($I(x, y)$) with Gaussian kernels ($G(x, y, k\sigma)$) at different scales ($k\sigma$).

$$L(x, y, k_i\sigma) = G(x, y, k_i\sigma) * I(x, y) . \quad (4.13)$$

Then it creates the difference (D) of successive Gaussian blurred images (difference of gaussian (DoG)) and identifies the key-points as the local maxima and minima of DoG images across the scales.

$$D(x, y, \sigma) = L(x, x, k_i\sigma) - L(x, x, k_j\sigma) . \quad (4.14)$$

Essentially each pixel in DoG images is compared with its 8 neighbor pixels at the same scale and 9 corresponding neighbor pixels in the neighboring scales. This pixel is identified as the key-point if it has the maximum or minimum value among all its neighbors.

In the next step, the identified key points are filtered to reject the ones with low contrast or those along the edges. The remaining key points are then assigned one or more orientations based on the local image gradient directions. In this step, the gradient magnitude and orientation of each neighboring pixel of the key point in the Gaussian blurred image L is calculated. For the sake of scale invariance, image $L(x, y)$ is selected at the same or closest scale (σ) of the key point of interest ($L(x, y, \sigma)$). The gradient magnitude and orientation of all the neighboring pixels in image $L(x, y, \sigma)$ are computed as follows:

$$m(x, y) = \sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2}, \quad (4.15a)$$

$$\theta(x, y) = \tan^{-1} \left(\frac{L(x, y+1) - L(x, y-1)}{L(x+1, y) - L(x-1, y)} \right). \quad (4.15b)$$

Then, using the computed magnitude and orientations of the pixels within a Gaussian circular weighted window G_L around the key point, an orientation histogram is formed. This histogram consists of 36-bins, each covering 10° . The orientation of each pixel in the neighborhood is accumulated to the histogram while weighted by its gradient magnitude and G_L . Finally, the key point is assigned with the dominant orientation corresponding to the highest peak of the histogram as well as the local peaks within 80% of the highest one (up to two peaks in total). In this way, there can be several key points at the same location and scale but with different orientations. The final key point descriptor is then created by taking a 16×16 neighborhood around the key point. This neighborhood is divided into 16 sub-blocks of 4×4 , and for each sub-block, an 8-bin orientation histogram

is created. Concatenation of these histograms (i.e. 16 regions) leads to a 128-dimensional feature vector as the final descriptor.

4.4.4 Feature representation

Different approaches for feature representation were discussed extensively in Sect. 3.4. As mentioned before, the main goal of this step is to represent the extracted features in the most efficient and effective manner. In our framework, we categorized the feature representation methods employed into two main groups: (i) low-level and (ii) high-level feature representation.

Low-level representation refers to the cases in which the features are represented in their original space, or in terms of some statistics derived from the original feature space, such as moments, histograms, or simply a concatenation of features from the local structures.

High-level representation. According to the mapping strategy chosen, low-level representation can lead to a high-dimensional feature space. High-level representation simplifies this space into a more discriminant lower space. Among the methods mentioned in Sect. 3.4, PCA, BoW, and Sparse Coded Features (SCF) can be used for this purpose.

Although the optimal number of PCA dimensions can be found by considering the minimum dimensions of the eigenvectors, which account for 95% of the sum of all the eigenvalues, in some experiments where PCA was employed, we chose to find this number via an exhaustive search through the feature dimensions. In this manner, all the possible numbers of dimensions (from 1 to the original size) are tried and the optimal number is found when the best classification accuracy is achieved. In a similar way, we found the optimal number of clusters for the BoW approach. As for the SCF approach, three sparsity levels $\lambda = \{2, 4, 8\}$ and different numbers of atoms $K = \{100, 200, \dots, 1000\}$ were considered.

4.4.5 Data Balancing

As noted before in Sect. 3.5, medical data, similar to other real datasets, face the class imbalance problem, and melanoma datasets are not an exception. Fig. 4.1 clearly demonstrates this fact. Our literature study showed that despite extensive research in melanoma classification, only a few studies tackled the issue of an imbalanced dataset [17, 38]. Barata *et al.* generated new synthetic samples by adding Gaussian noise with fixed parameters to the minority class samples [17]. Celebi *et al.* and Capdehourat *et al.* over-sampled their dataset using SMOTE [47] to improve the SE of their algorithm [38, 35]. We are convinced that this is an important step, and the effect of imbalanced data on the classification result should be considered. Therefore, we first balance the majority to a minority class ratio in the data space, and then explore the balancing strategies in the feature space (Sect. 3.5).

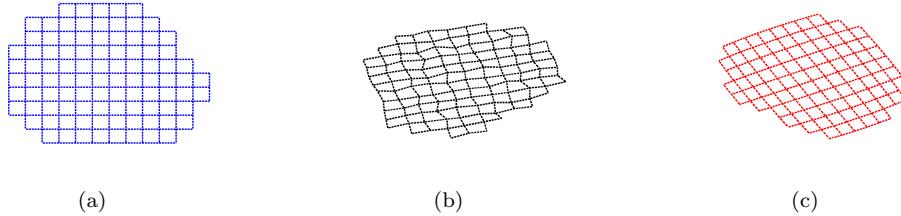


Figure 4.12: Data space transformation: (a) original synthetic data, (b) RDGM deformation, (c) BD deformation.

Data space over sampling DOS

These methods are related to the generation of new synthetic samples by modifying the original data before doing any feature extraction processes. DOS is performed on the original dataset by generating synthetic melanoma images based on the two types of deformation:

Random deformation using gaussian motion (RDGM) . Achieved by adding a random Gaussian motion $\mathcal{N}(\mu, \sigma) = (0, 5)$ at each pixel compounded with a global rotation of 80° .

Barrel deformation (BD) . Corresponds to a deformation of the original image using barrel distortion compounded with a global rotation of 145° .

These deformations are used because they are more likely to occur due to a non-planar surface of some body parts, skin wrinkles, and camera orientation. Furthermore, cubic b-spline interpolation is used with both methods to approximate non-integer points in the images. A synthetic example illustrating the results of these deformations is shown in Fig. 4.12.

Feature space (over/under)-sampling

These strategies are discussed extensively in Sect. 3.5. In the over-sampling approaches, such as SMOTE and Random Over-Sampling (ROS), minority samples are selected or synthetically generated to match the number of majority samples. Under-sampling approaches reject majority samples to match the number of minority samples. In this research, we tested all the over- and under-sampling methods mentioned in Sect. 3.5, with the value of k set to 3 for SMOTE, Neighborhood Cleaning Rule (NCR), NearMiss-1 (NM1), NearMiss-2 (NM2), and NearMiss-3 (NM3) algorithms.

4.4.6 Classification

In Sect. 3.6, we discussed different single learner and ensemble approaches. Among the aforementioned methods, we chose linear (RBF), kernel SVM and Linear Discriminant Analysis (LDA) as single learners, and Gradient Boosting (GB), RF, and weighted combination as ensembles. These classifiers are compared with each other in different experiments. The RF classifier is trained in all the experiments up to its maximum length without pruning. In the initial experiments, the number of RF trees is set to 500 while in the later ones the RF is trained with 100 trees. The GB classifier is used with an exponential loss function (See Eq. 3.21a), shrinkage (regularization) parameter of 0.01, and sub-sampling fraction of 0.7. The sub-sampling fraction indicates that in each iteration, only 70% of the training samples are randomly considered. The optimum regularization and soft margin parameter of the SVM classifier is set through a grid-search over the validation set.

Depending on the experiments, different cross-validation approaches (leave-one-out-cross-validation (LOOCV) or k fold cross-validation (k-CV)) are used as well. These approaches are explained in the following section along with their corresponding experiments.

Table 4.5: The outline and summary of the experiments performed. (seg) indicates segmentation and ({})) indicates the weighted combination ensemble. (tra., tes., val.) indicate the training, testing and validation percentage of the data, respectively.

	Dataset	Task	Pre-processing	Features	Mapping	Representation	Balancing	Classification	Evaluation tra.,tes.,val.
Experiment#1: Goal: Evaluation of shape, color and texture features, mapping and representation									
	Vienna	M vs D	pdf-seg	S, C_1, C_2 T_1, T_2, T_3 T_4, T_5	global local	High: BoW, PCA		RF GB SVM	10 round SE,SP 70%,15%,15%
Experiment#2: Goal: Effect of balance training and DOS balance strategy									
	Vienna	M vs D	Fusion-seg	S, C_1, C_2 T_1, T_2, T_3 T_4, T_5	global	Low	DOS	RF SVM	10 round SE,SP 60%,20%,20%
Experiment#3: Goal: Evaluate the effects of ensembles and our selected features on new dataset									
	PH ²	M vs D,B	seg	S, C_1, C_2 T_1, T_2, T_3 T_4, T_5	global	Low	DOS	RF {RF,SVM,LDA}	LOOCV SE,SP
Experiment#4: Goal: Evaluate the effects of balancing strategies									
	PH ²	M vs D,B	seg	C_1, C_2 T_1, T_3	global	Low	DOS, ROS, Tomek-Link (TL) SMOTE, Random Under-Sampling (RUS) NCR, NearMiss (1-3) SMOTE+Edited Nearest Neighbor (ENN) SMOTE+TL	RF	10-cv SE,SP, C 80%,20%
Experiment#5: Goal: Evaluate the effects of Sparse coding									
	PH ²	M vs D,B		C_1, C_2 C_3, T_5	local	High: SCF	ROS	RF	10-cv SE,SP 80%,20%

4.5 Experiments and Results

In this section, we present our tests and experiments with their corresponding results. Each subsection describes an experiment with its results, discussion and a brief conclusion. As noted before, we started our research using the Vienna dataset and later switched to the PH² dataset. Thus, initial experiments were performed on the Vienna dataset while the rest were based on PH². The experiments listed in the following subsections illustrate our progress in this research and our final framework, perfectly.

4.5.1 Experiment #1

Melanoma vs. Dysplastic Nevi. Balanced Subsets of the Vienna Dataset

Experiment #1 was performed on a subset of the Vienna dataset, in which we proposed a classification framework for differentiating M vs D lesions. This is a challenging task that has not been addressed sufficiently by the research community.

This experiment evaluated the performance of color and texture features, local and global mapping, BoW high-level representation as well as different classifiers for the melanoma vs. dysplastic lesion classification problem. Prior to the feature extraction, the lesions were segmented using the pdf-based method only. Then, the individual features described in Sect. 4.4.3 (except C_3) and their combinations were extracted using local and global mapping. Globally mapped features were extracted from the bounding box around the segmented lesions, while the locally mapped features were extracted from the patches within a grid. A grid was centered on each segmented lesion and only the patches in which the proportion of the lesion was greater than one third of the patch size (20 px \times 20 px) were selected for feature extraction. The extracted features were then presented in low- and high-level forms. PCA was used to reduce the dimensions of the globally mapped features, while BoW was used to simplify the dimensionality of the locally mapped features to a more discriminative space. The optimum dimensionality of the PCA was selected via an dichotomic search over the feature dimensions (i.e. ranging from quarter to three quarter of the original feature dimension). Similarly, the optimum number of “visual words” for the BoW approach was found empirically for each feature set and classification approach. Depending on the classifier and feature set, this number varied from a small (e.g. 15) to a large (e.g. 700) number of “visual words”. This was expected because classifiers such as RF can handle higher feature dimensions contrary to classifiers such as SVM that perform better on low-dimensional features. The extracted features were then classified using three classifiers: RF, GB, and RBF-SVM.

In this experiment, we did not explore the balancing strategies, however, in order to have a balanced dataset, we randomly selected 10 subsets from the Vienna dataset. Each subset was composed of a total of 180 lesions, including 90 melanoma and 90 dysplastic lesions. While the

Table 4.6: Experiment #1. The top of the table represents the classification results using globally mapped features with three classifiers: Random Forests (RF), Gradient Boosting (GB), and support Vector Machine (SVM) with RBF kernel. The bottom of the table represents the classification results of the combined texture features ($\{T_1, T_2, T_3, T_4\}$) over 10 sets. The results are given in terms of the sensitivity (SE) and specificity (SP) percentages.

Features	RF		GB		SVM	
	SE	SP	SE	SP	SE	SP
Individual Features						
T_4	90.77	60.00	92.31	51.54	86.92	66.15
C_2	86.15	58.46	86.15	56.15	83.07	60.77
C_1	89.23	55.38	83.85	62.31	79.23	73.85
T_3	81.54	68.46	83.85	67.69	82.31	66.15
T_1	80.77	65.38	82.31	69.23	84.61	56.15
T_2	87.69	60.77	81.54	56.92	76.15	63.85
S	60.77	56.92	60.00	56.15	56.15	62.31
Combined Features						
C_1, C_2	93.85	53.85	90.00	59.23	86.92	78.46
C_1, C_2, S	93.85	53.08	91.54	59.23	86.92	63.85
T_1, T_2	90.00	75.38	89.23	66.92	86.92	73.08
T_1, T_3	90.00	70.77	93.85	73.85	92.31	69.23
T_1, T_4	93.08	64.61	92.31	75.38	89.23	65.38
T_1, T_2, T_3, T_4	98.46	70.00	96.15	68.46	92.31	70.00
T_1, T_2, C_1, C_2	95.38	71.54	96.15	70.00	93.08	73.85
T_1, C_1, C_2	96.15	73.08	93.85	75.38	94.61	75.38
T_4, C_1, C_2	97.70	69.23	95.38	70.00	89.23	66.92
$T_1, T_2, T_3, T_4, C_1, C_2$	95.38	73.08	96.92	73.08	96.15	73.08
$T_1, T_2, T_3, C_1, C_2, S$	97.69	62.30	96.15	64.61	93.08	73.85
$T_1, T_2, T_3, T_4, C_1, C_2, S$	96.15	69.23	97.69	69.23	96.15	76.92
RF- $\{T_1, T_2, T_3, T_4\}$						
SE	100	92.31	100	100	92.31	100
SP	76.92	61.54	69.23	53.85	69.23	84.62
					61.54	61.54
					69.23	92.31
						100

Table 4.7: Experiment #1. Classification results using the BoW representation of locally mapped features and three classifiers: Random Forests (RF), Gradient Boosting (GB), and support Vector Machine (SVM) with RBF kernel. The results are indicated in terms of the sensitivity (SE) and specificity (SP) percentages.

Features	RF		GB		SVM	
	SE	SP	SE	SP	SE	SP
Individual Features						
T_1	80.77	68.46	80.77	66.15	88.46	46.92
T_4	78.46	50.77	76.15	56.16	78.46	53.85
T_3	75.38	65.38	72.31	60.00	81.54	66.92
C_2	81.54	53.85	72.31	63.85	63.85	60.00
C_1	72.31	67.69	68.46	67.69	70.77	57.69
T_2	70.70	55.38	66.15	60.77	57.69	46.15
T_5	71.54	73.85	73.85	70.00	73.85	70.00
Combined Features						
C_1, C_2	82.31	60.77	76.15	69.23	71.54	68.46
T_1, T_2	78.46	55.38	76.15	64.62	73.08	53.08
T_1, T_3	84.62	56.92	79.23	66.92	84.62	56.15
T_1, T_4	73.85	59.23	76.92	60.77	87.69	44.62
T_1, T_2, T_3, T_4	73.85	65.38	71.54	64.62	72.31	54.62
T_1, T_2, C_1, C_2	74.62	70.77	69.23	70.00	63.85	64.62
T_1, T_2, T_3	78.46	71.54	76.92	63.85	76.15	63.85
T_1, C_1, C_2	83.08	64.62	76.15	76.92	74.62	78.46
T_4, C_1, C_2	72.31	70.77	69.23	62.31	70.00	62.31
T_1, T_2, T_3, C_1, C_2	78.46	66.15	70.77	70.00	60.00	79.23
$T_1, T_2, T_3, T_4, C_1, C_2$	80.00	63.85	75.38	73.08	65.38	73.85

melanoma images were the same in all the subsets, the dysplastic moles were randomly selected from the 950 nevi in the dataset. For each subset, 70% of the data were used for training, 15% for the validation, and 15% for testing.

The validation set was only used for the SVM classifier to optimize its parameters, C and λ , via a grid-search. The validation set was ignored for RF and GB because RF uses bootstraps to create its tree and GB randomly selects a subset of the training data for each split, so both methods automatically generalize the data and do not need an additional validation set.

Nineteen feature sets (individual as well as combinations) were mapped globally and tested using the three classifiers. The corresponding results are shown in Table 4.6. In this table, the top section represents the classification results with all the features, while the bottom part represents the results of the combined texture features ($\{T_1, T_2, T_3, T_4\}$) over 10 sets. Table 4.7 shows the classification results with eighteen feature sets, excluding shape and including SIFT features, mapped locally and represented using the BoW approach.

Discussion

In the automated classification of malignant lesions such as melanoma, correct classification of cancer cases has high importance. For this reason, high sensitivity is considered as the priority measure. In Table 4.6, the individual features are sorted with reference to their performance. According to the results listed, RF yields high sensitivities for most cases (12 out of 19), outperforming GB and SVM. The highest SE is achieved for the combination of the texture features (T_1 , T_2 , T_3 , T_4) using the RF classifier (98.46%). In this case, the classifier reached a specificity of 70%, which is a very good result considering the difficulties of distinguishing melanoma from dysplastic nevi.

The classification results of this feature set over 10 randomly selected subsets are shown in the bottom part of Table 4.6. These measurements highlight how datasets can affect the classification results and how complicated it is to have a fair method comparison. The comparison of individual global features shows that HOG (T_4) outperforms all the other features. It is followed by opponent color angle and hue histogram (C_2), color statistics and RGB histogram (C_1) and Gabor filter (T_3) features. Taking into account the machine learning concepts, one can argue that SVM provides the best results considering that the difference between SE and SP is rather small. However, as stated above, in the case of cancer classification, it is more important not to miss any cancer case than to classify non-cancer cases.

Observing the results given in Table 4.6, it is obvious that texture features in general outperform color features. The results achieved by the combination of color features (C_1 , C_2) are highlighted using an italic font and the combination of texture features (T_1 , T_2 , T_3 , T_4) are underlined. It can be seen that the three classifiers perform better and achieve higher SE and SP with texture features as opposed to color descriptors. This could be explained by the difference in illumination and image acquisition procedures. The images in the dataset were obtained under different conditions without color calibration, so their color information is not accurately represented by the feature descriptors. This issue and the problems of faithful color reproduction by different devices were addressed in [151, 104, 95, 173, 77].

Table 4.7 shows the results obtained with the BoW representation of locally mapped features. Here, as in Table 4.6, the individual features are listed with reference to their performance. These results illustrate that CLBP (T_1) has a better performance in comparison with the other individual features (SE and SP of 83.33% and 60.51%, respectively) considering the average of the three classifiers. This result is followed by HOG (T_4), Gabor filter (T_3) and opponent color angle and hue histogram (C_2). The worst performance is demonstrated by all three classifiers with the GLCM (T_2) descriptor. The two highest individual performances were achieved using the T_1 and C_2 descriptors with the SVM and RF classifiers, respectively. These descriptors resulted in the sensitivity of 88.46% and 81.54%, respectively, although the SP of SVM with T_1

is very low (see Table 4.7).

It can also be observed that the combination of locally mapped features did not affect the SE as much as the the SP. The best results were obtained using the combination of two texture features, CLBP and Gabor filter (T_1, T_3). With this feature set, the RF classifier achieved the SE of 84.62% and SP of 56.92%. The combination of T_1 , C_1 and C_2 demonstrated the second highest SE of 83.08% with the SP of 64.62%. Finally, the SIFT features showed a stable performance with the three classifiers, with SE and SP around 73%. However, it falls behind the other feature sets with respect to the SE.

Overall, the globally mapped features result in a much better classification outcome than the BoW representation of locally mapped features. Moreover, the RF classifier proved its potential by having the best results in 12 out of the 19 and in 14 out of the 18 experiments for the first and second feature tests, respectively.

Conclusion

In this experiment, our automated framework for classification of melanoma against dysplastic nevi was tested. It evaluated the performance of locally and globally mapped features, and it was observed that the latter perform better. This is likely due to the quality of the images as the dataset was created back in 2001 with an older dermoscope and under varying illumination conditions. The results demonstrate the potential of texture features for melanoma detection and highlight the efficiency of the descriptors that have not been widely used in this field.

In particular, the HOG feature and the opponent color space outperformed the other descriptors as individual feature sets. The combination of all the texture features had a significantly better performance in comparison with color and shape features as well. In general, texture descriptors result in a better classification performance as opposed to color features. This is due to the fact that color is not an indicative and consistent descriptor throughout the whole dataset because of varying image acquisition conditions and the lack of color calibration. Finally, the assessment of the classifiers in this study confirms the potential of RF in comparison with GB and SVM.

4.5.2 Experiment #2

Melanoma vs. Dysplastic Nevi, Imbalance subset of Vienna Dataset

This experiment considered classification of melanoma against dysplastic lesions using a larger subset of Vienna dataset. This subset contains 98 melanoma and 993 dysplastic lesions.

This experiment was mainly performed to evaluate the impact of our balancing strategy, data space over-sampling (DOS). In order to reduce the effects of the class imbalance problem, we chose to synthetically generate some melanoma lesions using the DOS technique (see Sect. 4.4.5).

Table 4.8: A list of melanoma and dysplastic lesions in the training, testing and validation sets of each subset. 10 different subsets were created by randomly selecting # lesions from the dataset. If a melanoma lesion was selected to be a part of a specific group, its two deformed synthetic samples were also included in that set.

Set	# Melanoma	# $Melanoma_{D1}$	# $Melanoma_{D2}$	# Dysplastic
Training	58	58	58	174
Test	20	20	20	410
Validation	20	20	20	409

As mentioned before, this technique deforms the lesion grid using two approaches of barrel deformation (BD) and random deformation using gaussian motion (RDGM). Figure. 4.13 shows a melanoma lesions after each deformation.

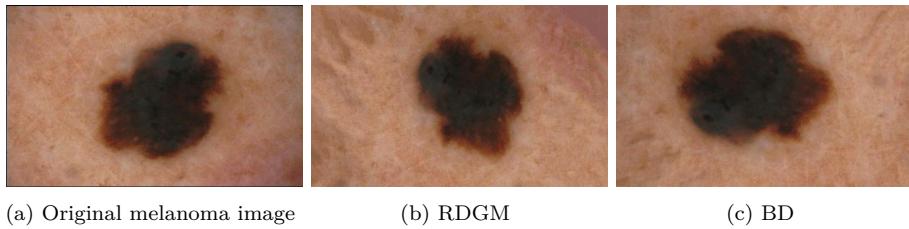


Figure 4.13: Example of a melanoma lesion from Vienna dataset and its deformed generated samples. (a) RDGM, (b) BD.

Using this approach, based on 98 original melanoma lesions, 196 synthetic images were created. From a total of 294 melanoma lesions (98×3), 60% were kept for training, 20% for testing and another 20% for validation. Therefore, to ensure balance training, the training set contained 174 melanoma lesions (58 original melanoma samples randomly selected from 98 lesions and their deformed images) and 174 dysplastic lesions randomly selected from 993 dysplastic nevi. Besides 60 melanoma lesions (20 original samples and their 40 deformed samples), both the validation and test sets contained half of the remaining dysplastic lesions. Table 4.8 tabulates the number of melanoma and dysplastic lesions in each set. Similar to our previous experiment, 10 different subsets were selected, while in each subset, the lesions were randomly selected for each of training, testing, and validation sets. The validation set was only used with the SVM classifier.

In this experiment, only global mapping is used and the features are extracted from a window bounded to lesion boundaries. All the lesions were segmented using a fusion approach rather than only a pdf-based. Individual and combinations of extracted features were further classified using RF and RBF-SVM classifiers. The results obtained are listed in Table. 4.9. The

Table 4.9: The classification results after applying the DOS balancing method to the Vienna M vs. D subset. The data were classified using RF and RBF-SVM. The results are indicated in terms of sensitivity (SE) and specificity (SP) percentages.

Features	RF		SVM	
	SE	SP	SE	SP
Individual Features				
T_1	72.00	77.97	69.00	68.63
T_2	73.50	87.19	70.50	89.32
T_3	74.17	82.96	65.83	74.65
T_4	74.67	90.54	78.00	78.78
C_1	80.33	92.20	61.67	74.40
C_2	80.17	92.05	73.50	76.60
S	72.67	81.25	62.50	69.10
Combined Features				
C_1, C_2	81.17	92.49	67.17	81.42
C_1, C_2, S	80.83	92.74	67.17	78.00
T_1, T_2	76.50	87.73	72.17	85.43
T_1, T_3	75.50	84.40	73.17	74.74
T_1, T_4	79.17	89.22	63.00	73.86
T_1, T_2, T_3, T_4	79.83	91.30	68.00	82.03
T_1, T_2, C_1, C_2	82.00	94.23	73.50	82.76
T_1, T_3, C_1, C_2	81.33	94.47	67.00	80.81
T_1, T_3, T_4	79.50	90.22	64.83	78.63
T_1, C_1, C_2	82.00	94.13	66.50	80.88
T_3, C_1, C_2	80.17	93.35	68.50	79.83
T_4, C_1, C_2	80.83	93.33	67.83	81.96
T_1, T_2, C_1, C_2, S	82.33	94.47	66.50	81.05
$T_1, T_2, T_3, T_4, C_1, C_2$	80.83	95.11	69.67	84.82
$T_1, T_2, T_3, C_1, C_2, S$	81.67	94.74	70.83	82.08
$T_1, T_2, T_3, T_4, C_1, C_2, S$	80.67	95.01	67.83	84.25

best performance concerning individual and combined features are highlighted in bold.

To be able to compare and evaluate the effects of having more melanoma samples in the balance training set, two different cases without using any balancing strategy, beside our initial proposed case with balancing strategy are considered: (i) case 1: DOS-balance training, (ii) case 2: balance-training, and (iii) case 3: imbalance-training. The second case considered balance training only using the original data. Accordingly, in this case, our training set included 58 melanoma and 58 dysplastic, the test set had 20 melanoma and 468 dysplastic, and the validation set contained 20 melanoma and 461 dysplastic. Case 3 considered imbalance training, by removing the synthetic samples from each of the training, testing and validation sets. For the sake of comparison of the three cases, only the results obtained by the best performing

feature sets (i.e. the results highlighted in gray cell colors in Table 4.9) are listed in Table 4.10.

The first part of this table shows the results obtained with the RF classifier for the three cases and the second part shows the results obtained with the SVM.

Discussion

The results listed in Table 4.10 indicate the benefits of DOS-balancing and balanced training in general. It is clear that imbalanced training, as expected, overfits the majority class. In this case, the performance of the SVM was poorer, but RF was also biased towards the majority class.

In the balanced training case (without the DOS balancing), the RF ensemble achieved a mean SE and SP of 69% and 67%, respectively. In turn, the SVM reached an average SE and SP of 61% and 71%, respectively. These results indicate that the classifiers performed slightly better than average, with many false positive (FP) samples (approximately more than 100 dysplastic lesions were classified as melanoma in each test). In comparison, in the DOS-balanced training case, the average SE and SP over all the features, were measured at 79% and 91%, respectively. The results show that besides increasing the SE, the amount of FP predictions was reduced notably (25–37 dysplastic lesions were misclassified). In general, the performance of the RF classifier was superior to the SVM in this experiment.

Conclusion

The tests carried out in this experiment highlight the importance of a balanced training and balancing strategies. Data space over-sampling method granted the possibility of a balanced training using more samples from the dataset without reducing the test set. Increasing the training set by a factor of 3 improved the results remarkably. And it may be expected that balanced training with more samples will lead to even better results.

4.5.3 Experiment #3

Melanoma .vs Benign and Dysplastic nevi, Ensemble approach on PH² dataset

This experiment was conducted to evaluate the impact of the ensembles and features selected in the experiments #1 and #2 on the PH² [135] dataset. This dataset contains 40 melanoma, 80 benign and 80 dysplastic lesions with their respective groundtruth segmentations, so there was no need to apply our segmentation algorithm. In this experiment, we used a subset of 193 images, which included 39 melanoma, 78 benign and 76 dysplastic lesions. Seven images were excluded due to artifacts.

Table 4.10: Comparison of the classification results on the DOS-balanced, balanced, and imbalanced training cases. Each case was tested with the RF and SVM classifiers. The first part of the table shows the results obtained with RF, while the second part lists the results achieved with SVM. The results are indicated in terms of the sensitivity (SE) and specificity (SP) percentages.

Features	RF classifier					
	DOS-balanced training		balanced training		imbalanced training	
	SE	SP	SE	SP	SE	SP
Individual Features						
T_3	74.17	82.96	63.50	64.82	23.00	94.35
T_4	74.67	90.54	59.50	62.48	20.00	93.64
C_1	80.33	92.20	74.00	71.39	39.50	92.96
C_2	80.17	92.05	71.00	58.44	38.00	92.84
Combined Features						
C_1, C_2	81.17	92.49	74.50	64.58	43.00	93.72
T_1, T_2, C_1, C_2	82.00	94.23	73.50	72.91	46.50	95.06
T_1, T_3, C_1, C_2	81.33	94.47	76.00	73.28	44.00	95.57
T_1, C_1, C_2	82.00	94.13	76.00	72.78	44.50	95.21
T_1, T_2, C_1, C_2, S	82.33	94.47	75.50	72.42	43.50	95.33
$T_1, T_2, T_3, T_4, C_1, C_2$	80.83	95.11	75.00	72.59	43.00	95.84
$T_1, T_2, T_3, C_1, C_2, S$	81.67	94.74	74.50	72.96	44.00	95.45
SVM classifier						
Features	DOS-balanced training		balanced training		imbalanced training	
	SE	SP	SE	SP	SE	SP
	65.83	74.65	71.00	58.89	4.50	99.00
T_3	78.00	78.78	47.50	62.33	5.00	97.53
T_4	61.67	74.40	45.00	85.76	3.00	99.49
C_2	73.50	76.60	66.50	65.20	4.00	99.63
Combined Features						
C_1, C_2	67.17	81.42	52.00	80.26	0.00	100.00
T_1, T_2, C_1, C_2	73.50	82.76	56.50	78.29	0.00	100.00
T_1, T_3, C_1, C_2	67.00	80.81	65.50	74.99	0.00	99.98
T_1, C_1, C_2	66.50	80.88	68.50	76.60	0.50	99.95
T_1, T_2, C_1, C_2, S	66.50	81.05	56.00	81.48	0.00	100.00
$T_1, T_2, T_3, T_4, C_1, C_2$	69.67	84.82	63.50	75.87	13.00	97.33
$T_1, T_2, T_3, C_1, C_2, S$	70.83	82.08	63.00	76.70	6.00	98.73

Table 4.11: Benchmark results on the PH² dataset using locally and globally mapped color and texture features [17]. The results are given in terms of the SE and SP percentages.

		Features			
		Color		Gradients	
		Global	Local	Global	Local
SE	90	93		93	88
SP	89	84		78	76

In Sect. 4.2, we mentioned the research studies that employed this dataset [17, 19, 167, 166]. Among these, Barata et al. [17] was highlighted as the main benchmark for future comparison. In this work, the authors used locally and globally mapped color and gradient features. The locally mapped features were represented by the BoW approach and were classified by means of AdB, the SVM and k-NN. The benchmark results reported in [17] are listed in Table 4.11. Since the BoW representation was already explored by the authors, and also due to the fact that BoW did not outperform the RF ensemble on global features in Experiment #1, we chose to test only the ensemble performance in this experiment.

Here, we used the same individual and combined features as in Experiment #1. These features were globally mapped from the lesions' segmented areas and were represented in their original feature space. RF and weighted combination ensembles were used. The RF ensemble was trained with 100 unpruned trees, while the weighted ensemble was created based on the combination of the SVM, an LDA, and RF. We also employed performance weighting, which makes the weight of each base learner proportional to its sensitivity performance on the validation set. We chose sensitivity instead of accuracy, because in cancer classification, SE has more importance than ACC.

The framework in this experiment was evaluated using LOOCV in line with the PH² benchmark [17]. In order to tackle the class imbalance problem, Barata. et al [17] suggested replicating melanoma features in the training set using additional Gaussian noise ($N(\mu, \sigma) = N(0, 0.0001)$). Our approach, similar to Experiment #2, was to use the data space and generate synthetic melanoma images by deforming the originals using the BD and RDGM methods. This technique yielded a nearly balanced dataset with 117 melanoma and 154 benign and dysplastic lesions. During the evaluation, if a melanoma sample was included in a test set, its two synthetic images were included as well. Besides, there was no need for the validation set for RF, but half of the training set was kept for the validation of the weighted combination ensemble. The results of this experiment are presented in Table 4.12.

Table 4.12: Classification results of the RF and weighted combination (RF, LDA, and SVM) ensemble approaches with the PH² dataset. The results are presented in terms of the SE and SP percentages.

Features	RF		{SVM,LDA,RF}	
	SE	SP	SE	SP
Individual Features				
T_1	62.39	82.47	54.78	75.97
T_2	72.65	88.39	64.10	82.47
T_3	82.05	85.71	63.38	77.92
T_4	88.03	90.91	62.07	85.71
C_1	88.89	94.16	59.83	93.51
C_2	89.74	90.91	75.21	90.26
S	76.92	90.26	62.39	86.36
Combined Features				
C_1, C_2	93.16	94.16	76.07	94.16
C_1, C_2, S	90.06	94.81	81.74	92.21
T_1, C_2	94.02	92.21	70.94	90.91
T_1, C_1, C_2	93.16	93.51	75.21	94.81
T_3, C_1, C_2	89.74	95.45	86.32	94.81
T_4, C_1, C_2	89.74	92.86	88.03	92.21
T_1, T_2, T_3, T_4	90.06	92.21	86.32	92.16
T_1, T_3, C_1, C_2	92.31	93.51	67.54	93.51
$T_1, T_2, T_3, T_4, C_1, C_2$	91.45	94.16	89.74	93.51
$T_1, T_2, T_3, T_4, C_1, C_2$	91.45	94.16	92.31	92.86
$T_1, T_2, T_3, T_4, C_1, C_2, S$	91.45	93.51	88.89	94.16

Discussion

The comparison between the current experiment's results and the benchmark reveals the potential of the proposed method. In Table 4.12, the figures showing a better outcome than the benchmark are highlighted in bold. The highest performance of the benchmark, as shown in Table 4.11, was achieved by the BoW representation of locally mapped color features with SE and SP of 93% and 84%, respectively. The global mapping of the gradient features showed the second best SE and SP of 93% and 78%, respectively. In this experiment, the combination of colors (C_1, C_2), CLBP and opponent color angle (T_1, C_2), and the CLBP and both color features (T_1, C_1, C_2) outperformed the benchmark. Among these, (T_1, C_2) achieved the best SE and SP of 94.02% and 92.21%, respectively. As for the ensemble performance, RF proved better than the weighted combination of RF, the LDA and the SVM. This could be due to the fact that with a validation set, the base learners of the weighted combination ensemble are trained on half of the training data, which may have weakened its performance.

Conclusion

We proposed an automated classification framework based on a global feature extraction and ensemble learning. The framework was designed for the differentiation of melanoma from benign and dysplastic lesions. The evaluation results from the PH² dataset demonstrated that ensembles such as RF outperform single learner methods. Using the RF ensemble on globally mapped color and texture features, our framework achieved the highest SE and SP of 94% and 92%, respectively. In general, the results confirm the role and importance of color features for the detection and discrimination of melanoma lesions. However, the significance of texture features omitted in previous research works cannot be denied.

4.5.4 Experiment #4

Melanoma vs. Benign and Dysplastic nevi, Study of balancing techniques on PH² dataset

This experiment studies the effect of balancing strategies on melanoma classification. In order to make future comparisons possible, we chose to work with the PH² dataset. The experiment was conducted on a subset of the original dataset with an imbalance ratio of 1:3, which was chosen deliberately to comply with the requirements of the DOS method. The DOS method synthetically generates two additional melanoma images by deforming the original sample. Subsequently, the subset was composed of 39 melanoma and 117 benign and dysplastic lesions, which were randomly selected.

Based on the results of Experiment #3, we considered two color features (color statistics (C_1) and the opponent color angle space (C_2)) and two texture features (CLBP (T_1) and Gabor filters (T_3)). Using these individual features and their combinations, the 9 following feature sets were compared; $\{C_1; C_2; C_{1,2}; T_1; T_3; T_{1,3}; \{T_1, C_{1,2}\}; \{T_2, C_{1,2}\}; \{T_{1,2}, C_{1,2}\}\}$. These features were globally extracted from the segmented part of each sample and were represented in their original space. Twelve different balancing techniques, DOS, ROS, SMOTE, RUS, TL, Clustering (CUS), NM1, NM2, NM3, NCR, SMOTE+ENN, and SMOTE+TL, were used to create a balanced training set. We used the RF classifier with 100 unpruned trees and the original feature dimension of size $M = \{144, 84, 228, 26, 48, 74, 254, 276, 302\}$.

In order to validate the classifier with a stratified sampling, we used a 10-fold cross-validation scheme. However, contrary to the common 10-fold cross-validation, at each iteration, eight folds were kept for training and two for testing. In order to find the best performance, similar to [19], we used a cost function, which provides a trade-off between the SE and SP. It is formulated as follows:

$$C = \frac{c_{10}(1 - SE) + c_{01}(1 - SP)}{c_{10} + c_{01}}, \quad (4.16)$$

where c_{10} and c_{01} are the costs of incorrectly classifying melanoma and non-melanoma lesions, respectively. As it is more important to correctly classify melanoma lesions (i.e., high SE), c_{10} should be penalized more. In order to achieve a high SE without significantly reducing the value of SP, Barata et al. proposed setting $c_{10} = 1.5 \times c_{01}$ and $c_{01} = 1$ [19]. We considered the same configuration for our cost function.

The classification results obtained in this experiment are reported in Table 4.13, which can be divided into three main parts that show the results for imbalanced data (IB), balancing in the data space (DOS) and balancing in the feature space. These strategies are separated by double horizontal lines. The strategies applied in the feature space are subdivided into either Over-Sampling (OS), Under-Sampling (US) or a combination of OS followed by US (see horizontal dashed line in Table 4.13). In this table, based on the cost function defined above, the best results for each feature set are highlighted in the shaded cells. Table 4.14 shows the cost values for each configuration. Strategies with low cost functions are synonymous with a better SE and SP trade-off.

Discussion

The results of this experiment indicate that balancing techniques are essential for improving the classification performance. Thus, the US techniques and their combination with OS performed better than the OS techniques alone. In turn, the OS techniques applied in the data space outperformed those applied in the feature space. Due to the similar characteristics of melanoma and dysplastic lesions, their images are expected to have overlapping feature spaces. Subsequently, the misleading samples could be removed using US and thus ensure a better performance.

Specifically to our purpose, the NM2 algorithm with the combination of all the features ($T_{1,3}, C_{1,2}$) with the lowest cost values, achieved the highest SE and SP of 91.25% and 81.67%, respectively. Using the same feature combination, RUS achieved the second lowest cost with a SE and SP of 92.50% and 78.33%, respectively. The NM2 algorithm also achieved the third lowest cost with the combination of the Gabor and the color features ($T_3, C_{1,2}$) resulting in a SE and SP of 92.5% and 77.50%, respectively.

Among the color descriptors, the opponent color angle and hue histogram feature descriptor, C_2 , achieved better results than the widely used color statistics, C_1 . In the texture domain, Gabor descriptor, T_3 , outperformed CLBP features, T_1 .

Conclusion

In this study, we analyzed the impact of data balancing techniques on the classification of malignant melanoma. We presented an extensive comparison of twelve OS and US techniques

Table 4.13: Results obtained with different balancing techniques for color and texture features using the RF classifier. The first and second highest results for each feature set are highlighted in gray colors.

Features	Color						Texture					
	C ₁		C ₂		C _{1,2}		T ₁		T ₃		T _{1,3}	
	SE	SP	SE	SP	SE	SP	SE	SP	SE	SP	SE	SP
Balancing techniques												
IB	52.50	89.58	75.00	88.75	71.25	87.50	38.75	91.67	60.00	96.25	66.25	93.75
DOS	93.75	66.67	80.00	86.25	82.50	87.08	43.75	83.75	72.50	90.00	70.00	91.67
ROS	55.00	80.83	80.00	84.17	72.50	85.42	42.50	82.08	60.00	89.17	66.25	87.92
SMOTE	60.00	82.50	78.75	84.58	75.00	70.00	56.25	74.17	61.25	87.50	84.17	87.08
RUS	72.50	72.92	86.25	80.00	78.75	80.00	67.50	53.33	76.25	76.25	85.00	78.75
TL	51.25	86.25	76.25	87.92	67.50	88.33	37.50	87.92	65.00	90.42	68.75	91.67
CUS	81.25	67.92	80.00	84.58	86.25	80.42	56.25	65.83	70.00	77.50	85.00	77.08
NM1	67.50	72.08	86.25	79.17	85.00	82.50	72.50	43.75	80.00	62.50	87.50	66.67
NM2	70.00	72.92	86.25	81.25	85.00	82.92	76.25	48.75	86.25	40.83	86.25	51.25
NM3	82.50	75.00	87.50	80.83	85.00	80.42	73.75	55.83	72.50	82.50	82.50	80.42
NCR	66.25	76.67	87.50	81.25	85.00	82.08	67.50	67.92	75.00	85.83	82.50	83.33
SMOTE + ENN	76.25	73.33	85.00	81.25	85.00	82.08	81.25	56.25	76.25	82.08	80.00	79.58
SMOTE + TL	75.00	73.75	83.75	82.50	87.50	80.83	72.50	59.17	77.50	82.08	78.75	78.75

Features	Combined					
	T _{1, C_{1,2}}		T _{3, C_{1,2}}		T _{1,3, C_{1,2}}	
	SE	SP	SE	SP	SE	SP
Balancing techniques						
IB	73.75	89.58	71.25	89.58	71.25	92.50
OS	77.50	87.08	81.25	88.33	78.75	88.33
ROS	75.00	85.42	73.75	86.25	73.75	85.83
SMOTE	78.75	85.00	73.75	84.58	73.75	85.00
RUS	91.25	75.00	85.00	78.75	92.50	78.33
TL	73.75	88.75	63.75	90.00	72.50	91.25
CUS	83.75	81.25	80.00	84.17	83.75	82.92
NM1	85.00	82.08	86.25	80.42	87.50	80.83
NM2	87.50	82.08	92.50	77.50	91.25	81.67
NM3	83.75	81.25	85.00	80.00	86.25	80.42
NCR	86.25	81.67	82.50	85.00	83.75	85.42
SMOTE + ENN	86.25	81.25	83.75	82.50	78.75	82.92
SMOTE + TL	85.00	82.08	77.50	82.92	88.75	82.50

Table 4.14: The classification cost, C , for different balancing techniques and feature sets.

Balancing techniques	Classification cost, C								
	C_1	C_2	$C_{1,2}$	T_1	T_3	$T_{1,3}$	$T_1, C_{1,2}$	$T_3, C_{1,2}$	$T_{1,3}, C_{1,2}$
IB	0.3267	0.1950	0.2225	0.4008	0.2550	0.2275	0.1992	0.2142	0.2025
DOS	0.1708	0.1750	0.1567	0.4025	0.2050	0.2133	0.1867	0.1592	0.1742
OS	0.3467	0.1833	0.2233	0.4167	0.2833	0.2508	0.2083	0.2125	0.2142
SMOTE	0.3100	0.1892	0.2133	0.3658	0.2825	0.2317	0.1875	0.2192	0.2175
RUS	0.2733	0.1625	0.2075	0.3817	0.2375	0.1750	0.1525	0.1750	0.1317
TL	0.3475	0.1908	0.2417	0.4233	0.2483	0.2208	0.2025	0.2575	0.2000
CUS	0.2408	0.1817	0.1608	0.3992	0.2700	0.1817	0.1725	0.1833	0.1658
NM1	0.3067	0.1658	0.1600	0.3900	0.2700	0.2083	0.1617	0.1608	0.1517
NM2	0.2883	0.1575	0.1583	0.3475	0.3192	0.2775	0.1467	0.1350	0.1258
NM3	0.2050	0.1517	0.1683	0.3342	0.2350	0.1833	0.1725	0.1700	0.1608
NCR	0.2958	0.1500	0.1617	0.3233	0.2067	0.1717	0.1558	0.1650	0.1558
SMOTE+ENN	0.2492	0.1650	0.1617	0.2875	0.2142	0.2017	0.1575	0.1675	0.1958
SMOTE+TL	0.2550	0.1675	0.1517	0.3283	0.2067	0.2125	0.1617	0.2033	0.1375

in both the feature and data space. The results obtained particularly highlight the advantage of balancing the training set over using the original data, particularly for the methods based on US (NM2, NCR) and a combination of OS and US in the feature space. Furthermore, OS in the data space outperformed the OS techniques in the feature space. Finally, this study showed once again that the combination of color and texture features outperforms any other feature combination.

4.5.5 Experiment #5

Malignant vs. Benign and Dysplastic nevi, Sparse coded features using PH² dataset

In this experiment, we evaluated the effects of sparse-coded features on melanoma classification. We proposed a more general framework that does not rely on image preprocessing and lesion segmentation. Subsequently, it consists only of the feature extraction, representation and classification steps. Figure 4.14 shows the proposed framework of this experiment.

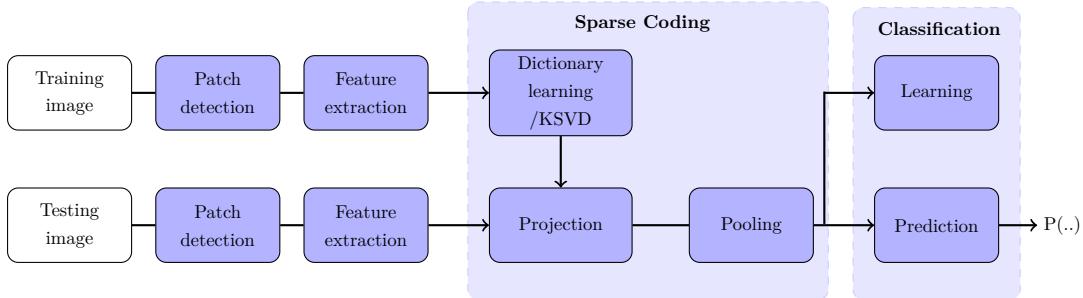


Figure 4.14: Proposed framework for an automated classification using sparse coded features.

Similar to experiments #3 and #4, we used a subset of the PH² dataset, containing 39 melanoma, 78 dysplastic and 76 benign lesions. Seven lesions were excluded due to artifacts in the images.

Low-level texture (SIFT) and color (hue and opponent color space angle histogram (C_2) and R, G, B intensity (C_3)) features were extracted from the patches in dermoscopic images via local mapping. The high-level descriptors are then computed using sparse-coded techniques. The main goal of sparse modeling is to efficiently represent the images as a linear combination of a few typical patterns, called atoms, selected from the dictionary (see Sect. 3.4). Here we intend to use a sparse representation of low-level features for melanoma classification.

Sparse coding consists of three main steps: (i) dictionary learning, (ii) low-level feature projection, and (iii) feature pooling. The dictionary (D) is learned using K-Singular Value Decomposition (SVD) which is a generalized version of k -means clustering that uses SVD. Once the dictionary is learned, each low-level feature extracted from a patch can be projected using D to form a set of sparse codes. This set is further max-pooled to build a final global descriptor to characterize the whole image. Finally, the descriptor obtained after max-pooling is used to train an RF classifier.

The high-level descriptors, is then computed using sparse coded techniques. The main goal of sparse modeling is to efficiently represent the images as linear combination of a few typical patterns, called atoms, selected from the dictionary (see Sect. 3.4).

In this experiment, we used a patch size of $10\text{ px} \times 10\text{ px}$ to extract the features from all the images. The class imbalance problem, similar to [17], was tackled by over-sampling the minority class. New samples were generated by randomly repeating original samples of the minority class with additional Gaussian noise $\mathcal{N}(0, 0.0001)$. The three low-level features were sparsely encoded considering three sparsity levels $\lambda = \{2, 4, 8\}$ and a different number of atoms $K = \{100, 200, \dots, 1000\}$. The classification was performed using a RF classifier with 1000 unpruned trees using the Gini criterion in a 10-fold cross-validation model. However, different

from the usual 10-fold cross-validation, 8 folds were kept for training (80%) and 2 for testing (20%) at each iteration. The results of this experiment are given in Table. 4.15 in terms of SE and SP.

Discussion

The highest classification rates with respect to each feature type and each sparsity level are highlighted in different shades of gray from dark to light. The results show that C_2 and SIFT sparse-coded features performed better with sparsity levels of 2 and 4, respectively, while C_3 was better with a sparsity level of 8.

In general, larger dictionary sizes led to better classification performance, independently of the feature type and the sparsity level. More precisely, dictionaries with more than 600 atoms are most suitable. Figure 4.15 illustrates the 12 best results. Although, the SIFT and C_2 sparse-coded features achieved a better classification performance in comparison with the C_3 , it must be noted that C_3 features represented the images in their simplest form and created comparable results.

Conclusion

In this experiment, we proposed a classification framework for melanoma lesions based on sparse representation of low-level features. This framework does not need the steps of image pre-processing or lesion segmentation and provides a more general algorithm to solve this problem. We proposed using the SIFT as a texture descriptor and the hue and opponent color space histogram as color descriptor (C_2). We also considered representing the images in their simplest form and considered the second color descriptor as R,G and B intensity values (C_3). An extensive comparison based on different dictionary sizes and several sparsity levels was carried out on the PH^2 dataset. The results highlighted the advantage of the proposed method, where the RF classifier and a sparse representation of SIFT features, with a dictionary size of 800 and a sparsity level of 2, achieved the highest performance in terms of SE and SP (100% and 90.3%, respectively).

4.6 Conclusion

In this chapter, we presented our automated recognition framework for the classification of melanoma lesions in dermoscopy images. The proposed framework consists of six main steps: image preprocessing, image mapping, feature extraction, feature representation, data balancing, and classification. For each of these steps, we made the following contributions:

Table 4.15: Results obtained with different numbers of atoms and sparsity levels. The first and second highest results for each sparsity level are highlighted in different shades of gray from dark to light color, respectively.

Features		C_2						C_3					
Sparsity level	Dictionary size	2		4		8		2		4		8	
		SE	SP	SE	SP	SE	SP	SE	SP	SE	SP	SE	SP
100	65.6	49.6	51.5	49.4	41.4	59.7	100	71.7	48.5	71.3	59.9	60	
200	50.1	59.3	52.7	53.5	51.5	50	57.2	64	95.8	86.6	71.4	72	
300	59.8	65.4	52.8	71	57.1	62.3	30	80	65.7	75.4	85.7	85.6	
400	67	78.6	62.6	81.3	69.9	76	38.5	66	85.8	77.3	78.6	91.4	
500	78.7	79	71.4	78.3	51.3	84	54.2	59.7	61.4	69.6	82.9	83.4	
600	98.6	82.5	68.7	89.6	64	89.9	48.7	78.4	50	64.6	91.4	89.3	
700	92.8	89.9	72.8	91.9	54.4	95.9	37.1	75.4	72.8	72	80	82.6	
800	92.9	81.4	100	88.4	78.5	89.7	40	70.9	58.6	80.1	97.2	83.9	
900	90	88	80	92	79.9	95.4	25.7	81.3	19.9	91.1	95.7	73.1	
1000	100	86.8	80	89.6	94.3	91.7	34.3	70.7	42.7	76.5	100	73.8	

Features		SIFT					
Sparsity level	Dictionary size	2		4		8	
		SE	SP	SE	SP	SE	SP
100	1.4	99.4	0	100	1.4	99.7	
200	65.6	47.5	34.3	78	20	95	
300	58.6	47.7	64.3	51.1	8.6	88.7	
400	62.8	74.6	59.9	64.7	71.3	58.2	
500	58.5	92	61.2	72.9	54	56.7	
600	85.8	86.6	61.4	73.3	51.3	53.4	
700	98.6	84.6	73	94.8	47	62.4	
800	100	90.3	97.1	93	48.5	72.7	
900	95.7	81.8	80	94.5	54.2	78.5	
1000	90	83.5	71.4	89.8	51.3	90.3	

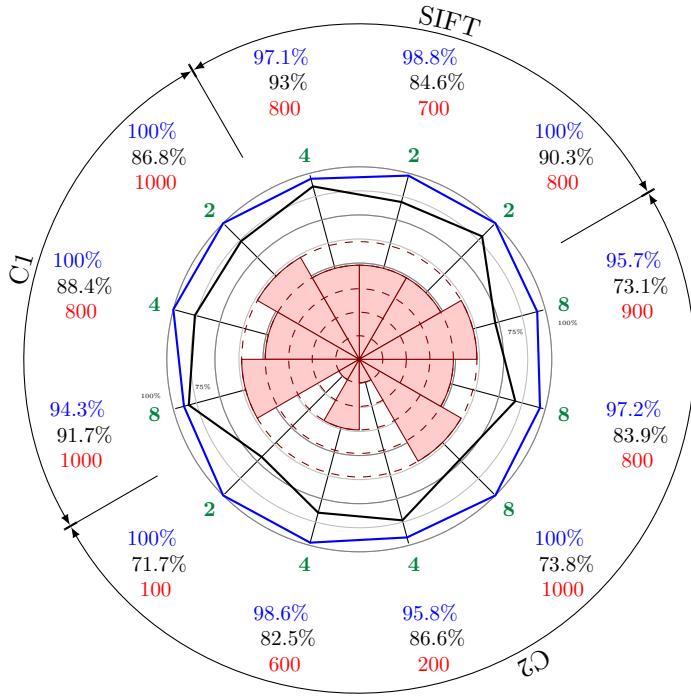


Figure 4.15: The 12 highest result achieved by RF classifier and sparse representation of SIFT, C_1 , and C_2 features with different **sparsity levels**. The results are illustrated in blue and black as sensitivity and specificity, respectively, while their associated dictionary size is represented in red. A comparison of the dictionary sizes is also presented in middle of the graph, which contains five levels with maximum dictionary sizes of 1000, 800, 600, 400, 200 for each level respectively.

- Proposed a hair removal algorithm based on morphological operations and exemplar-based inpainting.
- Developed a segmentation method based on a fusion of pdf-based, level-set, and FCM algorithms.
- Considered different texture and color descriptors beyond the clinical characteristics of the “ABCD” rule.
- Compared global and local feature extraction methods (mapping). The latter was implemented using a sliding window and dividing each sample into small patches.

- Applied the BoW and SCF techniques to reduce the complexity and dimensionality of the locally mapped features.
- Addressed the class imbalance problem by adapting the data space over-sampling approach to pigmented skin lesions and comparing it with the state-of-the-art over/under-sampling techniques in the feature space.
- Compared the performance of different learners and ensembles, such as SVM, GB, RF and a weighted combination.

The proposed framework was tested with respect to its individual steps in 5 different experiments. Experiments #1 and #2 were performed with the Vienna dataset, while the rest were carried out with the PH² dataset. The results obtained highlighted the superiority of the texture features such as Gabor filters, CLBP, HOG, and SIFT as compared to the state-of-the-art GLCM method. They also showed that the combination of color and texture features leads to a better performance. As to the global and local mapping results, we believe it to be highly dependant on the dataset, because in some cases, global feature extraction performs equal or superior to local extraction.

The importance of a balanced training, as well as the advantage of balance strategies in the feature space, was demonstrated in different experiments. In particular, it was shown that the US methods are more suited for melanoma recognition than OS. In terms of classifiers, the benefit and capability of an RF ensemble was shown throughout different experiments. In general, random forests is a less parameterized classifier, computationally less expensive, requiring minimum user interactions and obtaining the highest results. Finally, in Experiment #5, we tested the performance of Sparse Coded Features. This experiment showed that by using the simplest features (R, G, and B intensities), and without prior segmentation, it is possible to achieve acceptable results. The highest performance was achieved on the PH² dataset using SIFT features, a sparsity level of 2, and 800 atoms.

Our proposed framework is offline and requires learning stage. However, depending on the configuration setup, the time required varies. For instance, high-level representations using dictionary learning such as BoW or SCF require more time than low-level representation and, among the classifiers, the SVM requires more training time, since a grid-search method is used to optimize its parameters.

In line with our objective, the degree of robustness and generality of the proposed framework with respect to different datasets has been questioned. Although no test has been performed to confirm this aspect, we expect the use of ensembles such as AdB and RF, along with balancing techniques and pruning approaches [160] will be the solution to this problem.

4.6.1 Future Work

As an avenue for future research, we would like to explore several aspects of our framework in more detail, including:

- i An evaluation of the hair removal algorithm using annotated data.
- ii Extension and further studies of the segmentation algorithm.
- iii A comparison of sparse learned dictionaries with the Bag of Words model.
- iv A more detailed examination of balancing strategies.
- v A comparison of the tools developed with a larger dermoscopy dataset.
- vi An evaluation of the degree of robustness and generality of the proposed framework while dealing with data streams over time.
- vii Use of the recent techniques such as deep-learning for the classification.

Chapter 5

Automated Melanoma Classification with Polarimetric Images

While cross-polarized dermoscopes are used extensively by clinicians and dermatologists, other aspects of polarized imaging in the field of skin lesion analysis remains unexplored. The principles of polarization and state-of-the-art methods were discussed in Chap. 3. In this chapter, the methods proposed by the research community are divided into three categories: (i) partial Stokes polarimetry, (ii) full Stokes polarimetry, and (iii) Mueller polarimetry. As shown in [186], Mueller and full Stokes polarimetries are difficult to adjust for real-time and in-vivo imaging. Therefore, we propose a novel dermocope based on partial Stokes polarimetry, which was tested in the Melanoma Unit of the Hospital Clinic of Barcelona. Using the images obtained with our polarimetry device, their polarization properties were analyzed and an automated Computer-Aided Diagnosis (CAD) system for the differentiation of melanoma lesions was proposed.

The remainder of this chapter is organized as follows. A detailed description of the proposed polarimeter dermoscope is given in Sect. 5.1. Then, the quality and characteristics of the images acquired using this system are presented in Sect. 5.2. Our dataset and the proposed framework are presented in Sect. 5.3 and Sect. 5.4, respectively. Sect. 5.5 shows the experiments along with the results obtained, and finally, conclusions are drawn in Sect. 5.6.

5.1 Polarimeter Dermoscope

Figures 5.1 and 5.2 show the internal design of the polarimeter dermoscope and our first prototype, respectively. This dermoscope consists of several different parts: a camera, a lens, a focusing mechanism, a polarized state generator (PSG) and polarized state analyzer (PSA) units, and an electronic controller. This system provides configurable acquisitions allowing for the adjustment of different parameters, including the number of acquisitions and polarization angles in both the PSG and PSA units. The desired configuration can be set via a USB connector.

The system is designed to fit the distance sensor and the coupling mechanism in any of the Canon EOS models. Its focusing system placed right after the camera coupling mechanism, is manual. As shown in Fig.5.2a, the device has a switch and two buttons. The switch turns the device on and off while the buttons are for controlling the acquisition. The first button lights up the illumination system and allows visualization of the acquisition in-vivo using a liveview approach. The second button starts the acquisition process.

This polarimetric dermoscope has three connectors: (i) the connector for the signal exchange with the camera, (ii) the USB connector, and (iii) the power supply connector. The USB connector is used to set an automatic image acquisition program, while the power supply connector serves to charge the dermoscope for autonomous use. The optical system of the PSA unit consists of a dermoscope with a fixed macro lens, Schneider Componon-S 4.0/80, which has the magnifier with ratio of 1 (1:1). The lens is followed by a polarizer with a rotating system driven by a stepper motor programmable via the USB connector.

The PSG unit consists of a ring of white LEDs (model: Cree SMD white LED) as the illumination system (Fig. 5.3a). This system is followed by a diffuser for uniform illumination and a fixed polarizer. The fixed polarizer is positioned so that the first image is acquired with cross-polarized illumination (the polarizer of the PSA unit is positioned at 90° with respect to the horizontal axis of the fixed polarizer). For the image acquisition, the skin lesion should be in contact with the borosilicate window.

The dermoscope's focusing is done manually by rotating the ring placed just after the camera's coupling mechanism, as shown in Figs. 5.2 and 5.1b. Figure 5.3b shows the board for electronic control, while its position is shown in Fig. 5.1.

5.2 Polarimeter Images

As mentioned previously, the proposed polarimetric dermoscope is able to rotate the polarizer automatically in front of the camera at different angles. In order to obtain the first three Stokes parameters, the polarizer is positioned automatically at the three angles of 90° , 45° , and 0°

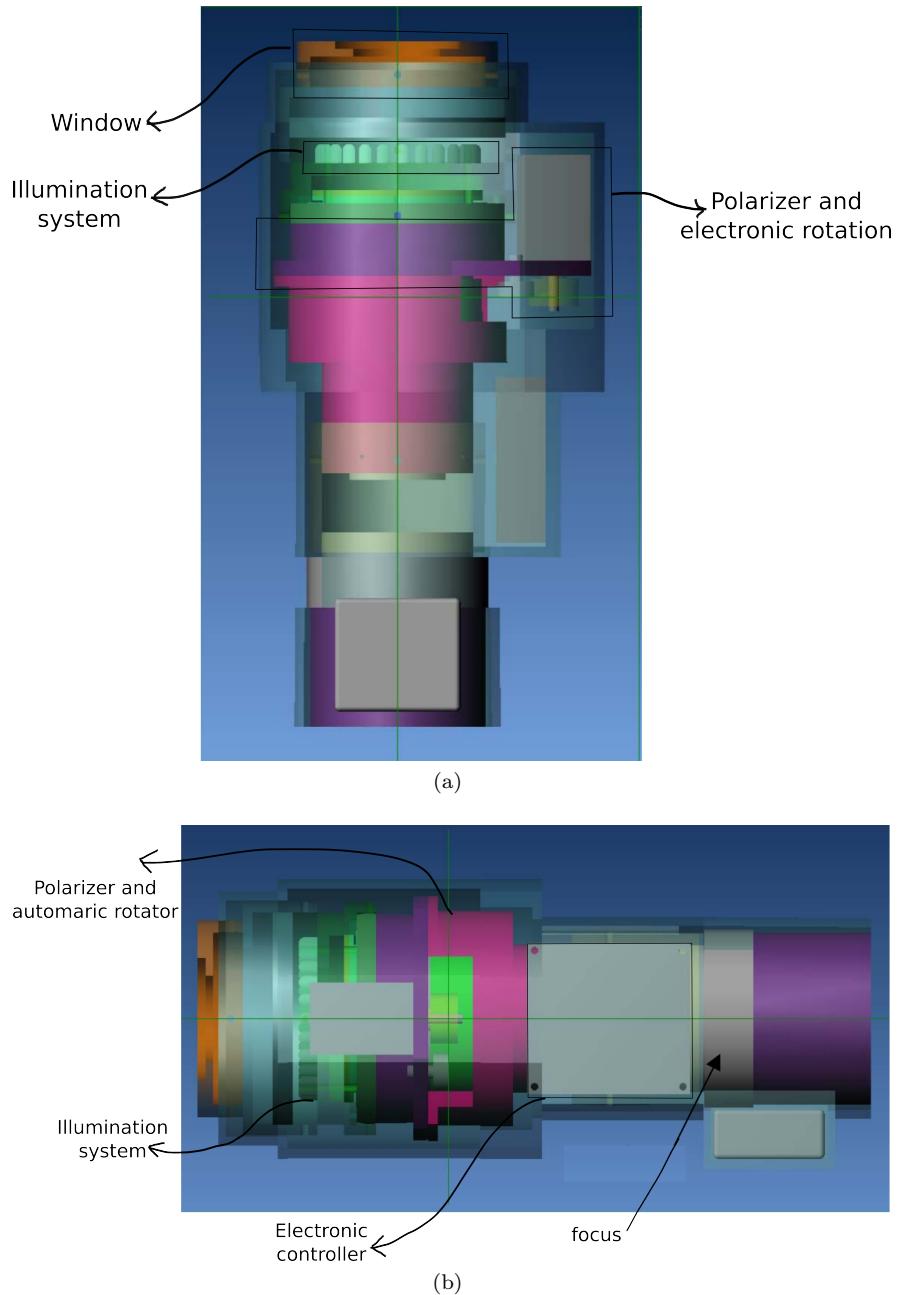
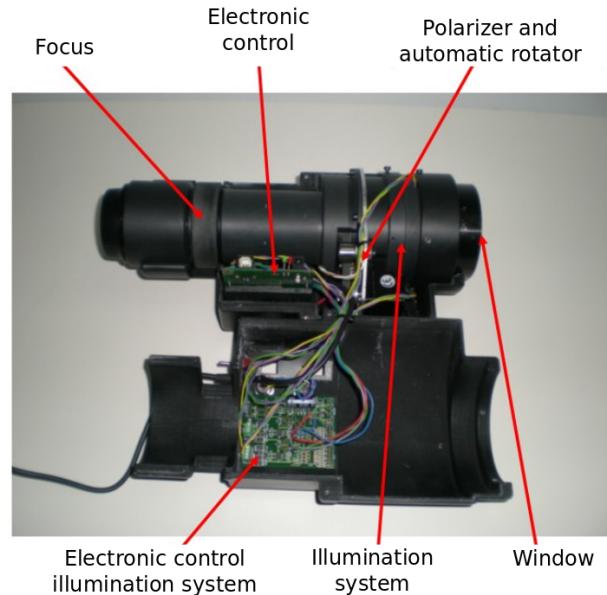


Figure 5.1: Internal design of the polarimeter dermoscope.



(a)



(b)

Figure 5.2: Polarimeter dermoscope first prototype.

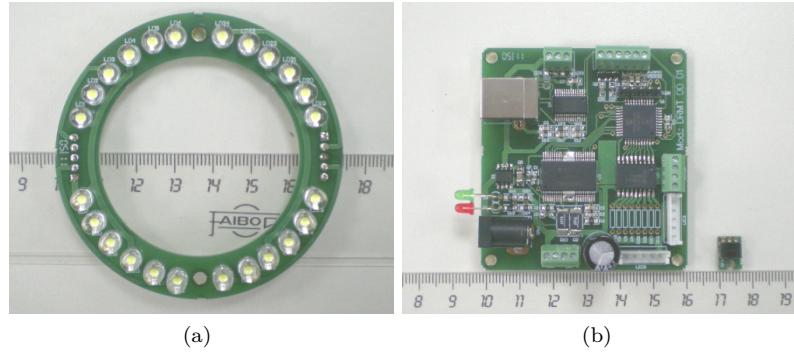


Figure 5.3: (a) Ring of white LEDs for the illumination system of the PSG unit. (b) Electronic board of the polarimeter dermoscope.

with respect to the horizontal axis of the polarizer in front of the illumination system (PSG). Using these angles, three images are acquired at each acquisition.

The detector used by the dermoscope is Canon EOS 40D, with a 10.1 Mpx complementary metal oxide semiconductor (CMOS) sensor and acquires 14-bit raw images of size 3908×2602 px ($1\text{px} \approx 36 \mu\text{m}^2$). The images are saved in Canon Raw-CR2 and JPEG formats. The left column in Fig. 5.4 shows images of one series saved in JPEG format.

In polarization calculations, accurate original measurements are essential, so working with the raw data instead of JPEG is required. Thus, the RAW data are analyzed with the Unidentified Flying Raw (UFRaw)¹ software. This software uses DCRaw (David Coffin's)² conversion utility to read different Raw formats. In this research, the UFRaw software is applied to convert Raw-CR2 formats to standard-RGB (sRGB) color images. The software uses camera white balance and a gamma correction of 0.45 without any further adjustment to convert Raw-CR2 to sRGB Tagged Image File Format (TIFF) images. The right column in Fig. 5.4 shows the converted versions of the Raw-CR2 images.

As shown in Fig. 5.4, the specular reflections are removed when the PSA unit's polarizer is positioned at 90° (I_{90° , cross-polarized image) and are evident when the polarizer is positioned at 0° (I_{0°). The three images acquired in each acquisition are used to compute the first three Stokes parameters (see Eq. 2.15). Prior to calculating the Stokes parameters, the three images are registered together. A detailed description of the registration step is given in Sect. 5.3. Due to the large size of the images, the Stokes parameters are calculated for a region of interest around each lesion.

¹<http://ufraw.sourceforge.net/>

²<https://www.cybercom.net/~dcoffin/draw/>

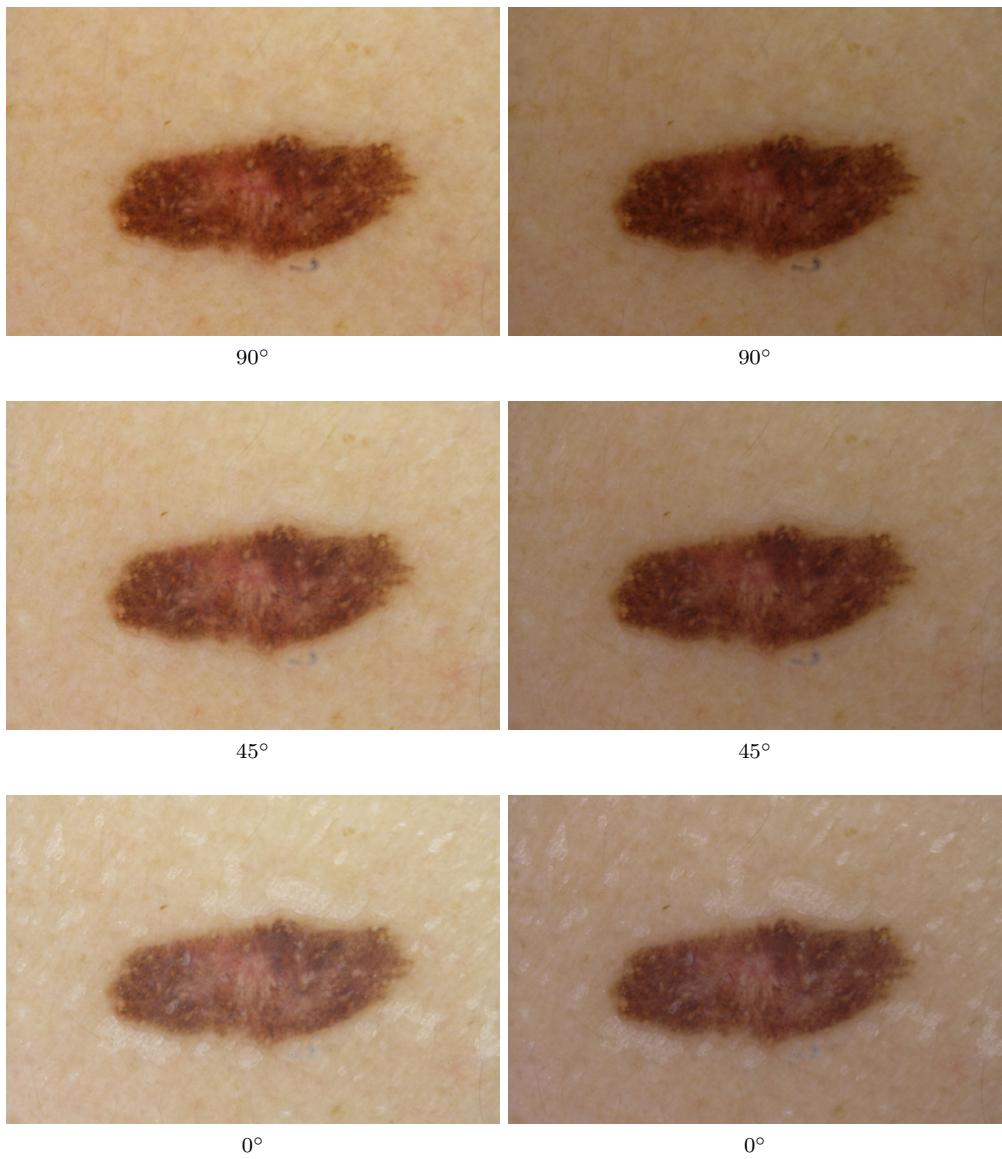


Figure 5.4: Example of a polarimetric acquisition. The left column shows images saved in JPEG format and the right column illustrates sRGB TIFF images obtained from the conversion of Raw-CR2 images with the UFRaw software. Three images are acquired while the polarizer in the PSA unit is positioned at 90°, 45°, and 0° with respect to the horizontal axis of the PSG polarizer.

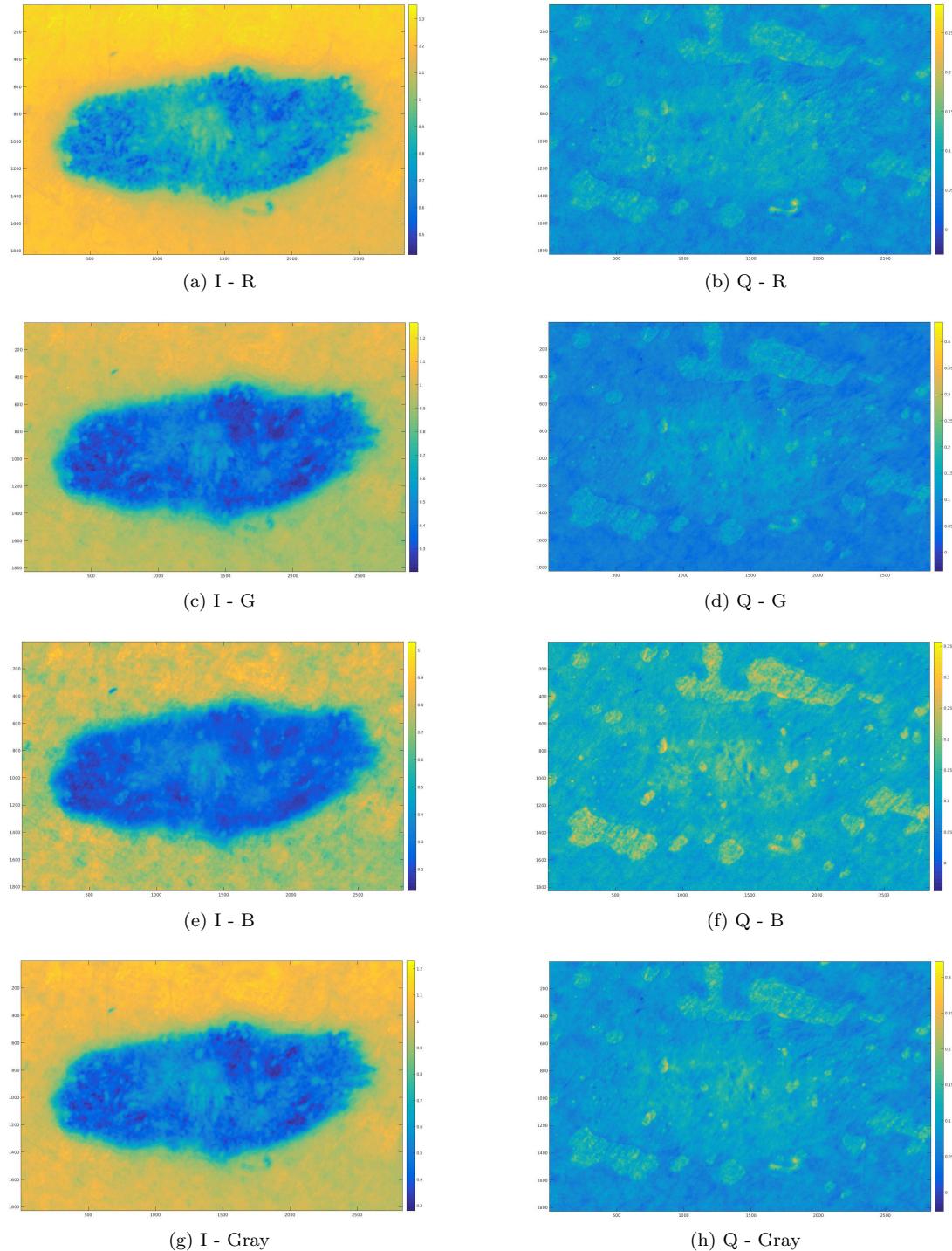


Figure 5.5: First and second Stokes parameters, I and Q in right and left columns, respectively, obtained from a polarimetric acquisition.

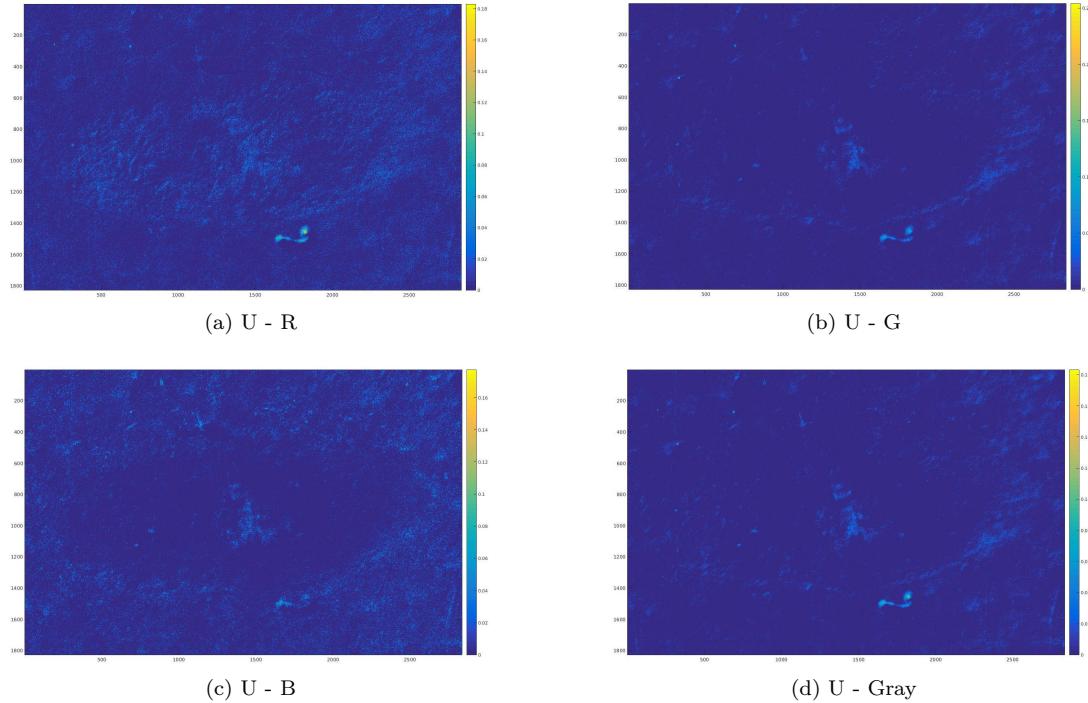


Figure 5.6: Third Stokes parameters (U) obtained from a polarimetric acquisition.

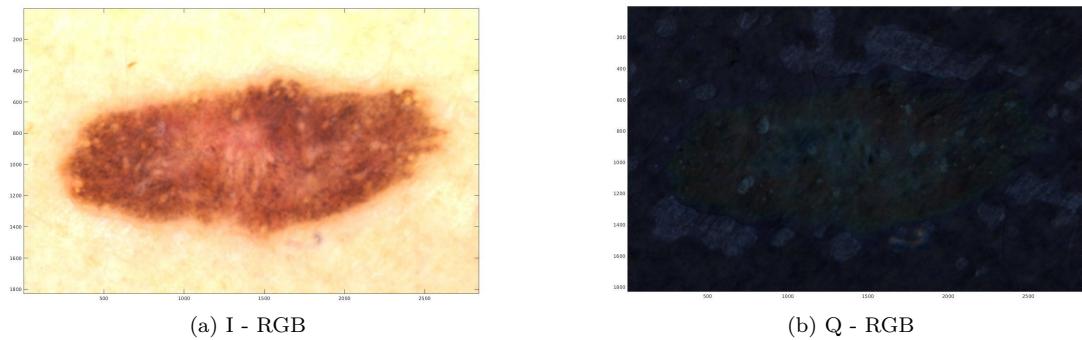


Figure 5.7: I and Q from Stokes parameters in RGB models.

The first Stokes parameter, $I = HI = I_{0^\circ} + I_{90^\circ}$, adds the scattered light to the camera from the epidermis and papillary dermis ($I_{0^\circ} = I_{\parallel}$) with the back-scattered light from the deeper layer of the dermis ($I_{90^\circ} = I_{\perp}$). Figure 5.5 shows this parameter in R, G, B and gray channel. The second Stokes parameter, $Q = HQ = I_{0^\circ} - I_{90^\circ}$, removes the light portion back-scattered from the deeper layer of dermis and only contains information from superficial layers (see Fig. 5.5, third and fourth rows). The third Stokes parameter, $U = 2I_{45^\circ} - I_{0^\circ} - I_{90^\circ}$, measures the difference between linearly 45° and -45° polarized back-scattered light (see Fig. 5.6). The perceived colors in the RGB representation of the first Stokes parameter, I , indicate that blue and green wavelengths are more absorbed by the melanin components in comparison with the red wavelength. The bluish color in RGB representation of the Q image is evidence of the same fact and is due to the difference of blue intensities between I_{0° and I_{90° , which is higher than the difference between red and green intensities (higher depolarization in blue channel). Figure 5.7 indicates these observations.

Using the Stokes parameters obtained, basic polarimetry properties, such as the degree of linear polarization (DOLP) and the angle of linear polarization (AOLP) can be obtained:

$$DOLP = \frac{(U^2 + Q^2)^{0.5}}{I}, \quad (5.1)$$

$$AOLP = 0.5 * \arctan2(U, Q). \quad (5.2)$$

Figure 5.8 and 5.9 illustrate the DOLP and AOLP in each color channel and grayscale. (the same lesion as in Fig. 5.4). As shown in the second column in Fig. 5.8 and 5.9, the AOLP has very low intensities (the angle is almost 0°). This is due to the construction and direction of the incident light. Since the incident light is illuminated orthogonally to the skin, the majority of the reflected light is redirected in the same direction and the difference of the angle is very low. The sample shown is a benign nevus, so for comparison, Figs. 5.10, 5.12 & 5.11 show the first three Stokes parameters and the two polarization properties in R,G,B channels and grayscale for a melanoma sample, respectively.

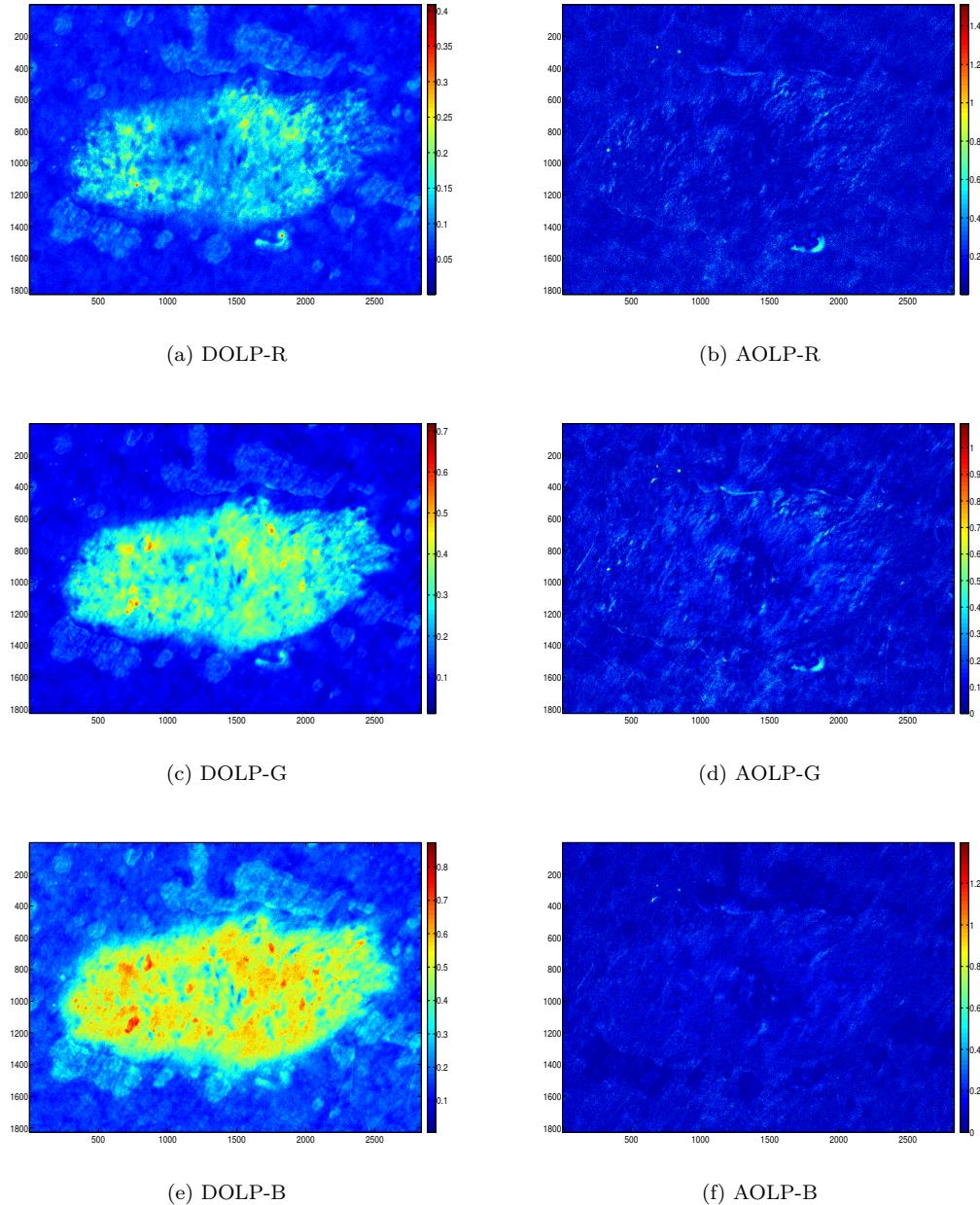


Figure 5.8: DOLP and AOLP obtained from I, Q, and U in the three color channels.

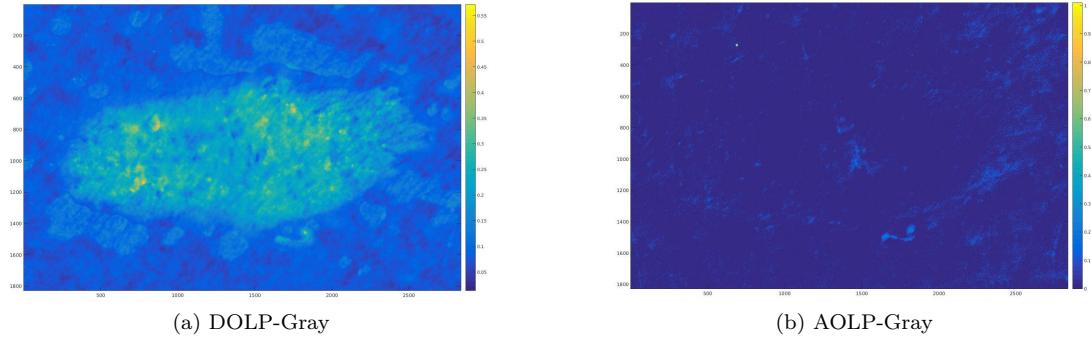


Figure 5.9: DOLP and AOLP obtained from mono-channel, grayscale of I_{90° , I_{45° , and I_{0° .

Table 5.1: Summary of the polarimetric dataset acquired with the prototype.

Lesion	Histopathology	# Samples
Melanoma	✓	22
Dysplastic	✓	19
Basal cell carcinoma	✓	3
Keratosis	✓	4
Benign	✓	8
	-	141

5.3 Dataset

This polarimetry dermoscope prototype has been tested in the Melanoma Unit of the Clinic Hospital of Barcelona during the past three years. In this time, series of polarimetry images of over 200 lesions were acquired. Due to focusing issues, some images had to be discarded, and, as a result, only 197 lesions were considered. In this set, 56 lesions were provided with a histopathology conclusion and 141 were assumed benign without histopathology. From the lesions for which a biopsy was performed, 22 were melanoma and 34 were classified either as dysplastic nevus, keratosis, basal cell carcinoma or benign lesions. Table 5.1 lists the total number of different lesion types.

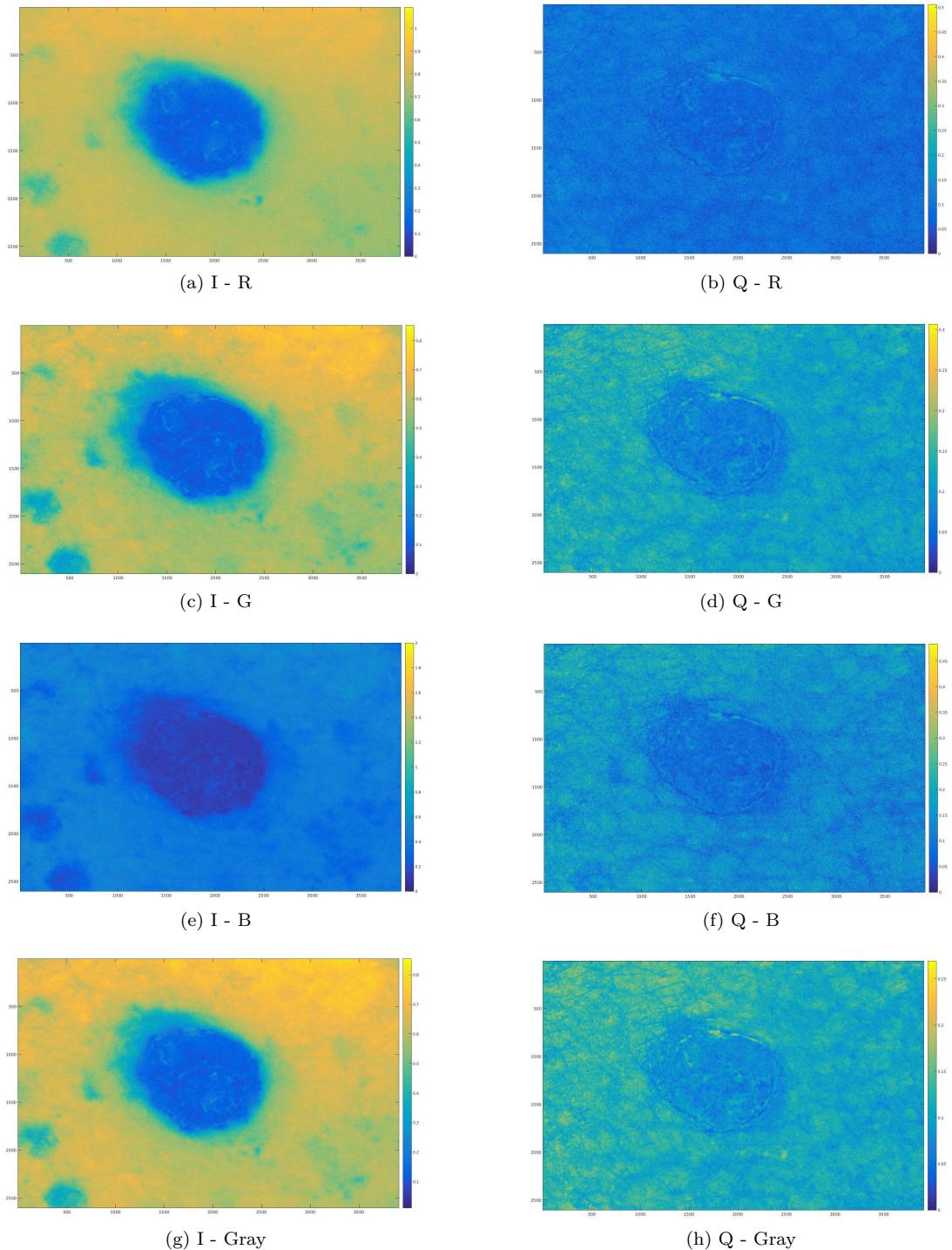


Figure 5.10: First and second Stokes parameters (I, Q) for a melanoma sample, right and left column respectively.

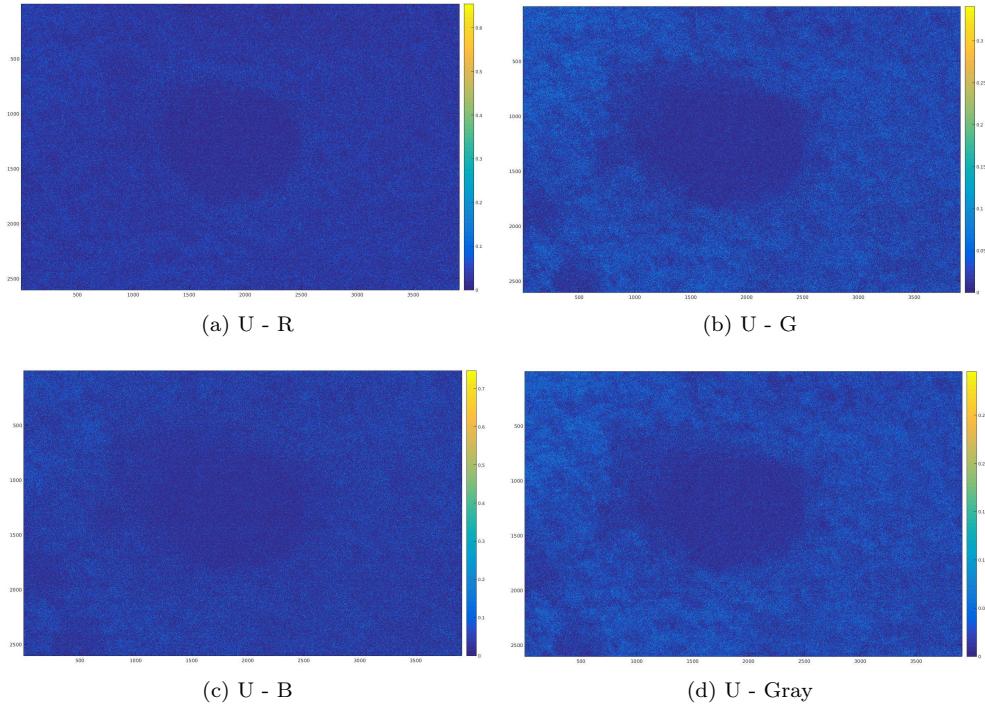


Figure 5.11: Third Stokes parameter (U) for a melanoma sample.

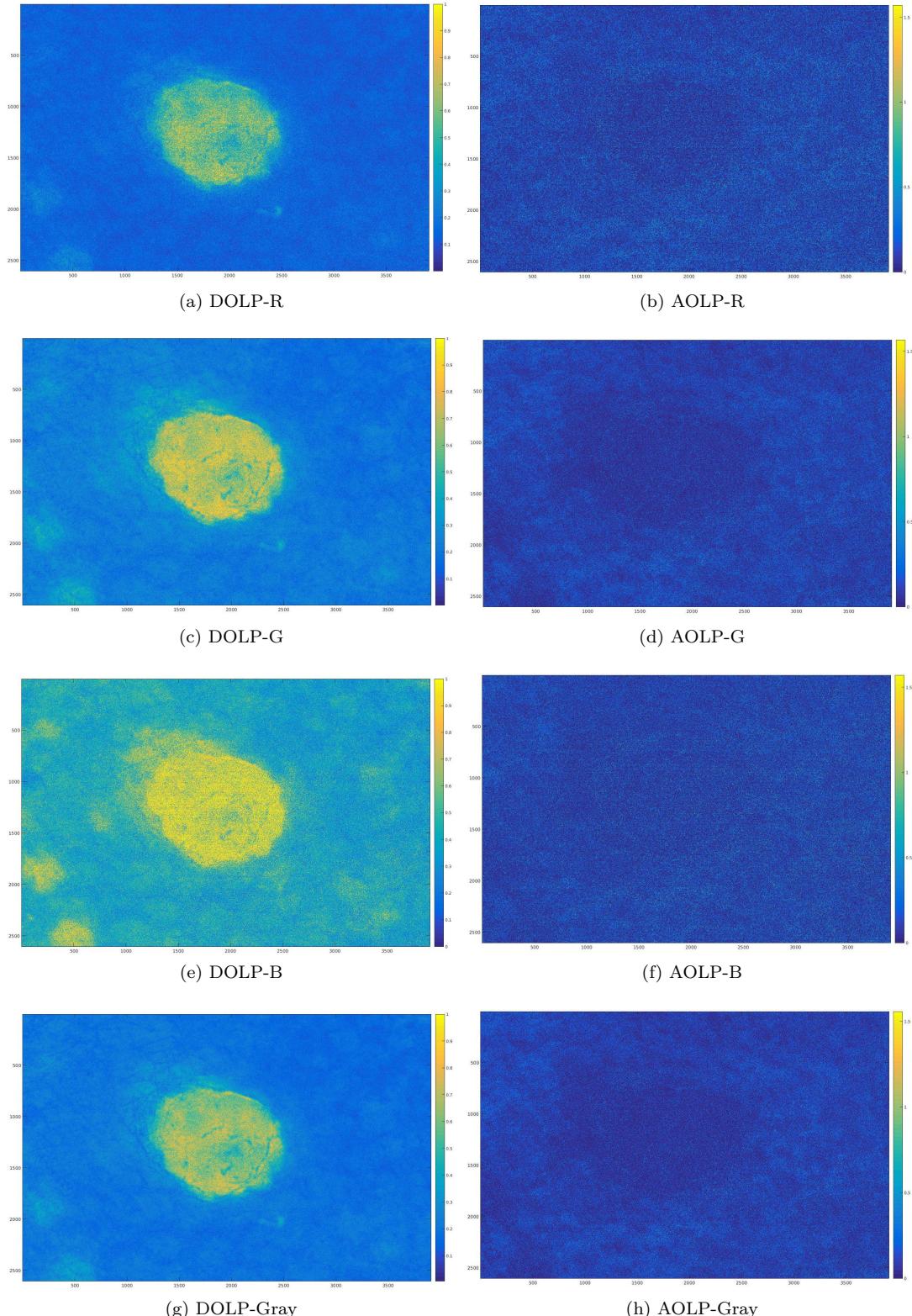


Figure 5.12: DOLP and AOLP obtained from I, Q, and U in the three color channels and grayscale.

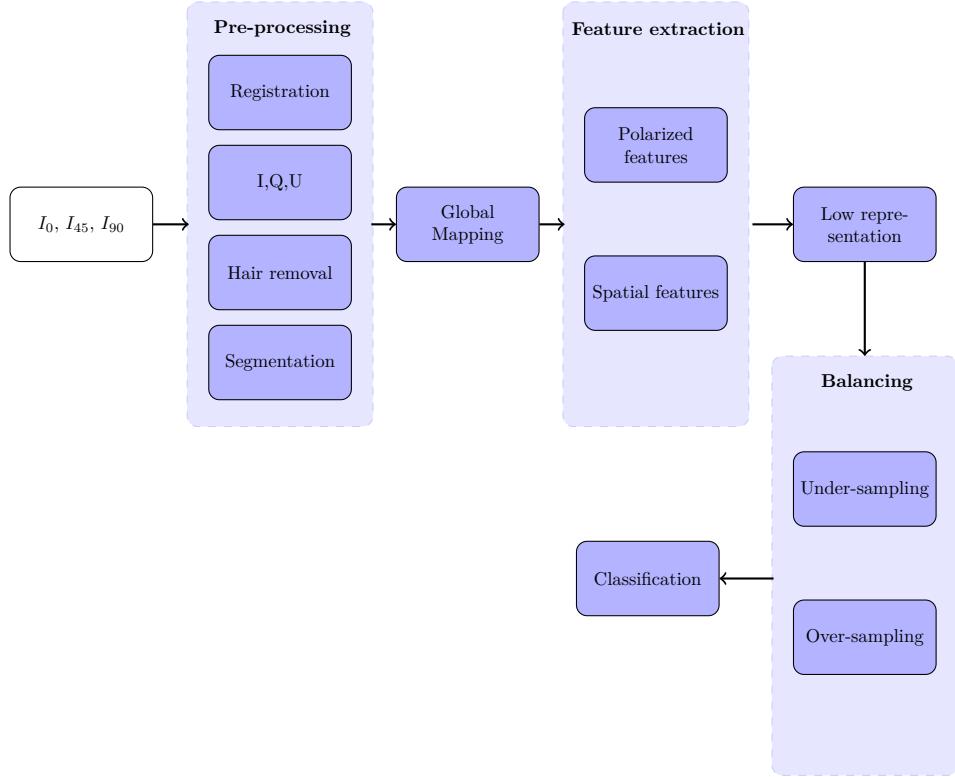


Figure 5.13: Proposed framework for automatic classification of melanoma lesions based on polarimetric images.

5.4 Methodology

Figure 5.13 shows our proposed framework for the automatic classification of melanoma lesions using polarized images. Similar to the framework for dermoscopy images (see Fig. 4.4), this framework is also composed of 6 main steps: pre-processing, mapping, feature extraction, representation, balancing and classification. These steps are explained in the following.

5.4.1 Pre-processing

For each lesion, three polarized images are acquired using the device proposed. As mentioned above, the polarizer in the PSA unit rotates automatically during the image acquisition, so the quality and accuracy of the images in terms of polarization angles is ensured. However, since the device is positioned and focused manually during the acquisition process, the images may

be out of focus and/or misaligned. When there is a focusing problem, the images are simply excluded from the dataset, whereas in the case of misalignment, the images must undergo a registration procedure. Then the registered images are processed to acquire the first Stokes parameters, I, Q, and U (see Sect. 5.2), and to remove any artifacts. Finally, the combination of the new triplet images and the acquired images is used for lesion segmentation. All these stages are explained below.

Registration

Image registration is a crucial step prior to further processing: for optimal polarimetric measurements, the intensities retrieved from all the images must correspond to the same spatial locations. Thus, the three acquired images should be perfectly aligned.

Image registration uses geometric transformations to align an unregistered image (moved image) to the reference image (fixed image). In this research, we consider the second acquired image I_{45° as the fixed image (reference image), so the first and third images are registered onto this reference. The registration is performed by means of a transformation model that relates the unregistered image to the reference frame. Rigid (affine) and non-rigid, or elastic, transformations are two common transformation models. The transformation model required here should preserve the shapes and borders of the lesion without deforming it. These criteria is required to preserve the intrinsic criteria of the lesions with the assumption that they have not been deformed in the acquisition process. In this regard, affine rigid transformation is opted since it preserves the ratios of distances and angles (thanks to its linearity).

Thus, our registration framework consists of four main steps: (i) feature detection, (ii) feature matching, (iii) mapping estimation, and (iv) image transformation [214].

Feature detection refers to the detection of distinct aspects of each image such as corners, closed boundaries, line intersections, and highly textured patches. Different detectors, such as Harris, Hessian, Scale-Invariant Feature Transform (SIFT), or Speeded Up Robust Features (SURF), are used for finding salient and distinct points in images. Once the feature locations are known, different descriptors can be used for their unique representation, e.g. SIFT or SURF [24].

Feature matching refers to establishing correspondences between features detected in the unregistered and reference images. Feature descriptors are usually matched based on similarity measurements such as correlation or distance ratio.

Mapping estimation provides a mapping function that transforms the unregistered image to the fixed image frame. The parameters of the mapping function (e.g., homography) are calculated based on the feature correspondences from the previous step [214].

Image transformation serves to actually transform the unregistered image to the reference frame based on the mapping function estimated in the previous step.

In this work, SURF was used for feature detection and description, correlation was employed as the similarity measure, and the mapping function (homography) was built as an affine transformation between corresponding feature points. Finally, using the estimated rigid (affine) transformation, the unregistered I_{90° and I_{0° were registered to the reference frame (I_{45°).

The principle of the SURF detector and descriptor is briefly explained in the following.

Speeded Up Robust Features (SURF) introduced by Bay et al. [24], is a fast and robust feature detector and descriptor. SURF differs from the SIFT mainly in their computation speed. This detector uses the Hessian matrix for selecting feature location and scale. Equation 5.3 illustrates the formulation of the Hessian matrix for one location (x, y) in a given image I , at a defined scale σ :

$$H(x, y, \sigma) = \begin{bmatrix} L_{xx}(x, y, \sigma) & L_{xy}(x, y, \sigma) \\ L_{xy}(x, y, \sigma) & L_{yy}(x, y, \sigma) \end{bmatrix}, \quad (5.3)$$

where L_{xx} , L_{xy} , L_{yy} are the convolutions of the second-order Gaussian derivatives of I at a given location (x, y) . The second-order Gaussian derivatives are approximated with box filters [24] (see Fig. 5.14b). The computational advantage of SURF is mainly due to the use of the *integral image* instead of the original image in the convolution process. The integral image (I_Σ) at a given location is the sum of all the pixels in a rectangular area at that location (see Eq. 5.4). Figure 5.14a shows how Eq. 5.4 can be efficiently computed using the integral image.

$$I_\Sigma(x, y) = \sum_{i=0}^{i \leq x} \sum_{j=0}^{j \leq y} I(i, j). \quad (5.4)$$

The SURF keypoint description process is divided into two steps [24]. First, a consistent orientation for each detected point is chosen based on the information in the circular region around. The orientation is selected based on the response of the Haar wavelet in this circular neighborhood. After weighting the Haar wavelet responses using a Gaussian kernel, the final orientation is calculated based on the sum of all the responses within a sliding orientation window of size $\pi/3$. In the second step, a square region with the selected orientation is considered around each keypoint. The defined square region is then divided into 16 sub-regions (4×4), from which the horizontal and vertical responses of the Haar wavelet are calculated (d_x and d_y , respectively) [24]. The sum and absolute sum of

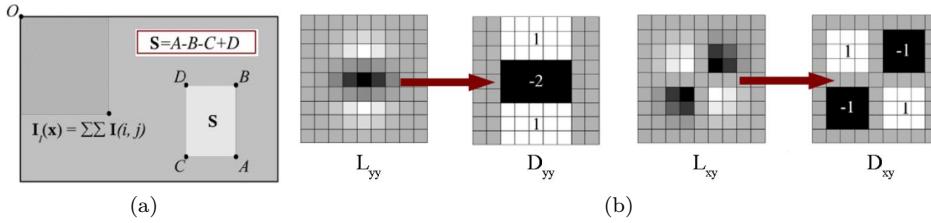


Figure 5.14: (a) Integral image. (b) Approximation of the second-order Gaussian derivatives (L_{yy} , L_{xy}) using box filters(D_{yy} , D_{xy}). The images are taken from [24].

the detected responses from each sub-region define the four-dimensional descriptor, which leads to the final descriptor of size 64.

Hair removal

The hair removal algorithm we proposed to use with dermoscopy images is applied here as well. The two steps of this algorithm: (i) hair detection using morphological operations and (ii) inpainting using an exemplar method are explained in Sect. 4.4.1. Using this algorithm, only the cross-polarized image, I_{90° , serves for the hair mask creation. And, assuming a good image registration, the same mask is used to inpaint the detected hairs in the other images. Figure 5.15 shows some examples where the hair removal algorithm was used.

Segmentation

Thanks to having several images of the same lesion with different contrasts, I_{0° , I_{45° , and I_{90° , and a polarimetric image DOLP, we can apply a rather simple segmentation algorithm: Otsu thresholding [144]. This algorithm considers that a given image contains two classes of pixels, foreground and background, and attempts to find the intensity threshold that separates the two classes. The optimum threshold is set so that the within-class variance is minimized. This method calculates all the possible threshold values and measures pixel characteristics (mean, variance and weights) on each side of the threshold. Then the within-class variance is calculated for each level, and the level with the minimum within-class variance is considered as the threshold. Equation 5.5 shows how this variance is calculated for two classes in one level:

$$\sigma_W^2 = W_b \sigma_b^2 + W_f \sigma_f^2, \quad (5.5)$$

where W_b and W_f are the weights of the background and foreground, respectively, and σ_b , σ_f represent the variance of each class.

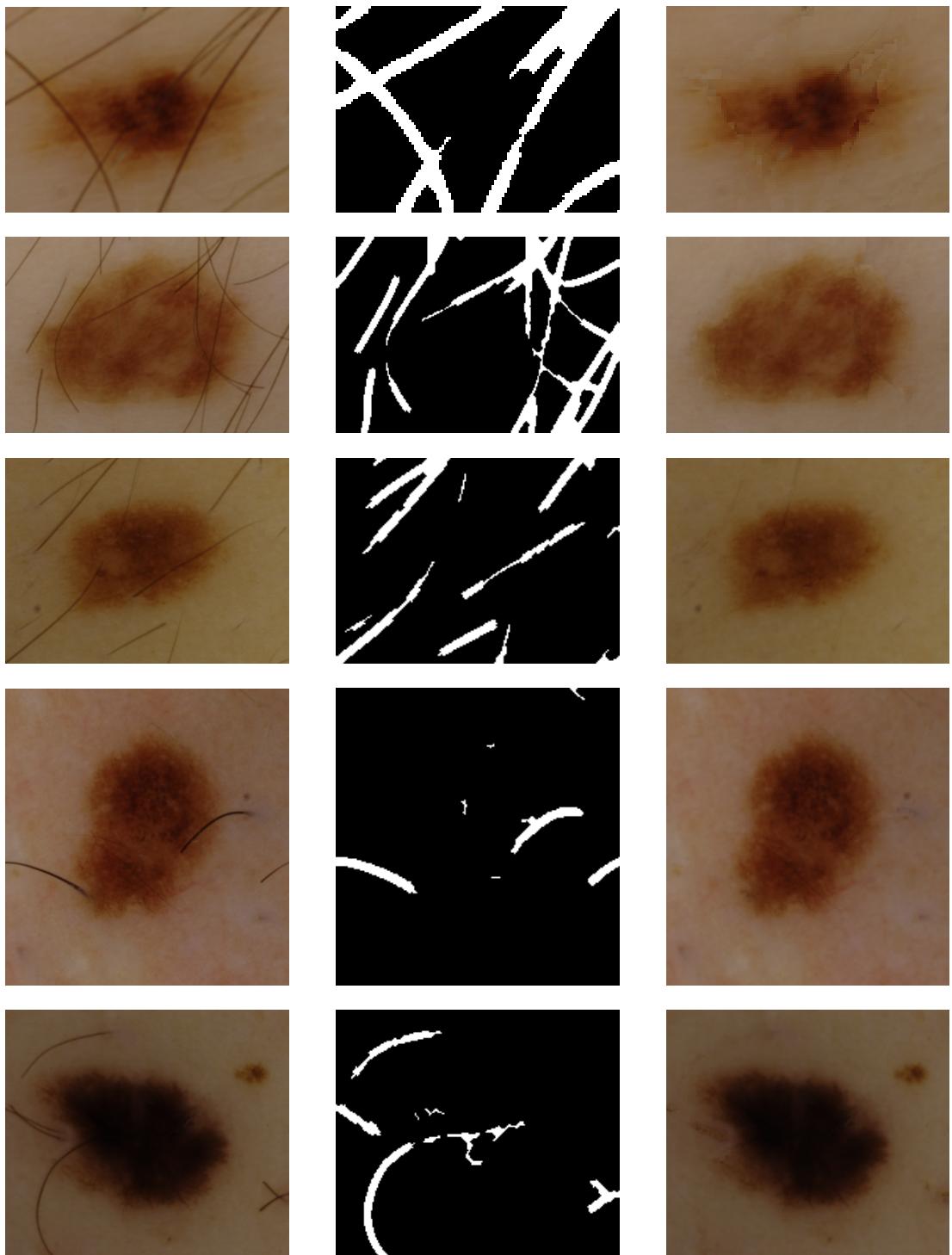


Figure 5.15: Hair removal algorithm used on several polarimetric images.

The computation of the within-class variance can be expensive. So instead, the between-class variance can be used since its maximization is equivalent to the minimization of the within-class variance [144]. This computation is shown in Eq. 5.6.

$$\sigma_B^2 = W_b W_f (\mu_b - \mu_f)^2, \quad (5.6)$$

where, μ_b and μ_f are the mean of the background and foreground classes, respectively.

Using the Otsu algorithm, the larger area in the segmented image is considered as the main region of interest. The four segmentations are then combined, and the final lesion mask contains only the areas present in at least three out of the four segmentations. In addition, the resulting segmentation is optimized using morphological dilation with a disk structuring element of size 2 px. Thus, the majority voting on the segmentations obtained from a series of images overcomes the drawbacks of a simple algorithm such as Otsu.

In our approach, three segmentation regions are considered for each lesion: internal, external and their combination (combined). The internal region is the output of the majority voting after the morphological step. To obtain the external region, the boundary of the internal segmentation is detected and expanded using the dilation operation with a square of size 50 pixels as the structuring element. Considering the internal and external segmented regions, the combined segmented area is simply a combination of the two aforementioned regions. Figure 5.16 shows some segmentation results obtained using this approach.

5.4.2 Mapping

Since finding the polarimetric properties of the entire lesion is the main goal of this research, we opted for global mapping (see Sect. 4.4.2) Therefore, the features are extracted from either the segmented lesion or the region of interest bound to the lesion's edges.

5.4.3 Feature extraction

Due to the nature of the images, two sets of features were extracted: polarimetric and spatial (dermoscopy) features. Table 5.2 illustrates a list of descriptors for each category. The spatial features are those that can be obtained from color dermoscopy images, so they were only extracted from the cross-polarized image (I_{90}). With reference to our previous experiments and the framework for dermoscopy images, two color and four texture features were considered: color statistics (C_1), opponent angle and hue histogram (C_2), Completed Local Binary Pattern (CLBP) (T_1), Grey-Level Co-Occurrence Matrix (GLCM) (T_2), Gabor filters (T_3), and Histogram of Oriented Gradients (HOG) (T_4). These features are extracted from the region of interest bound to the edges of the entire segmented lesion (similar to the global features in the

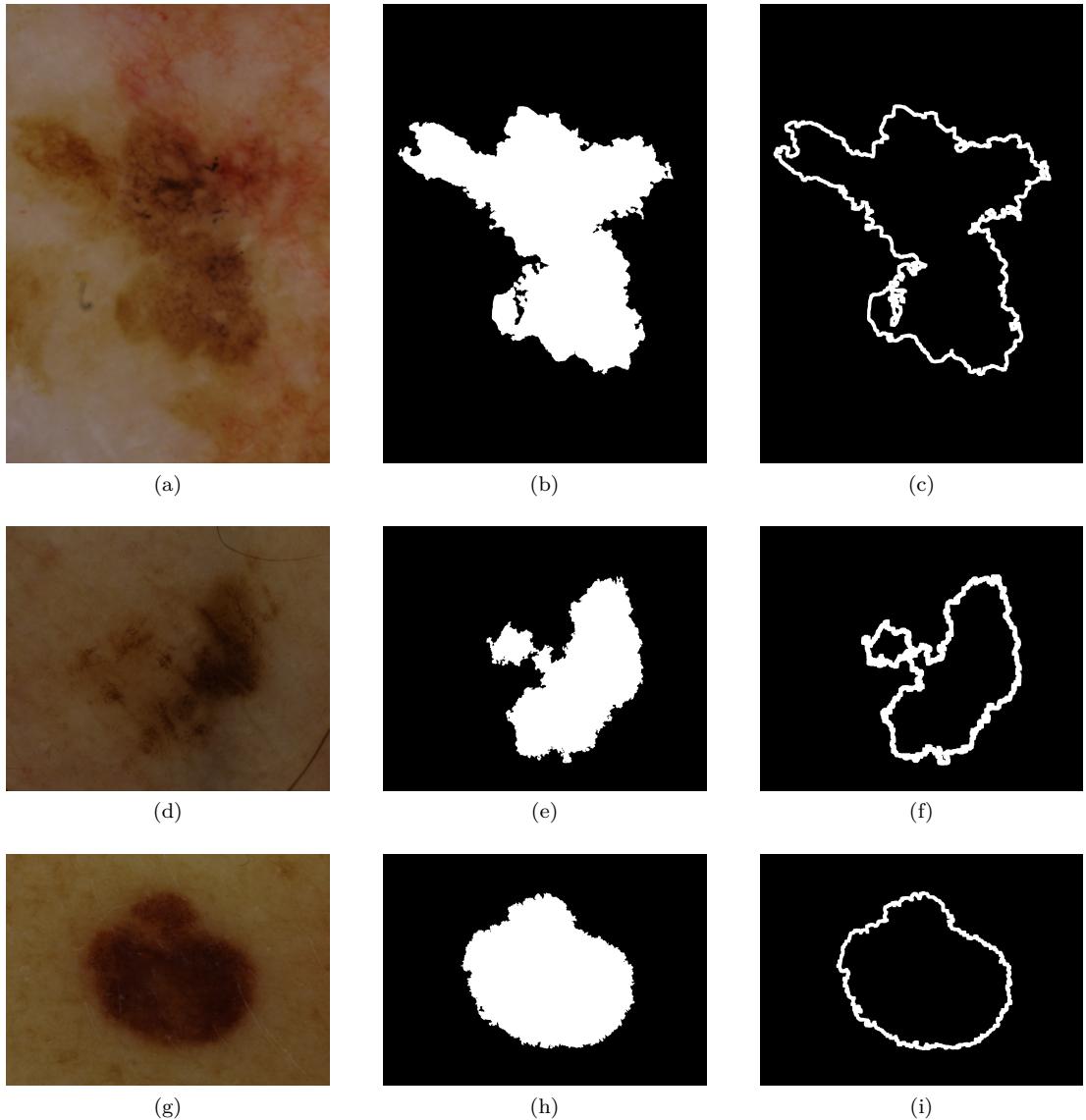


Figure 5.16: Examples of segmentations on polarimetric images of pigmented skin lesions. The leftmost column shows the cross-polarized image (I_{90°) acquired with our device. The middle and rightmost columns show the internal and external segmentation regions, respectively.

dermoscopy framework). For the sake of comparison, we used the same extraction parameters. See Chapter 4, Sect. 4.4.3 for a detailed explanation of these features.

Polarimetric features, as the name suggests, are extracted from polarimetric images. In the current stage of our algorithm, the polarimetric features are extracted from mono-channel inputs, grayscale of I_{90° , I_{45° , and I_{0° . After primary observations, we considered using the DOLP, polarization intensity (Pol_{int}), and the third Stokes parameter U . The last two parameters are intensity independent, therefore prior to feature extraction the obtained images are normalized. Equation 5.9 shows the formulation of these parameters once again:

$$DOLP = \frac{\sqrt{U^2 + Q^2}}{I}, \quad (5.7)$$

$$\text{Pol}_{int} = \sqrt{U^2 + Q^2}, \quad (5.8)$$

$$U = 2I_{45^\circ} - I_{0^\circ} - I_{90^\circ}. \quad (5.9)$$

Figure 5.17 shows the images of these parameters for the two melanoma samples. The polarimetric features are extracted from the segmented area of the lesions. As explained above, each lesion is segmented into external and internal areas, and therefore, the polarimetric features are extracted for internal, external and combined regions. The features for external and internal regions are annotated with the extension $_E$ and $_I$, respectively.

5.4.4 Feature representation

The primary objective of the proposed classification framework is to evaluate the polarimetric features in comparison with their spatial counterparts. Thus, we considered the simplest forms of feature representation. The DOLP, Pol_{int} , and U features were characterized by (i) histograms and (ii) statistical moments (mean, standard deviation, skewness, and kurtosis) of the polarimetric feature distribution. The representations based on histograms and moments are denoted in Table 5.2 with the extensions H and M , respectively. The U feature was also represented by the magnitude of gradients (U_{Gmag}). The gradient of the gray image is obtained by convolving the image with the Sobel mask in the horizontal and vertical directions. By doing so, a gradient vector for each point is obtained ($[G_x \ G_y]^T$), from which the gradient magnitude (G) and phase (θ) are calculated using the following formulation (here, only the gradient magnitude is used):

$$\|G(x)\| = \sqrt{G_x^2(x) + G_y^2(x)}, \quad (5.10)$$

$$\theta(x) = \text{atan2}(G_y(x), G_x(x)). \quad (5.11)$$

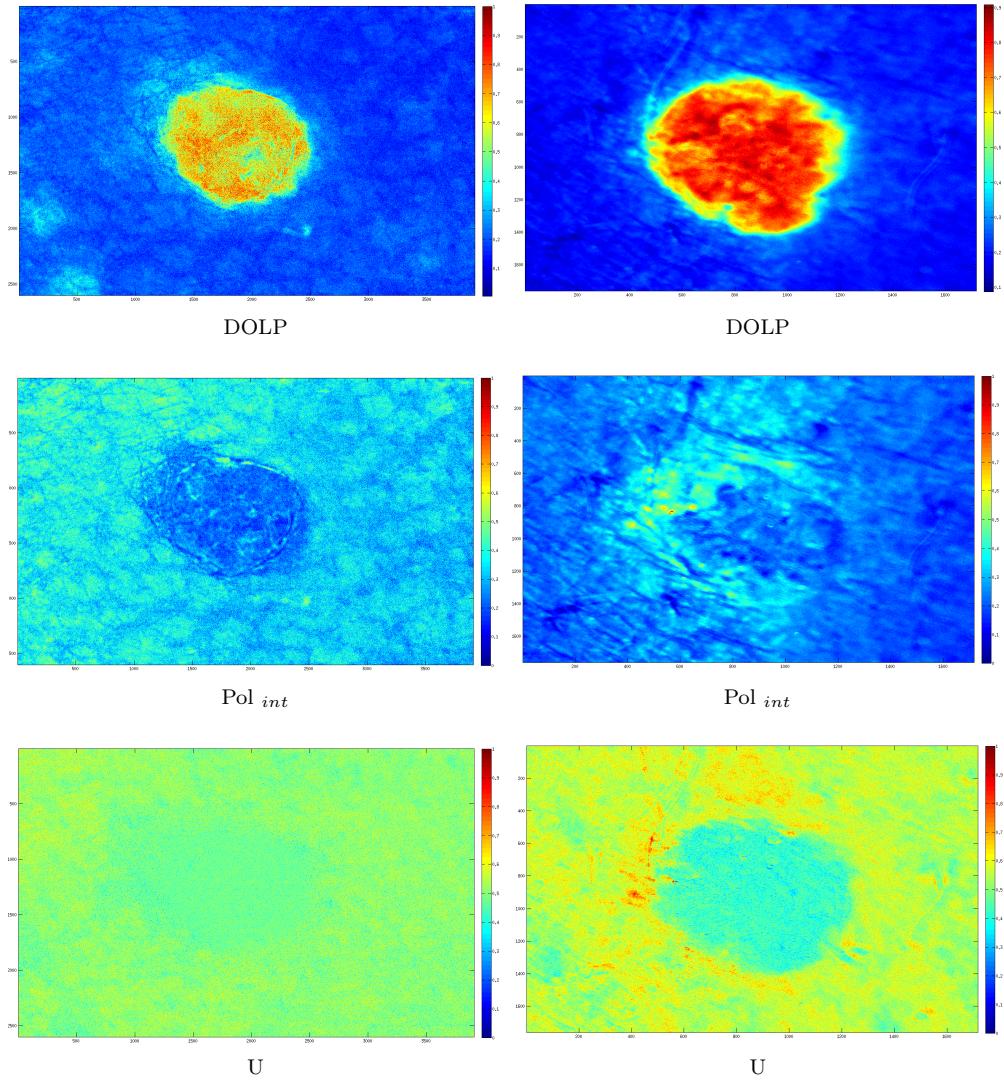


Figure 5.17: Images of polarimetric features for two melanoma lesions: DOLP, Pol *int*, and U.

Table 5.2: Summary of the features used in the automatic classification of melanoma from polarimetric images.

Spatial Features	
Feature	Index
Color	
Color mean and variance along RGB, HSV and LAB	C_1
Color histogram in RGB ($b = 42$) ¹	
Opponent color space angle and hue histogram ($b = 42$)	C_2
Texture	
Completed Local Binary Pattern ²	T_1
Gray-Level Co-occurrence Matrix ($\theta = \{0, \pi/4, \pi/2, 3\pi/4\}$, $D = 9$ pxls, $G = 32$) ³	T_2
Gabor Filter ($s = 4$, $\theta = \{\pi/6, \pi/3, \pi/2, 2\pi/3, 5\pi/6, \pi\}$)	T_3
Histogram of Oriented Gradients	T_4
Polarimetric Features	
Feature	Index
DOLP	
Histogram of internal segmentation	D_{H_I}
Histogram of external segmentation	D_{H_E}
Histogram of segmented lesion	D_H
Moments of internal segmentation	D_{M_I}
Moments of external segmentation	D_{M_E}
Moments of segmented lesion	D_M
Pol $_{int}$	
Histogram of internal segmentation	P_{H_I}
Histogram of external segmentation	P_{H_E}
Histogram of segmented lesion	P_H
Moments of internal segmentation	P_{M_I}
Moments of external segmentation	P_{M_E}
Moments of segmented lesion	P_M
U	
Histogram of internal segmentation	U_{H_I}
Histogram of external segmentation	U_{H_E}
Histogram of segmented lesion	U_H
Moments of internal segmentation	U_{M_I}
Moments of external segmentation	U_{M_E}
Moments of segmented lesion	U_M
Magnitude of gradients	U_{Gmag}

¹ b stands for the number of bins.

² 24 neighbourhood, rotation invariant, uniform and normalized histogram.

³ D stands for distance in pixels and G is the quantized number of gray levels.

5.4.5 Balancing

As illustrated in Table 5.1, our polarimetric dataset is imbalanced. In order to reduce the effect of the imbalance problem, prior to classification, over and under-sampling techniques in the feature space are applied. The same techniques, previously explained in Sect. 3.5, are used here as well: Random Over-Sampling (ROS), Synthetic Minority Over-Sampling (SMOTE), Random Under-Sampling (RUS), Clustering (CUS), Neighborhood Cleaning Rule (NCR), NearMiss-1 (NM1), NearMiss-2 (NM2), NearMiss-3 (NM3), Tomek-Link (TL), SMOTE+Edited Nearest Neighbor (ENN), and SMOTE+TL.

5.4.6 Classification

In this framework, the only classifier we applied was the Random Forests ensemble. This decision was made based on the quantitative results obtained in the experiments with the classification framework for dermoscopy images, where the RF ensemble showed the best results.

5.5 Experiments and Results

In order to evaluate the framework developed, we performed four experiments. Similar to Chapter 4, the description of the experiments, their results, discussion and a brief conclusion are presented in the following subsections. Due to the dataset imbalance problem and the small number of melanoma in comparison with non-melanoma samples (benign, dysplastic and other lesions), the first two experiments were performed using a balanced subset of the data and the leave-one-out-cross-validation (LOOCV) evaluation. For the other two experiments, we used balancing strategies with 10 different larger subsets. Table 5.3 summarizes the configuration of each experiment.

As indicated in the table, the choice of pre-processing, mapping, feature extraction, representation, and classification strategies is the same for all the experiments. All the images were first registered and body hair, if present, was removed. Then the lesions were segmented, producing the external and internal lesion regions. According to the global mapping approach, the spatial features were extracted from the combined (external and internal) region, while the polarimetric features were extracted either from the internal, external or combined areas. The polarimetric features were represented using histograms and moments. Individual spatial and polarimetric features, as well as their combinations, were tested in all the experiments following the three main steps. First, the polarimetric (step 1) and spatial features (step 2) were compared in their own groups. Then, the best of the polarimetric features were combined with the best spatial features and their performance evaluated (step 3). The classification task was

performed using the RF classifier in all the experiments. The Random Forests classifier was trained with 100 unpruned trees.

Table 5.3: Outline and summary of the performed experiments. The abbreviations “tra.” and “tes.” indicate percentages of the data for training and testing, respectively. In this table, the group of suspicious lesions (not melanoma), which were sent to histopathology, is indicated by S , and \sim indicates that the common configuration applies.

	Dataset Size	Task	Pre-processing	Features	Mapping	Representation	Balancing	Classification	Evaluation tra.,tes.
Common:			Registration Hair removal Segmentation	Polarized Spatial	Global	Low: Histogram Moments		RF	SE, SP
Experiment#1: Goal: Evaluation of polarized and spatial features	40	M vs B	\sim	\sim	\sim	\sim	-	\sim	LOOCV
Experiment#2: Goal: Evaluation of polarized and spatial features	40	M vs S+B	\sim	\sim	\sim	\sim	-	\sim	LOOCV
Experiment#3: Goal: Evaluate the effects of balancing strategies	164	M vs B	\sim	\sim	\sim	\sim	-	\sim	10-fold 75%,25%
Experiment#4: Goal: Evaluate the effects of balancing strategies	176	M vs S+B	\sim	\sim	\sim	\sim	-	\sim	10-fold 75%,25%

5.5.1 Experiment #1

Melanoma vs. Benign lesions. LOOCV validation on a subset of the data.

Experiment #1 was performed to evaluate the potential of polarimetric and spatial features for the classification of M vs. B lesions. In this experiment, the portion of the dataset with dysplastic lesions, basal cell carcinoma and keratosis images was excluded. Due to the small number of melanoma samples (22), 10 subsets of 40 samples were considered. Each subset contained 20 melanoma and 20 benign lesions, randomly selected from the original set of 22 melanoma and 149 benign lesion images, respectively. Since a dataset of this size is relatively small, we used the LOOCV for validation.

Table 5.4 illustrates the results obtained in this experiment, indicated in terms of SE and SP. The first part of the table shows the evaluation of polarimetric features (step 1), while the classification performance when using spatial features (step 2) is shown in the second part. The best performing features are highlighted in gray. In step 3, the five leading polarimetric feature sets are used in conjunction with all the spatial features, and their results are shown in the 3rd part of the Table 5.4.

Discussion

Our proposed device, unlike previously proposed partial polarimeter devices [106, 11, 186], acquires three images ($I_{0^\circ}, I_{90^\circ}, I_{45^\circ}$). Thus it provides extra information (U) and completed measurements in terms of DOLP and AOLP. Subsequently, nineteen individual polarized feature and five combinations based on the information provided are used in step 1. The moments and histogram features of internal, external and combined regions of the DOLP ($D_H, D_{H_I}, D_{H_E}, D_M, D_{M_I}, D_{M_E}$), Pol_{int} ($P_H, P_{H_I}, P_{H_E}, D_M, P_{M_I}, P_{M_E}$), and the third Stokes parameter ($U_H, U_{H_I}, U_{H_E}, U_M, U_{M_I}, U_{M_E}$), plus its gradient magnitude U_{Gmag} , were used as individual features. The combined feature sets included all the polarimetric features extracted from the internal ($\{D, P, U\}_I$), external ($\{D, P, U\}_E$) and combined (Pol_a) regions, represented as histograms ($\{D, P, U\}_H$) and moments ($\{D, P, U\}_M$).

As illustrated in the first part of Table 5.3 (shaded in gray), five feature sets, including the DOLP extracted as histograms from the entire lesion (D_H) and the internal region (D_{H_I}), the combination of features extracted from external ($\{D, P, U\}_E$), internal ($\{D, P, U\}_I$) and combined (Pol_a) regions, outperformed the others. Nevertheless, they achieved SE and SP of 66.5% and 70%, respectively, which is better than average but cannot be considered a good result.

In step 2, the performance of spatial features was evaluated. As indicated in the second part of Table 5.3, GLCM (T_2) and color statistics (C_1) outperformed the others showing above average results, while the rest of the features performed poorly. The results of step 3 of the

Table 5.4: Results of experiment #1. This experiment classifies M vs. B lesions using polarimetric and spatial features, as well as their combinations (see the three sections of the table). The results are indicated in terms of SE and SP.

Polarized Features								
list	SE	SP	list	SE	SP	list	SE	SP
D_H	63.0	75.5	D_{H_E}	66.7	60.5	D_{H_I}	67.0	72.0
D_M	62.5	66.5	D_{M_E}	56.5	67.5	D_{M_I}	62.5	64.0
P_H	56.5	67.0	P_{H_E}	49.0	59.0	P_{H_I}	53.0	66.5
P_M	51.0	65.5	P_{M_E}	49.0	57.0	P_{M_I}	61.0	64.5
U_H	60.0	63.5	U_{H_E}	57.5	58.5	U_{H_I}	57.5	61.0
U_M	48.5	65.5	U_{M_E}	44.0	57.5	U_{M_I}	51.0	63.0
U_{Gmag}	58.0	62.5	$\{D, P, U\}_H$	67.0	65.5	$\{D, P, U\}_M$	64.0	68.0
$\{D, P, U\}_E$	68.5	67.0	$\{D, P, U\}_I$	67.5	69.5	Pol_a	67.0	66.5

Spatial Features								
list	SE	SP	list	SE	SP	list	SE	SP
T_1	55.5	62.5	T_2	66.0	61.5	T_3	57.5	65.0
T_4	54.5	58.5	C_1	62.0	68.0	C_2	56.5	61.5
T_1, T_2, T_3, T_4	49.5	59.0	C_1, C_2	61.0	70.5			

Polarized and Spatial Features								
list	SE	SP	list	SE	SP	list	SE	SP
T_1, D_H	64.0	75.5	$T_1, \{D, P, U\}_E$	67.0	68.5	T_1, Pol_a	66.5	66.0
T_2, D_H	59.5	74.0	$T_2, \{D, P, U\}_E$	65.0	69.0	T_2, Pol_a	65.5	67.0
T_3, D_H	58.0	73.5	$T_3, \{D, P, U\}_E$	67.0	68.0	T_3, Pol_a	67.0	67.0
T_4, D_H	58.0	74.5	$T_4, \{D, P, U\}_E$	64.5	67.5	T_4, Pol_a	66.5	67.0
C_1, D_H	63.0	73.5	$C_1, \{D, P, U\}_E$	63.5	68.5	C_1, Pol_a	66.5	66.5
C_2, D_H	58.5	74.0	$C_2, \{D, P, U\}_E$	66.5	70.0	C_2, Pol_a	65.5	66.0
T_1, D_{H_I}	62.5	73.5	$T_1, \{D, P, U\}_I$	67.5	66.5	$\{\text{Spa}, \text{Pol}\}_a$	66.5	67.5
T_2, D_{H_I}	58.5	70.5	$T_2, \{D, P, U\}_I$	63.5	69.0			
T_3, D_{H_I}	56.0	71.5	$T_3, \{D, P, U\}_I$	66.5	66.0			
T_4, D_{H_I}	57.5	70.0	$T_4, \{D, P, U\}_I$	68.0	71.5			
C_1, D_{H_I}	65.0	71.0	$C_1, \{D, P, U\}_I$	67.5	71.5			
C_2, D_{H_I}	56.0	71.5	$C_2, \{D, P, U\}_I$	66.0	67.5			

Table 5.5: Results of experiment #2. This experiment classifies M vs. S+B lesions using polarimetric and spatial features, as well as their combinations (see the three sections of the table). The results are indicated in terms of SE and SP.

Polarimetric Features									
list	SE	SP	list	SE	SP	list	SE	SP	
D_H	63.5	58.5	D_{H_E}	42	48.2	D_{H_I}	65	62.5	
D_M	46	52	D_{M_E}	49.5	56	D_{M_I}	53.5	58.5	
P_H	36.5	38	P_{H_E}	29.5	35	P_{H_I}	42.5	42.5	
P_M	38.5	42	P_{M_E}	27.5	41	P_{M_I}	50.5	51	
U_H	50.5	46	U_{H_E}	54	45	U_{H_I}	47.5	44	
U_M	27.5	38	U_{M_E}	34.5	45	U_{M_I}	37.5	37.5	
U_{Gmag}	48	42	$\{D, P, U\}_H$	44.5	45.5	$\{D, U, P\}_M$	34.5	35	
$\{D, U, P\}_E$	47	51	$\{D, U, P\}_I$	40.5	37	Pol_a	41	45	

Spatial Features									
list	SE	SP	list	SE	SP	list	SE	SP	
T_1	49.5	58.5	T_2	58	52	T_3	55	54.5	
T_4	52	51	C_1	44.5	44.5	C_2	46	45	
T_1, T_2, T_3, T_4	50.5	48.5	C_1, C_2	41	38.5				

Polarimetric and Spatial Features									
list	SE	SP	list	SE	SP	list	SE	SP	
T_1, D_H	60.5	59.5	T_1, D_{H_I}	64.0	64.0	$\{\text{Spa}, \text{Pol}\}_a$	39.5	42.5	
T_2, D_H	58.5	56.0	T_2, D_{H_I}	64.0	61.5				
T_3, D_H	56.5	53.0	T_3, D_{H_I}	58.0	61.5				
T_4, D_H	54.0	56.5	T_4, D_{H_I}	56.0	57.5				
C_1, D_H	48.5	50.5	C_1, D_{H_I}	52.0	53.5				
C_2, D_H	57.0	51.5	C_2, D_{H_I}	60.5	59.0				

experiment (3rd section of Table 5.3) indicate that the combination of spatial and polarimetric features improve the classification performance. This is evident for all the spatial features, except T_2 , which demonstrated almost the same performance with and without polarimetric features. Among all the combinations, the union of T_4 and $\{D, P, U\}_I$ achieved the best performance with SE and SP of 68% and 71.5%, respectively. Overall, the results indicated that the combination of $\{D, P, U\}_I$, $\{D, P, U\}_E$, and Pol_a with spatial features yielded better results.

Conclusion

This experiment was performed to evaluate the potential of polarimetric features for the differentiation of M vs. B lesions. Although the average SE and SP of the best performing features are below 70% and 75%, respectively, the results and comparisons indicated that polarimetric

features outperform spatial features and improve the performance of traditional dermoscopy features, while combined with them.

5.5.2 Experiment #2

Melanoma vs. Suspicious and Benign lesions. LOOCV validation on a subset of the data.

Experiment #2 was performed to evaluate polarimetric features for the differentiation of M vs. S+B lesions. Due to the similar characteristics of melanoma and suspicious lesions, this case is more challenging than the previous experiment. All the lesions excluded from the dataset in the previous experiment are considered suspicious because they were sent for biopsy by dermatologists. Similar to the previous experiment, 10 subsets of 40 samples (20 melanoma, 10 benign and 10 suspicious lesions) were selected. The lesions were randomly selected from the original dataset. Table 5.5 shows the results obtained in the three steps of the experiment (see the three respective parts of the table).

Discussion

As expected, due to the challenging similarities in the samples' appearance, the classification performance decreased in comparison with the previous experiment. Among the polarimetric features, the DOLP histogram features extracted from the internal area (D_{H_I}) and entire lesions (D_H) achieved the highest performances (highlighted in gray color in Table 5.5). As shown in the second part of Table 5.5, the spatial features performed poorly. However, their performance increased after they were combined with the best of the polarimetric features. The combination of T_1 and D_{H_I} achieved the highest SE and SP of 64% and 64%, respectively.

Conclusion

This experiment was performed to evaluate the potential of polarized features for classification of M vs. S+B lesions. The results showed that few polarized features outperform the spatial features. However, the results with an average SE and SP of 64% and 64% are not satisfactory. In general, in comparison to the previous experiment, it is apparent that polarized features are more discriminative for differentiation of melanoma and benign lesions, rather than for melanoma and suspicious lesions.

5.5.3 Experiment #3

Melanoma vs. Benign lesions. A comparison of balancing techniques.

Experiment #3 was performed with the same objective as experiment #1. However, instead of using a small portion of the dataset, we used more samples and tackled the imbalance problem

using balancing techniques. The Over-Sampling and Under-Sampling (US) techniques, as well as their combination in the feature space, were tested on a subset of 164 lesions, similar to the experiment #4 in Sect. 4.5.4. Thus 10 different subsets containing 20 melanoma and 144 benign lesions were randomly selected from the original dataset. This was done in order to create a 10-fold validation, where, for each fold, 75% of the data was kept for training and the rest was used for testing. It should be noted that this 10-fold validation is different from the k fold cross-validation (k-CV) in that it creates a more generalized dataset.

Different balancing techniques, including ROS, SMOTE, RUS, TL, CUS, NM1, NM2, NM3, NCR, SMOTE+ENN, and SMOTE+TL, were tested. However, from which only the US techniques such as RUS, CUS, and NM3 responded to the problem. Similar to the previous experiments, we first investigated the effects of balancing strategies on the polarimetric and spatial features (steps 1 and 2), and then combined the best strategies in step 3. Table 5.6 shows the results from steps 1 and 2, while Table 5.7 illustrates the outcome of step 3.

Discussion

The evaluation of the polarimetric/spatial features and balancing strategies (see Table 5.6) shows that RUS and NM3 are more effective and have fewer cases of overfitting in comparison with CUS. In step 1, the best performance was achieved by P_H and NM3 with SE and SP of 74% and 65%, respectively. Significant results among all the polarimetric features are highlighted in gray in the table. Among these features, D_{H_I} , P_H , P_M , $\{D, P, U\}_M$, $\{D, P, U\}_E$, and all the combinations (Pol_a) were selected to be tested in step 3. It must be noted that, apart from D_{H_I} , the rest of these features were not selected in experiment #1.

The performance analysis of spatial features revealed that GLCM (T_2) outperformed the others by showing slightly better results than average, similar to experiment #1. This was followed by the combination of color features (C_1 , C_2).

As indicated in Table 5.7, the combination of polarimetric and spatial features with balancing techniques leads to better results. Once again, RUS and NM3 demonstrated minimum overfitting at the cost of a lower SE. In this experiment, the combination of T_1 and T_2 with the $\{D, P, U\}_E$ achieved the highest SE and SP, 82% and 61.1%, and 80% and 61.9%, respectively. In general, the union of $\{D, P, U\}_E$ and spatial features was better with an average SE of 76.6% and SP of 60.5%. The union of P_H and spatial features showed the second best result.

Conclusion

The outcome of this experiment in comparison to experiment #1 clearly shows the benefits of balancing strategies. The average SE and SP of the best performing features are above 75%

Table 5.6: The results obtained in steps 1 and 2 of experiment #3. This experiment evaluates the effect of the polarimetric features and balancing strategies on the classification of M vs. B. The table is divided into two sections for the polarimetric and spatial features. The results are indicated in terms of SE and SP.

Polarimetric Features						
Balancing	RUS		CUS		NM3	
Features	SE	SP	SE	SP	SE	SP
D_H	58.0	46.4	80.0	29.1	48.0	54.7
D_{H_E}	56.0	46.9	78.0	38.6	48.0	52.8
D_{H_I}	52.0	57.5	62.0	63.6	68.0	59.7
D_M	58.0	58.6	70.0	63.1	52.0	58.1
D_{M_E}	56.0	43.9	54.0	62.5	52.0	57.8
D_{M_I}	72.0	66.7	84.0	40.8	72.0	64.7
P_H	68.0	57.8	84.0	37.8	74.0	65.8
P_{H_E}	62.0	67.2	86.0	46.9	64.0	69.1
P_{H_I}	66.0	64.7	68.0	64.4	52.0	67.8
P_M	62.0	66.4	68.0	62.8	72.0	61.4
P_{M_E}	62.0	61.7	66.0	64.4	48.0	63.8
P_{M_I}	56.0	52.8	86.0	32.5	58.0	52.8
U_H	58.0	46.9	82.0	28.6	54.0	46.9
U_{H_E}	52.0	52.2	78.0	39.7	60.0	50.8
U_{H_I}	62.0	51.7	64.0	52.2	60.0	61.1
U_M	68.0	47.8	66.0	45.3	66.0	52.2
U_{M_E}	56.0	51.1	52.0	49.1	58.0	51.1
U_{M_I}	56.0	45.8	76.0	38.6	56.0	48.3
U_{Gmag}	64.0	55.3	70	66.9	62.0	60.0
$\{D, P, U\}_H$	66.0	59.44	84.0	35.8	66.0	67.5
$\{D, P, U\}_M$	70.0	60.0	64.0	62.8	70.0	61.9
$\{D, P, U\}_E$	72.0	60.0	84.0	28.9	58.0	64.7
$\{D, P, U\}_I$	56.0	42.5	84.0	23.6	64.0	46.7
Pol _a	64.0	45.2	68.0	52.8	74.0	50.0
Spatial Features						
T_1	62.0	50.3	66.0	51.94	70.0	50.0
T_2	68	61.7	62.0	62.8	64.0	59.4
T_3	50	47.8	74.0	31.9	62.0	55.5
T_4	56	47.5	44.0	45.8	54.0	48.3
C_1	70	47.5	76.0	35.5	62.0	57.5
C_2	70	46.7	64.0	41.9	54.0	58.6
T_1, T_2, T_3, T_4	58.0	48.3	64.0	37.8	54.0	53.1
C_1, C_2	68.0	66.7	78.0	33.1	66.0	64.2

Table 5.7: Results obtained in step 3: combinations of the best polarimetric and spatial features. The results are indicated in terms of SE and SP.

Polarimetric and Spatial Features						
Balancing	RUS		CUS		NM3	
Features	SE	SP	SE	SP	SE	SP
T_1, D_{HI}	70.0	48.3	88.0	23.0	56.0	54.4
T_2, D_{HI}	62.0	47.2	100.0	18.3	56.0	49.4
T_3, D_{HI}	48.0	52.7	100.0	16.4	52.0	53.0
T_4, D_{HI}	58.0	52.7	98.0	22.8	64.0	50.2
C_1, D_{HI}	52.0	50.6	88.0	28.6	40.0	61.6
C_2, D_{HI}	56.0	51.4	88.0	24.7	52.0	51.9
T_1, P_H	72.0	62.8	90.0	25.0	68.0	63.3
T_2, P_H	76.0	66.9	84.0	31.7	76.0	68.0
T_3, P_H	74.0	68.9	84.0	40.6	70.0	64.1
T_4, P_H	76.0	61.1	88.0	33.1	72.0	63.3
C_1, P_H	72.0	66.7	84.0	34.7	66.0	67.2
C_2, P_H	70.0	64.2	84.0	50.6	62.0	69.4
T_1, P_M	70.0	51.7	76.0	60.6	70.0	56.1
T_2, P_M	78.0	56.9	78.0	56.1	72.0	55.5
T_3, P_M	76.0	56.9	74.0	66.9	62.0	58.0
T_4, P_M	70.0	50.0	92.0	33.0	68.0	51.9
C_1, P_M	64.0	50.0	72.0	29.1	56.0	58.6
C_2, P_M	68.0	53.6	78.0	29.1	66.0	52.2
$T_1, \{D, P, U\}_M$	86.0	55.2	76.0	60.0	76.0	56.6
$T_2, \{D, P, U\}_M$	66.0	53.3	68.0	49.7	68.0	54.7
$T_3, \{D, P, U\}_M$	86.0	55.8	70.0	63.0	68.0	59.4
$T_4, \{D, P, U\}_M$	68.0	48.6	92.0	30.6	72.0	53.9
$C_1, \{D, P, U\}_M$	60.0	55.8	56.0	45.3	58.0	58.6
$C_2, \{D, P, U\}_M$	64.0	49.4	76.0	40.8	66.0	57.8
$T_1, \{D, P, U\}_E$	82.0	61.1	90.0	27.8	66.0	60.8
$T_2, \{D, P, U\}_E$	82.0	60.6	84.0	26.9	70.0	56.4
$T_3, \{D, P, U\}_E$	66.0	65.3	90.0	24.4	76.0	61.1
$T_4, \{D, P, U\}_E$	76.0	60.6	88.0	25.6	70.0	61.1
$C_1, \{D, P, U\}_E$	74.0	57.2	84.0	32.8	76.0	61.9
$C_2, \{D, P, U\}_E$	80.0	58.0	88.0	33.0	76.0	62.8
T_1, Pol_a	72.0	60.5	84.0	26.1	66.0	67.8
T_2, Pol_a	62.0	64.4	80.0	35.0	72.0	63.3
T_3, Pol_a	80.0	61.9	82.0	31.7	58.0	63.6
T_4, Pol_a	76.0	58.6	84.0	32.8	64.0	64.1
C_1, Pol_a	74.0	58.8	84.0	30.8	64.0	65.8
C_2, Pol_a	62.0	59.5	75.0	51.1	63.0	64.1

and 60%, respectively. In other words, balancing strategies helped to improve the SE without overfitting the classifier.

5.5.4 Experiment #4

Melanoma vs. Suspicious and Benign lesions. A comparison of balancing techniques.

This experiment is a revision of experiment #2, where, instead of small portions of the dataset and LOOCV validation, we applied balancing strategies in conjunction with a 10-fold validation on a larger data subset. Thus, a subset of 176 lesions containing 20 melanoma, 12 suspicious and 144 benign lesions was randomly selected from the original dataset. Similar to the previous experiment, in each fold, 75% of the data was used for training while the rest was kept for testing.

After testing different balancing strategies, the same methods as in the previous experiment were considered: RUS, CUS and NM3. The results of using these techniques in steps 1 and 2 of this experiment are shown in Table 5.8.

The combinations of the best performing polarimetric and spatial features were then tested in step 3. The results obtained in this step are shown in Table 5.9.

Discussion

The performance analysis of the polarimetric features (see the first part of Table 5.7) shows that the features extracted from $\text{Pol}_{int}(P)$ in conjunction with NM3 and RUS achieve better results. The best performance was achieved by P_{H_E} with SE and SP of 72% and 60.7% followed by P_{M_E} with SE and SP of 70% and 65.1%, respectively. The D_{M_I} , P_H , P_{H_E} , P_{M_E} , and $\{D, P, U\}_H$ features, highlighted in the table, as well as Pol_a , were considered to be used in combinations in step 3.

Among spatial features, T_2 once again outperformed the other features with all three balancing strategies. This feature achieved the highest performance with SE of 78% and SP of 61%.

The combinations of polarimetric and spatial features are given in Table 5.9. Their comparison shows that the union of $\{D, P, U\}_H$ and spatial features yields a better classification than any other combination. Overall, the best performances in this step were achieved with the SE and SP above 70% and 60%, respectively, which demonstrates a clear improvement in sensitivity in comparison with experiment #2.

As for the balancing techniques, RUS and NM3 resulted in less overfitting than CUS. However, the best performance was achieved with the union of T_3 and P_{M_E} when the CUS technique was used (SE and SP of 74% and 71%, respectively).

Table 5.8: Results obtained in steps 1 and 2 of experiment #4. This experiment evaluates the effect of polarimetric features and balancing strategies on the classification of M vs. S+B. The table is divided into two sections for polarimetric and spatial features. The results are indicated in terms of SE and SP.

Polarimetric Features						
Balancing	RUS		CUS		NM3	
Features	SE	SP	SE	SP	SE	SP
D_H	46.0	47.4	84.0	25.6	52.0	39.7
D_{H_E}	48.0	49.7	82.0	36.4	42.0	50.7
D_{H_I}	44.0	60.3	42.0	62.3	50.0	57.1
D_M	48.0	62.0	50.0	60.7	46.0	53.1
D_{M_E}	38.0	51.8	34.0	55.9	40.0	52.5
D_{M_I}	60.0	61.3	76.0	39.5	64.0	71.8
P_H	60.0	66.7	74.0	37.9	64.0	70.5
P_{H_E}	72.0	60.7	82.0	46.4	62.0	66.4
P_{H_I}	58.0	65.4	62.0	67.4	62.0	68.2
P_M	64.0	58.2	58.0	74.1	60.0	64.1
P_{M_E}	58.0	61.8	64.0	62.8	70.0	65.1
P_{M_I}	60.0	48.2	86.0	29.5	62.0	47.7
U_H	56.0	45.4	78.0	24.8	60.0	49.7
U_{H_E}	56.0	48.2	74.0	36.7	68.0	50.7
U_{H_I}	60.0	49.7	58.0	51.0	54.0	56.9
U_M	52.0	53.3	64.0	45.4	62.0	46.9
U_{M_E}	60.0	54.4	50.0	48.7	48.0	54.8
U_{M_I}	62.0	51.5	78.0	39.2	56.0	55.4
U_{Gmag}	44.0	59.5	44.0	63.3	44.0	57.7
$\{D, P, U\}_H$	70.0	62.6	74.0	33.1	64.0	66.0
$\{D, P, U\}_M$	64.0	53.3	50.0	59.2	58.0	63.6
$\{D, P, U\}_E$	64.0	62.0	78.0	28.2	56.0	65.4
$\{D, P, U\}_I$	50.0	45.1	90.0	20.3	66.0	47.7
Pol_a	58.0	53.3	74.0	57.9	56.0	53.8
Spatial Features						
T_1	56.0	51.0	54.0	56.9	58.0	50.7
T_2	60.0	66.9	72.0	68.2	78.0	61.0
T_3	50.0	51.5	66.0	33.0	52.0	48.9
T_4	62.0	51.2	44.0	51.5	56.0	51.5
C_1	46.0	54.3	62.0	38.2	56.0	52.6
C_2	42.0	54.8	68.0	40.7	46.0	54.1
T_1, T_2, T_3, T_4	58.0	48.7	74.0	33.8	50.0	55.6
C_1, C_2	62.0	61.0	78.0	34.1	58.0	63.6

Table 5.9: Results obtained in step 3: combinations of the best polarimetric and spatial features. The results are indicated in terms of SE and SP.

Polarimetric and Spatial Features						
Balancing	RUS		CUS		NM3	
Features	SE	SP	SE	SP	SE	SP
T_1, D_{M_I}	64.0	53.3	62.0	55.9	68.0	49.7
T_2, D_{M_I}	54.0	50.0	56.0	55.1	54.0	51.0
T_3, D_{M_I}	68.0	58.9	68.0	68.2	62.0	64.3
T_4, D_{M_I}	60.0	48.7	70.0	40.5	58.0	54.8
C_1, D_{M_I}	44.0	56.9	66.0	39.4	48.0	56.6
C_2, D_{M_I}	56.0	52.0	50.0	47.4	52.0	55.6
T_1, P_{H_E}	66.0	69.7	72.0	28.9	72.0	64.1
T_2, P_{H_E}	62.0	66.9	76.0	31.0	64.0	70.2
T_3, P_{H_E}	64.0	65.3	74.0	36.1	62.0	69.7
T_4, P_{H_E}	62.0	61.8	80.0	27.4	66.0	64.8
C_1, P_{H_E}	56.0	67.7	72.0	37.7	66.0	64.8
C_2, P_{H_E}	64.0	62.8	70.0	46.6	62.0	66.4
T_1, P_H	64.0	63.6	72.0	31.2	66.0	69.7
T_2, P_H	62.0	70.0	76.0	30.2	68.0	66.1
T_3, P_H	66.0	64.1	76.0	36.7	62.0	69.5
T_4, P_H	66.0	60.0	74.0	33.6	66.0	68.2
C_1, P_H	70.0	60.0	68.0	38.2	64.0	66.9
C_2, P_H	62.0	66.1	76.0	48.7	64.0	73.3
T_1, P_{M_E}	62.0	50.0	72.0	59.5	62.0	55.9
T_2, P_{M_E}	66.0	55.9	60.0	56.1	68.0	62.0
T_3, P_{M_E}	70.0	60.0	74.0	71.3	72.0	67.9
T_4, P_{M_E}	60.0	52.0	68.0	42.0	58.0	55.4
C_1, P_{M_E}	52.0	49.2	62.0	40.7	44.0	55.6
C_2, P_{M_E}	52.0	52.5	80.0	30.7	50.0	52.5
$T_1, \{D, P, U\}_H$	56.0	60.5	78.0	23.3	68.0	60.5
$T_2, \{D, P, U\}_H$	64.0	64.6	74.0	26.4	68.0	63.5
$T_3, \{D, P, U\}_H$	66.0	67.9	74.0	30.0	68.0	66.1
$T_4, \{D, P, U\}_H$	66.0	57.1	70.0	26.9	70.0	65.1
$C_1, \{D, P, U\}_H$	66.0	63.3	74.0	30.0	64.0	61.2
$C_2, \{D, P, U\}_H$	72.0	60.2	72.0	35.3	68.0	59.7
T_1, Pol_a	58.0	64.1	72.0	31.5	66.0	68.7
T_2, Pol_a	70.0	64.8	76.0	37.1	66.0	70.5
T_3, Pol_a	74.0	59.4	70.0	31.5	64.0	65.9
T_4, Pol_a	66.0	64.1	78.0	33.3	56.0	67.6
C_1, Pol_a	70.0	65.1	78.0	31.2	62.0	62.3
C_2, Pol_a	58.0	62.3	74.0	33.3	66.0	64.8

Conclusion

In this experiment, we studied the effect of the US balancing strategies and polarimetric features on the classification of M .vs S+B lesions. The results, contrary to those obtained in experiment #2, were more satisfactory and comparable with the results of experiment #3. Similar to our previous experiments, the combination of polarimetric and spatial features led to a better classification performance.

5.6 Conclusion

In this chapter, we presented the first prototype of our polarimetric dermoscope and the automated melanoma classification framework. The polarimetric dermoscope is designed to automatically capture three polarized images, which later are used to define the first three Stokes parameters (I , Q , U) and some polarization characteristics, such as the degree of linear polarization (DOLP), polarization intensity (Pol_{int}), and angle of linear polarization (AOLP). The three color polarized images are acquired while the polarizer of the PSA unit is automatically rotated 90° , 45° , and 0° with reference to the horizontal axis of the fixed polarizer in the PSG unit.

The polarimetric dermoscope developed was tested over the past three years in the Melanoma Unit of the Clinic Hospital of Barcelona, where images of more than 200 lesions were documented and recorded using this device. From the polarimetric images acquired, 197 cases were suitable for use in further studies. These images were studied in terms of their potential for both visual screening and information for accurate automated classification.

Concerning the visual screening, it was observed that using the current setup of the dermoscope, only the three Stokes parameters and the DOLP can be used, whereas the AOLP was excluded from further studies due to its low intensity values (almost 0). The low intensity values of this parameter are caused by the light source direction in the current setup. Studying the remaining parameters revealed no evident visual difference between melanoma and the rest of the lesions. It was also observed that in the current batch of images, some air bubbles and artifacts are apparent in I_{0° and I_{45° , which make further analysis and understanding of the colors in the Stokes parameters image less accurate.

Although no evident visual differences between melanoma and the rest of the lesions were observed in the primary studies, the polrimetric images were used in the next study as a source for automatic recognition. In this regard, we proposed an automatic recognition system which takes advantage of polarimetric and spatial features. After several studies, three polarimetric features extracted from the gray level polarimetric images were considered: the DOLP (D), Pol_{int} (P) and the third Stokes parameter (U). These features were extracted from the external,

internal or combined segmented regions of each lesion and were then represented using either histograms or moments. In total, nineteen individual and combined polarimetric and 8 spatial features were tested in several experiments. The spatial features were extracted only from the cross-polarized (I_{90°) image.

In order to test the potential of the polarimetric features, we performed 4 experiments: two on the classification of M vs. B lesions (experiments #1 and #3) and two on the classification of M vs. S+B lesions (experiments #2 and #4). Experiments #1 and #2 were performed using a small subset of 40 lesions, the LOOCV validation, without considering balancing strategies. The results of experiment #1 showed that polarimetric features outperformed their spatial counterparts, whereas the outcome of the spatial features improved when they were combined with polarimetric features. However, the benefits of polarimetric features were not as evident in experiment #2. In this experiment, except for several polarimetric and combined features, the rest of the measurements were below expectation. The poor results are possibly due to the lack of training data and a correlated feature space between the two classes.

Subsequently, in the next experiment, we tested the balancing strategies while larger subsets of data were used. The study of different over and under-sampling balancing techniques revealed that only a few US techniques were responsive to the defined problem: RUS, NM3 and CUS. These algorithms were tested in experiments #3 and #4. The results of these experiments showed an improvement in SE by a minimum of 5%. It was also concluded that some polarimetric features, such as Pol *int* (P), in combination with balancing strategies can achieve comparable results when used for the classification of M vs. S+B lesions. In general, the DOLP (D) features achieved better results in experiments #1 and #2 without balancing techniques, while the US techniques showed the potential of Pol *int* (P) features in the third and fourth experiments.

Observing the results from all the experiments and comparing them with those reported in the previous chapter, one might question the values of the SE and SP achieved. It should be noted that the same color and texture features (spatial features) were used on the cross-polarized images and their results were lower than those of the polarimetric features. This low performance is likely due to a lower magnification in the cross-polarized images. Higher magnification of these images will lead to a better performance in both polarimetric and spatial features.

The research and material presented in this chapter were based on our first study of polarimetry images and a prototype of our developed dermoscope. Our observations and studies of the first try opens new avenues and possible improvements that are explained in the following section.

5.6.1 Future work

In future research, we would like to explore several aspects including:

- Acquiring a series of images with the prior application of an ultra-sound gel or water to the skin surface. Using gel or water on the skin will ensure that images I_{0° and I_{45° are air bubble free, and will allow a better color understanding in the polarimetric images.
- Adapting the current dermoscope for acquiring more polarized images. Using more images to estimate the Stokes and polarimetric parameters will reduce measurement errors.
- Automating the acquisition and focus of the camera to eliminate the need of hand holding and manual focus. If implemented, there will be no need for image registration, and it will positively affect measurement accuracy.
- Use of division-of-focal-plane polarimeter sensor instead of rotating the polarizer coupled with the camera. The DOLP polarimeters are imaging sensors constructed by joining polarization filters and imaging elements to create a new array of polarized sensitive pixels [148]. Using such a system eliminates the need of registration and electrical rotation of the polarizers at the cost of lower resolution.
- Increasing the magnification power of the dermoscope to improve the differentiation potential of the polarimetric and spatial features.
- Evaluating the automated classification framework with the polarized images of color channels rather than in gray scale.

Chapter 6

Conclusion

Skin cancer is a worldwide major health problems. According to the World Health Organization (WHO), over 2 to 3 million non-melanoma cancer and 132,000 melanoma cancer cases occurs annually around world. Among, the different types of skin cancer, a melanoma is the deadliest. This cancer is incurable in its advanced stages with less than 20% survival rate of 5-10 years [183]. However, if discovered in its early stages, it can be treated easier than any other cancer. Subsequently, to facilitate an early diagnosis, numerous studies and research have been carried out by dermatologists and scientists.

The clinical prognosis of early stage melanoma is commonly based on a set of rules and guidelines such as “ABCDE” [9] or the Glasgow 7-point checklist [9]. These criteria are meant for a human reader to visually inspect an image of a skin lesion and characterize the lesion based on visual cues. However, due to its nature, this is challenging and prone to errors. Consequently, double reading and Computer-Aided Diagnosis (CAD) systems are proposed to assist dermatologists and clinicians. Numerous CAD systems have been proposed by the computer science community in the past decade. These algorithms take advantage of image processing and machine learning techniques to mimic the main criteria of a clinical diagnosis and provide a diagnosis that can assist dermatologists. The algorithms developed are mostly based on dermoscopy modality, one of the major skin screening techniques. Dermoscopy takes advantage of cross-polarized properties to remove specular reflections and capture information from the epidermis and papillary dermis.

Due to the importance of an early diagnosis of melanoma, the objective of this research was set to develop a CAD system for the classification of melanoma lesions. However, instead of using the well-known dermoscopy modality, going beyond cross-polarized characteristics and taking advantage of a full or partial polarimetry system was considered. Accordingly, we reviewed the literature on the computer science applications related to melanoma diagnosis.

These studies confirmed our previous expectations: (i) extensive use of dermoscopic images as the source of information, (ii) use of classical approaches for feature extraction which mimic the characteristics of “ABCD” rule. These findings greatly motivated us more to pursue our original objectives:

- Develop a general Computer-Aided Diagnosis system for melanoma lesions based on dermoscopy and polarimetry
- Analyze the full potential of polarimetry for skin imaging and the automated classification of melanoma lesions.

In the following, first a summary of this thesis is presented, then our main contributions and limitations are discussed and finally, future avenues of research are presented.

6.1 Summary of the thesis

This thesis starts with an introduction to the structure of human skin, a definition of pigmented skin lesions and, in particular, melanoma lesions. After describing melanoma cancer and the clinical ways of diagnosis, the well-used skin imaging methods are discussed which leads us to our main motivations as described above.

In this regard, to find a feasible way of developing a polarimeter system, we studied the polarization properties and the literature reviews related to polarimetry imaging for biological tissues. Chapter 2 represents our findings and a review of the state of the art in this field. Although the main goal of this thesis is to find new imaging techniques for skin cancer through the use of polarimetry system, in order to develop an automated classification framework, a study of previous studies and steps of CAD systems was necessary. It was also important to develop a framework which can perform well on dermoscopy images. Subsequently, in Chapter 3 the basics of machine learning and image processing tools necessary for developing an automated recognition were represented. Based on the conclusions drawn from the state of the art related to CAD systems for detection of melanoma, our new automated framework for dermoscopy modality was proposed in Chapter 4. The proposed framework considers different aspects of machine learning and automated classification such as:

- Pre-processing that, due to the nature of the problem, includes both hair-removal and segmentation.
- Two mapping strategies (global and local).
- Different feature extraction techniques (well-known texture features along with shape and color features common to the problem of skin lesions).

- Low representation of the features and high representation via well-known techniques such as Bag of Words (BoW) and Sparse Coded Features (SCF).
- Different feature and data space balancing techniques to overcome the imbalance problem. The imbalance problem appears quite frequently in real world applications such as medical imaging (samples of malignant cases are fewer than benign cases).
- The variety of classifiers such as the support Vector Machine (SVM), Gradient Boosting (GB), and Random Forests (RF).

The above aspects were considered to provide a general framework that is data independent and can be easily adapted for different datasets. These aspects were tested with 5 different experiments, using two datasets. Considering the problem of melanoma classification and, based on the results obtained from the five experiments, some general conclusions are drawn such as:

- Texture features are as important as the traditional color and shape features.
- It is crucial to use balance training and, for the specific case of skin lesions, under-sampling techniques are proven to be more effective.
- A Random Forests classifier is a capable tool and its efficiency was proven through a variety of experiments.
- High representations of the features via well-known techniques such as BoW and SCF are effective, however, they are more costly considering time and complexity.

Considering our findings from previous chapters, Chapter 5 presents our developed polarimetric dermoscope and our adapted automatic classification framework based on polarimetric images. We propose a first polarimetric dermoscope with the ability to acquire the three color images required to measure the first three Stokes parameters. This device was tested in the Hospital Clinic of Barcelona over the past three years. Using the acquired images with the proposed device, an automated classification framework was proposed to consider polarized features besides dermoscopy (spatial) features. The proposed CAD system was tested over four different experiments using our collected dataset. The experiment results conclude the potential of polarized features and drawbacks of the first prototype (see Sect. 5.6).

6.2 Contributions

The following are the major contributions of this thesis:

- An automated classification framework of melanoma lesions using dermoscopic images. The framework considers various aspects of the CAD system and proposes different techniques for each step including, hair removal, segmentation, global and local mapping, feature extraction, feature representation, balancing and finally, classification. Towards the development of this framework an extensive study of different feature extraction approaches, balancing methods and classifiers were made.
- The first partial polarimetric dermoscope able to acquire three polarized images with automated polarizer rotations (I_{90° , I_{45° and I_{0°) and an automated classification framework of melanoma classification using polarimetric images. This framework considers polarized features such as the degree of linear polarization (DOLP), Pol_{int} and some Stokes parameters as well as the classical dermoscopic features as the source of information for the differentiation of melanoma lesions.

Some aspects of the aforementioned contributions are published in [157, 161, 159, 158], while others are still in progress [156, 155].

6.3 Limitations

This section lists the major limitations observed through this thesis.

- Data collections and collaboration with the hospital.
Collecting medical data is challenging and time consuming for both the researcher and clinicians and unfortunately three years of doctoral research might not be enough to build a reasonable collection of data. Due to the difficulties and drawbacks of the software and devices developed, data acquisition is a difficult and time consuming task for the clinicians. Moreover, generally due to internal policies of hospitals, it is not possible for the researcher to attend the acquisition process.
- Manual focus and holding of the polarimetric dermoscope.
The current system requires manual focus before each acquisition and the necessity of holding the dermoscope with hand on the skin during the acquisition. This requirements often makes the acquired set out of focus and not aligned.
- Low magnification of the polarimetric dermoscope and the existence of air bubbles in some cases of I_{0° and I_{45° . Although a magnifier is used to the acquire images, in comparison to traditional dermoscopes, a stronger magnifier is required, as mentioned previously, the use of water or gel will remove possible air bubbles and reflections in these images.

6.4 Future work

Future work regarding each developed frameworks was previously discussed (see Sect. 5.6 & Sect. 4.6). However, regarding all the objectives, some feasible avenues, with respect to the two categories of hardware and software can be considered.

In terms of the developed software and CAD system besides performing more tests and analysis using larger datasets, the latest approach of machine learning techniques such as deep learning can be employed [150, 52]. There is also room for improvement in different aspects of the software developed such as hair removal, registration, balancing and etc. Last but not least the developed software can be improved to be applicable in real time.

In terms of hardware or the developed polarimetric dermoscope, besides considering improvement of the focus and camera magnification, water and gel can be applied to the skin prior to acquisition to minimize artifacts in I_{0° and I_{45° . Most importantly, continued testing and collaboration with hospitals is required in order to improve the device and overcome drawbacks.

Bibliography

- [1] Q. Abbas, E.M. Celebi, and I.F. Garcia. Computer-aided pattern classification system for dermoscopy images. *Skin Research Technology*, 18(3):278–289, 2012.
- [2] Q. Abbas, E.M. Celebi, I.F. Garcia, and W. Ahmad. Melanoma recognition framework based on expert definition of abcd for dermoscopic images. *Skin Research Technology*, 19(1):e93–e102, 2013.
- [3] Q. Abbas, ME Celebi, and I.F. Garcia. Hair removal methods: A comparative study for dermoscopy images. *Biomedical signal processing and control*, 6(4):395–404, 2011.
- [4] Q. Abbas, M.E. Celebi, and I.F. Garcia. Skin tumor area extraction using an improved dynamic programming approach. *Skin Research Technology*, 18(2):133–142, 2012.
- [5] Q. Abbas, I.F. Garcia, E.M. Celebi, and W. Ahmad. A feature-preserving hair removal algorithm for dermoscopy images. *Skin Research Technology*, 19:e27–e36, 2013.
- [6] Q. Abbas, I.F. Garcia, E.M. Celebi, W. Ahmad, and Q. Mushtaq. Unified approach for lesion border detection based on mixture modeling and local entropy thresholding. *Skin Research Technology*, 19(3):314–319, 2013.
- [7] Q. Abbas, I.F. Garcia, and M. Rashid. Automatic skin tumour border detection for digital dermoscopy using a new digital image analysis scheme. *British Journal of Biomedical Science*, 67(4):177–183, 2010.
- [8] Q. Abbas, I.F. Garcia, and M. Rashid. Unsupervised skin lesions border detection via two-dimensional image analysis. *Computer Methods and Programs in Biomedicine*, 27(1):65–78, 2010.
- [9] N.R. Abbasi, H.M. Shaw, D.S. Rigel, R.J. Friedman, W.H. McCarthy, I. Osman, A.W. Kopf, and D. Polsky. Early diagnosis of cutaneous melanoma: revisiting the abcd criteria. *Jama*, 292(22):2771–2776, 2004.

- [10] M. Aharon, M. Elad, and A. Bruckstein. K-svd: an algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Trans. Signal Processing*, 54(11):4311–4322, 2006.
- [11] M. Anastasiadou, A.D. Martino, D. Clement, F. Liège, B. Laude-Boulesteix, N. Quang, J. Dreyfuss, B. Huynh, A. Nazac, L. Schwartz, et al. Polarimetric imaging for the diagnosis of cervical cancer. *physica status solidi (c)*, 5(5):1423–1426, 2008.
- [12] R.R. Anderson and J.A. Parrish. The optics of human skin. *Journal of Investigative Dermatology*, 77(1):13–19, 1981.
- [13] M.R. Antonelli. *Biomedical applications of polarimetric imaging contrast. Initial studies for scattering media and human tissues*. PhD thesis, 2011.
- [14] D. Arthur and S. Vassilvitskii. k-means++: the advantages of careful seeding. In *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 1027–1035. Society for Industrial and Applied Mathematics, 2007.
- [15] B. Auffarth, M. López, and J. Cerquides. Comparison of redundancy and relevance measures for feature selection in tissue classification of ct images. In *ICDM*, pages 248–262. Springer, 2010.
- [16] R. Azzam. Photopolarimetric measurement of the mueller matrix by fourier analysis of a single detected signal. *Optics Letters*, 2(6):148–150, 1978.
- [17] C. Barata, M.A. Figueiredo, E.M. Celebi, and J.S. Marques. Color identification in dermoscopy images using gaussian mixture models. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3611–3615. IEEE, 2014.
- [18] C. Barata, J.S. Marques, and M.E. Celebi. Towards an automatic bag-of-features model for the classification of dermoscopy images: The influence of segmentation. In *8th International Symposium on Image and Signal Processing and Analysis (ISPA)*, pages 274–279. IEEE, 2013.
- [19] C. Barata, J.S. Marques, and J. Rozeira. The role of keypoint sampling on the classification of melanomas in dermoscopy images using bag-of-features. In *Pattern Recognition and Image Analysis*, pages 715–723. Springer, 2013.
- [20] C. Barata, M. Ruela, M. Francisco, T. Mendonça, and J.S. Marques. Two systems for the detection of melanomas in dermoscopy images using texture and color features. *IEEE Journal of Systems*, 8(3):965–979, Sept 2014.

- [21] CAZ Barcelos and VB Pires. An automatic based nonlinear diffusion equations scheme for skin lesion segmentation. *Applied Mathematics and Computation*, 215(1):251–261, 2009.
- [22] G.E. Batista, A.L. Bazzan, and M.C. Monard. Balancing training data for automated annotation of keywords: a case study. In *WOB*, pages 10–18, 2003.
- [23] G.E. Batista, R.C. Prati, and M.C. Monard. A study of the behavior of several methods for balancing machine learning training data. *ACM Sigkdd Explorations Newsletter*, 6(1):20–29, 2004.
- [24] H. Bay, T.uytelaars, and L. Van Gool. Surf: Speeded up robust features. In *European Conference on Computer Vision*, pages 404–417. Springer, 2006.
- [25] C. Becker, R. Rigamonti, V. Lepetit, and P. Fua. Supervised feature learning for curvilinear structure segmentation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 526–533. Springer, 2013.
- [26] A. Benassi, S. Cohen, and J. Istad. Identifying the multifractional function of a gaussian process. *Statistics & probability letters*, 39(4):337–345, 1998.
- [27] C. Benvenuto-Andrade, S.W. Dusza, A.L.C. Agero, A. Scope, M. Rajadhyaksha, A.C. Halpern, and A.A. Marghoob. Differences between polarized light dermoscopy and immersion contact dermoscopy for the evaluation of skin lesions. *Archive of Dermatology*, 143(3):329–338, 2007.
- [28] V. Berenguer, D. Ruiz, and A. Soriano. Application of hidden markov models to melanoma diagnosis. In J.M. Corchado, S. Rodriguez, J. Llinas, and J.M. Molina, editors, *International Symposium on Distributed Computing and Artificial Intelligence*, volume 50, pages 357–365, Berlin, 2009. Springer.
- [29] J.C. Bezdek, R. Ehrlich, and W. Full. Fcm: The fuzzy c-means clustering algorithm. *Computers & Geosciences*, 10(2):191–203, 1984.
- [30] Christopher M Bishop. *Pattern recognition and machine learning*. Springer, 2006.
- [31] B. Boulbry, T.A. Germer, and J.C. Ramella-Roman. A novel hemispherical spectro-polarimetric scattering instrument for skin lesion imaging. In *Biomedical Optics*, pages 60780R–60780R. International Society for Optics and Photonics, 2006.
- [32] L. Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

- [33] M. Burroni, P. Sbano, G. Cevenini, M. Risulo, G. Dell'Eva, P. Barbini, C. Miracco, M. Fimiani, L. Andreassi, and P. Rubegni. Dysplastic naevus vs. in situ melanoma: digital dermoscopy analysis. *Birtish Journal of Dermatology*, 152(4):679–684, 2005.
- [34] G. Capdehourat, A. Corez, A. Bazzano, R. Alonso, and P. Musé. Toward a combined tool to assist dermatologists in melanoma detection from dermoscopic images of pigmented skin lesions. *Pattern Recognition Letters*, 32(16):2187–2196, 2011.
- [35] G. Capdehourat, A. Corez, A. Bazzano, and P. Musé. Pigmented skin lesions classification using dermatoscopic images. In *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, pages 537–544. Springer, 2009.
- [36] E.M. Celebi, Q. Wen, S. Hwang, H. Iyatomi, and G. Schaefer. Lesion border detection in dermoscopy images using ensembles of thresholding methods. *Skin Research Technology*, 19(1):e252–e258, 2013.
- [37] M.E. Celebi, Y.A. Aslandogan, W.V. Stoecker, H. Iyatomi, H. Oka, and X. Chen. Unsupervised border detection in dermoscopy images. *Skin Research Technology*, 13(4):454–462, 2007.
- [38] M.E. Celebi, S. Hwang, H. Iyatomi, and G. Schaefer. Robust border detection in dermoscopy images using threshold fusion. In *IEEE International Conference on Image Processing (ICIP)*, pages 2541–44, Piscataway, NJ, 2010. IEEE Press.
- [39] M.E. Celebi, H. Iyatomi, G. Schaefer, and W.V. Stoecker. Approximate lesion localization in dermoscopy images. *Skin Research Technology*, 15(3):314–322, 2009.
- [40] M.E. Celebi, H. Iyatomi, G. Schaefer, and W.V. Stoecker. Lesion border detection in dermoscopy images. *Computerized Medical Imaging and Graphics*, 33(2):148–153, 2009.
- [41] M.E. Celebi, H.A. Kingravi, H. Iyatomi, Y.A. Aslandogan, W.V. Stoecker, R.H. Moss, et al. Border detection in dermoscopy images using statistical region merging. *Skin Research Technology*, 14(3):347–353, 2008.
- [42] M.E. Celebi, H.A. Kingravi, H. Iyatomi, J. Lee, Y.A. Aslandogan, WV Stoecker, et al. Fast and accurate border detection in dermoscopy images using statistical region merging. In Josien P. W. Pluim and Joseph M. Reinhardt, editors, *Proc. SPIE*, volume 6512 of *Medical Imaging: Image Processing*. SPIE, 2007.
- [43] M.E. Celebi, H.A. Kingravi, B. Uddin, H. Iyatomi, Y.A. Aslandogan, W.V. Stoecker, and R.H. Moss. A methodological approach to the classification of dermoscopy images. *Computerized Medical Imaging and Graphics*, 31(6):362–373, 2007.

- [44] M.E. Celebi, H.A. Kingravi, and P.A. Vela. A comparative study of efficient initialization methods for the k-means clustering algorithm. *Expert Systems with Applications*, 40(1):200–210, 2013.
- [45] M.E. Celebi and A. Zornberg. Automated quantification of clinically significant colors in dermoscopy images and its application to skin lesion classification. *IEEE Journal of Systems*, 8(3):980–984, Sept 2014.
- [46] N.V. Chawla. Data mining for imbalanced datasets: An overview. In *Data mining and knowledge discovery handbook*, pages 853–867. Springer, 2005.
- [47] N.V. Chawla, K.W. Bowyer, L.O. Hall, and W.P. Kegelmeyer. Smote: synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, pages 321–357, 2002.
- [48] S.S. Chen, D.L. Donoho, and M.A. Saunders. Atomic decomposition by basis pursuit. *SIAM journal on Scientific Computing*, 20(1):33–61, 1998.
- [49] A. Chiem, A. Al-Jumaily, and R.N. Khushaba. A novel hybrid system for skin lesion detection. In M. Palaniswami, Slaven Marusic, and Yee Wei Law, editors, *International Conference on Intelligent Sensors, Sensor Networks and Information (ISSNIP)*, pages 567–572, Piscataway, NJ, 2007. IEEE Press.
- [50] D.A. Clausi. An analysis of co-occurrence texture statistics as a function of grey level quantization. *Canadian Journal of Remote Sensing*, 28(1):45–62, 2002.
- [51] N. Codella, J. Cai, M. Abedini, R. Garnavi, A. Halpern, and J.R. Smith. Deep learning, sparse coding, and svm for melanoma recognition in dermoscopy images. In *Machine Learning in Medical Imaging*, pages 118–126. Springer, 2015.
- [52] E. Collett. Measurement of the four stokes polarization parameters with a single circular polarizer. *Optics Communications*, 52(2):77–80, 1984.
- [53] E. Compain, S. Poirier, and B. Drevillon. General and self-consistent method for the calibration of polarization modulators, polarimeters, and mueller-matrix ellipsometers. *Applied Optics*, 38(16):3490–3502, 1999.
- [54] A. Criminisi, P. Perez, and K. Toyama. Object removal by exemplar-based inpainting. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2, pages II–721. IEEE, 2003.
- [55] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, pages 886–893. IEEE, 2005.

- [56] J.G. Daugman. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *JOSA A*, 2(7):1160–1169, 1985.
- [57] G.M. Davis, S.G. Mallat, and Z. Zhang. Adaptive time-frequency decompositions. *Optics Express*, 33(7):2183–2191, 1994.
- [58] G.R. Day and R. Barbour. Automated melanoma diagnosis: where are we at? *Skin Research Technology*, 6(1):1–5, 2000.
- [59] G.R. Day and R.H. Barbour. Automated skin lesion screening—a new approach. *Melanoma Research*, 11(1):31–35, 2001.
- [60] A. De Martino, E. Garcia-Caurel, B. Laude, and B. Drévillon. General methods for optimized design and calibration of mueller polarimeters. *Thin Solid Films*, 455(456):112–119, 2004.
- [61] O. Debeir, C. Decaestecker, J.L. Pasteels, I. Salmon, R. Kiss, and P. Van Ham. Computer-assisted analysis of epiluminescence microscopy images of pigmented skin lesions. *Cytometry*, 37(4):255–266, 1999.
- [62] A.P. Dhawan, R. Gordon, and R.M. Rangayyan. Nevoscopy: three-dimensional computed tomography of nevi and melanomas in situ by transillumination. *IEEE Trans. Medical Imaging*, 3(2):54–61, 1984.
- [63] Atam P. Dhawan. Early detection of cutaneous malignant melanoma by three-dimensional nevoscopy. *Computer Methods and Programs in Biomedicine*, 21(1):59–68, 1985.
- [64] T.G. Dietterich and G. Bakiri. Solving multiclass learning problems via error-correcting output codes. *Journal of Artificial Intelligence Research*, pages 263–286, 1995.
- [65] H.C. Do and G. Sapiro. Segmenting skin lesions with partial-differential-equations-based image processing algorithms. *IEEE Trans. Medical Imaging*, 19(7):763–767, 2000.
- [66] S. Dreiseitl and M. Binder. Do physicians value decision support? A look at the effect of decision support systems on physician opinion. *Artificial Intelligence in Medicine*, 33(1):25–30, 2005.
- [67] S. Dreiseitl, L. Ohno-Machado, H. Kittler, S. Vinterbo, H. Billhardt, and M. Binder. A comparison of machine learning methods for the diagnosis of pigmented skin lesions. *Journal of Biomedical Informatics*, 34(1):28–36, 2001.

- [68] M. Dubreuil, S. Rivet, B. Le Jeune, et al. Snapshot mueller matrix polarimeter by wavelength polarization coding. *Optics Express*, 15(21):13660–13668, 2007.
- [69] S. Džeroski and B. Ženko. Is combining classifiers with stacking better than selecting the best one? *Machine Learning*, 54(3):255–273, 2004.
- [70] M. Elad. *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*. Incorporated Springer Publishing Company, 1st edition, 2010.
- [71] M. Elad and M. Aharon. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Trans. Image Processing*, 15(12):3736–3745, 2006.
- [72] B. Erkol, R.H. Moss, R. Joe Stanley, W.V. Stoecker, and E. Hvatum. Automatic lesion boundary detection in dermoscopy images using gradient vector flow snakes. *Skin Research Technology*, 11(1):17–26, 2005.
- [73] K.J. Ezawa, M. Singh, and S.W. Norton. Learning goal oriented bayesian networks for telecommunications risk management. In *ICML*, pages 139–147, 1996.
- [74] M. Faal, M. Baygi, M. Hossein, and E. Kabir. Improving the diagnostic accuracy of dysplastic and melanoma lesions using the decision template combination method. *Skin Research Technology*, 19(1):e113–e122, 2013.
- [75] P.M. Ferreira, T. Mendonça, and P. Rocha. A wide spread of algorithms for automatic segmentation of dermoscopic images. In *Pattern Recognition and Image Analysis*, pages 592–599. Springer, 2013.
- [76] F. Ferri, P. Pudil, M. Hatef, and J. Kittler. Comparative study of techniques for large-scale feature selection. *Pattern Recognition in Practice IV*, pages 403–413, 1994.
- [77] A. Fidalgo Barata, E. Celebi, and J. Marques. Improving dermoscopy image classification using color constancy. *IEEE Journal of Biomedical and Health Informatics*, PP(99):1–1, 2014.
- [78] M.G. Fleming, C. Steger, J. Zhang, J. Gao, A.B. Cognetta, L. Pollak, and C.D Dyer. Techniques for a structural analysis of dermatoscopic imagery. *Computerized Medical Imaging and Graphics*, 22(5):375–389, 1998.
- [79] A.M. Forsea, V. Del Marmol, E. de Vries, E.E. Bailey, and A.C. Geller. Melanoma incidence and mortality in europe: new estimates, persistent disparities. *Birtish Journal of Dermatology*, 167(5):1124–1130, 2012.

- [80] Yoav Freund, Robert E Schapire, et al. Experiments with a new boosting algorithm. In *ICML*, volume 96, pages 148–156, 1996.
- [81] R.J. Friedman, D.S. Rigel, and A.W. Kopf. Early detection of malignant melanoma: The role of physician examination and self-examination of the skin. *CA: a cancer journal for clinicians*, 35(3):130–151, 1985.
- [82] J. Fröhlauf, B. Leinweber, R. Fink-Puches, V. Ahlgrimm-Siess, E. Richtig, I. Wolf, et al. Patient acceptance and diagnostic utility of automated digital image analysis of pigmented skin lesions. *Journal of the European Academy of Dermatology and Venereology*, 26(3):368–372, 2012.
- [83] D. Gabor. Theory of communication. part 1: The analysis of information. *Journal of the Institution of Electrical Engineers-Part III: Radio and Communication Engineering*, 93(26):429–441, 1946.
- [84] H. Ganster, P. Pinz, R. Rohrer, E. Wildling, M. Binder, and H. Kittler. Automated melanoma recognition. *IEEE Trans. Medical Imaging*, 20(3):233–239, 2001.
- [85] R. Garnavi, M. Aldeen, and J. Bailey. Computer-aided diagnosis of melanoma using border-and wavelet-based texture analysis. *IEEE Trans. Information Technology in Biomedicine*, 16(6):1239–1252, 2012.
- [86] R. Garnavi, M. Aldeen, M.E. Celebi, G. Varigos, and S. Finch. Border detection in dermoscopy images using hybrid thresholding on optimized color channels. *Computerized Medical Imaging and Graphics*, 35(2):105–115, 2011.
- [87] N. Ghosh, A. Pradhan, P.K. Gupta, S. Gupta, V. Jaiswal, and R.P. Singh. Depolarization of light in a multiply scattering medium: Effect of the refractive index of a scatterer. *Physical Review E*, 70(6):066607, 2004.
- [88] N. Ghosh and I.A. Vitkin. Tissue polarimetry: concepts, challenges, applications, and outlook. *Journal of Biomedical Optics*, 16(11):110801, 2011.
- [89] N. Ghosh, M. Wood, S. Li, R.D. Weisel, B.C. Wilson, R. Li, and I.A. Vitkin. Mueller matrix decomposition for polarized light assessment of biological tissues. *Journal of Biophotonics*, 2(3):145–156, 2009.
- [90] N. Ghosh, M.F.G. Wood, and I.A. Vitkin. Mueller matrix decomposition for extraction of individual polarization parameters from complex turbid media exhibiting multiple scattering, optical activity, and linear birefringence. *Journal of Biomedical Optics*, 13(4):044036–044036, 2008.

- [91] L. Giancardo. *Automated fundus images analysis techniques to screen retinal diseases in diabetic patients*. PhD thesis, Université de Bourgogne, 2011.
- [92] S. Gilmore, R. Hofmann-Wellenhof, and H.P. Soyer. A support vector machine for decision support in melanoma recognition. *Experimental Dermatology*, 19(9):830–835, 2010.
- [93] D.H. Goldstein. Mueller matrix dual-rotating retarder polarimeter. *Applied Optics*, 31(31):6676–6683, 1992.
- [94] D.H. Goldstein. *Polarized Light*. Optical Engineering. Marcel Dekker, 2003.
- [95] D.D. Gómez, C. Butakoff, B.K. Ersboll, and W. Stoecker. Independent histogram pursuit for segmentation of skin lesions. *IEEE Trans. Biomedical Engineering*, 55(1):157–161, 2008.
- [96] Zhenhua Guo and David Zhang. A completed modeling of local binary pattern operator for texture classification. *IEEE Trans. Image Processing*, 19(6):1657–1663, 2010.
- [97] A. Gutenev, VN Skladnev, and D. Varvel. Acquisition-time image quality control in digital dermatoscopy of skin lesions. *Computerized Medical Imaging and Graphics*, 25(6):495–499, 2001.
- [98] D. Gutkowicz-Krusin, M. Elbaum, P. Szwarcowski, and AW Kopf. Can early malignant melanoma be differentiated from atypical melanocytic nevus by in vivo techniques? *Skin Research Technology*, 3(1):15–22, 1997.
- [99] G.A. Hance, S.E. Umbaugh, R.H. Moss, and W.V. Stoecker. Unsupervised color image segmentation: with application to skin tumor borders. *IEEE. Engineering in Medicine and Biology Magazine*, 15(1):104–111, 1996.
- [100] R.M Haralick, K. Shanmugam, and Dinstein. Textural features for image classification. *IEEE Transactions on Systems, Man and Cybernetics*, (6):610–621, 1973.
- [101] H. He, E. Garcia, et al. Learning from imbalanced data. *Knowledge and Data Engineering, IEEE Transactions on*, 21(9):1263–1284, 2009.
- [102] A. Hielscher, A. Eick, J. Mourant, D. Shen, J. Freyer, and I. Bigio. Diffuse backscattering mueller matrices of highly scattering media. *Optics Express*, 1(13):441–453, 1997.
- [103] H. Iyatomi, M.E. Celebi, G. Schaefer, and M. Tanaka. Automated color calibration method for dermoscopy images. *Computerized Medical Imaging and Graphics*, 35(2):89–98, 2011.

- [104] H. Iyatomi, H. Oka, M.E. Celebi, M. Hashimoto, M. Hagiwara, M. Tanaka, and K. Ogawa. An improved internet-based melanoma screening system with dermatologist-like tumor area extraction algorithm. *Computerized Medical Imaging and Graphics*, 32(7):566–579, 2008.
- [105] H. Iyatomi, H. Oka, M. Saito, A. Miyake, M. Kimoto, J. Yamagami, et al. Quantitative assessment of tumour extraction from dermoscopy images and evaluation of computer-based extraction methods for an automatic melanoma diagnostic system. *Melanoma Research*, 16(2):183–190, 2006.
- [106] S.L Jacques, J.C. Ramella-Roman, and K. Lee. Imaging skin pathology with polarized light. *Journal of Biomedical Optics*, 7(3):329–40, 2002.
- [107] G. James and T. Hastie. The error coding method and picts. *Journal of Computational and Graphical Statistics*, 7(3):377–387, 1998.
- [108] R. Jolivot. *Développement d'un outil d'imagerie dédié à l'acquisition, à l'analyse et à la caractérisation multispectrale des lésions dermatologiques*. PhD thesis, Université de Bourgogne, 2011.
- [109] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: active contour models. *International Journal of Computer Vision*, 1(4):321–331, 1988.
- [110] H.L. Kaufman. *The melanoma book: a complete guide to prevention and treatment*. Gotham, 2005.
- [111] Kimia Kiani and Ahmad R. Sharafat. E-shaver: An improved DullRazor® for digitally removing dark and light-colored hairs in dermoscopic images. *Computers in Biology and Medicine*, 41(3):139–145, 2011.
- [112] K. Korotkov and R. Garcia. Computerized analysis of pigmented skin lesions: a review. *Artificial Intelligence in Medicine*, 56(2):69–90, 2012.
- [113] Martin Kreutz, Maik Anschütz, Thorsten Grünendick, Andreas Rick, Stefan Gehlen, and Klaus Hoffmann. Automated diagnosis of skin cancer using digital image processing and mixture-of-experts. *Biomedizinische Technik/Biomedical Engineering*, 46(s1):376–377, 2001.
- [114] Jorma Laurikkala. *Improving identification of difficult small classes by balancing class distribution*. Springer, 2001.

- [115] T. Lee, V. Ng, R. Gallagher, A. Coldman, and D. McLean. DullRazor[®]: A software approach to hair removal from images. *Computers in Biology and Medicine*, 27(6):533–543, 1997.
- [116] G. Lemaître, R. Martí, J. Freixenet, J.C. Vilanova, P.M. Walker, and F. Meriaudeau. Computer-aided detection and diagnosis for prostate cancer based on mono and multi-parametric mri: a review. *Computers in Biology and Medicine*, 60:8–31, 2015.
- [117] C. Li, R. Huang, Z. Ding, J.C Gatenby, D.N. Metaxas, and J.C. Gore. A level set method for image segmentation in the presence of intensity inhomogeneities with application to mri. *IEEE Trans. Image Processing*, 20(7):2007–2016, 2011.
- [118] X. Li and G. Yao. Mueller matrix decomposition of diffuse reflectance imaging in skeletal muscle. *Applied Optics*, 48(14):2625–2631, 2009.
- [119] Z. Liu, J. Sun, M. Smith, L. Smith, and R. Warr. Unsupervised sub-segmentation for pigmented skin lesions. *Skin Research Technology*, 18(2):77–87, 2012.
- [120] David G Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.
- [121] J.Q. Lu, X.H. Hu, and K. Dong. Modeling of the rough-interface effect on a converging light beam propagating in a skin tissue phantom. *Applied Optics*, 39(31):5890–5897, 2000.
- [122] S.Y. Lu and R.A. Chipman. Interpretation of mueller matrices based on polar decomposition. *JOSA A*, 13(5):1106–1113, 1996.
- [123] R.M. MacKie. Cutaneous microscopy in vivo as an aid to preoperative assessment of pigmented lesions of the skin. *British journal of plastic surgery*, 25:123–129, 1972.
- [124] Rona M MacKie. An aid to the preoperative assessment of pigmented lesions of the skin. *Birtish Journal of Dermatology*, 85(3):232–238, 1971.
- [125] I. Maglogiannis and C.N. Doukas. Overview of advanced computer vision systems for skin lesions characterization. *IEEE Trans. Information Technology in Biomedicine*, 13(5):721–733, 2009.
- [126] S.G. Mallat and Z. Zhang. Matching pursuits with time-frequency dictionaries. *IEEE Trans. Signal Processing*, 41(12):3397–3415, 1993.
- [127] S. Manhas, M.K. Swami, H.S. Patel, N. Uppal, A. and Ghosh, and P.K. Gupta. Polarized diffuse reflectance measurements on cancerous and noncancerous tissues. *Journal of Biophotonics*, 2(10):581–587, 2009.

- [128] I. Mani and I. Zhang. knn approach to unbalanced data distributions: a case study involving information extraction. In *Proceedings of Workshop on Learning from Imbalanced Datasets*, 2003.
- [129] B.S. Manjunath and W.Y. Ma. Texture features for browsing and retrieval of image data. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 18(8):837–42, Aug 1996.
- [130] S Marčelja. Mathematical description of the responses of simple cortical cells. *Journal of the Optical Society of America*, 70(11):1297–1300, 1980.
- [131] R. Marchesini, A. Bono, C. Bartoli, M. Lualdi, S. Tomatis, and N. Cascinelli. Optical imaging and automated melanoma detection: questions and answers. *Melanoma Research*, 12(3):279–286, 2002.
- [132] A.A. Marghoob, L.D. Swindle, C. Moricz, F. Negron, B. Slue, A.C. Halpern, and A.W. Kopf. Instruments and new technologies for the in vivo diagnosis of melanoma. *Journal of the American Academy of Dermatology*, 49(5):777–797, 2003.
- [133] Aleix M Martínez and Avinash C Kak. Pca versus lda. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 23(2):228–233, 2001.
- [134] J.A. McGrath and J. Uitto. Anatomy and organization of human skin. *Rook’s Textbook of Dermatology, Eighth Edition*, pages 1–53, 2010.
- [135] T. Mendonça, P.M. Ferreira, J.S. Marques, A.R.S. Marcal, and J. Rozeira. Ph 2-a dermoscopic image database for research and benchmarking. In *Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE*, pages 5437–5440. IEEE, 2013.
- [136] S. Merler, C. Furlanello, B. Larcher, and A. Sboner. Tuning cost-sensitive boosting and its application to melanoma diagnosis. In *Multiple classifier systems*, pages 32–42. Springer, 2001.
- [137] M. Mete, S. Kockara, and K. Aydin. Fast density-based lesion detection in dermoscopy images. *Computerized Medical Imaging and Graphics*, 35(2):128–136, 2011.
- [138] S. Mika, B. Schölkopf, A.J. Smola, K.R. Müller, M. Scholz, and G. Rätsch. Kernel pca and de-noising in feature spaces. In *NIPS*, volume 4, page 7. Citeseer, 1998.
- [139] Kevin P Murphy. *Machine learning: a probabilistic perspective*. MIT press, 2012.
- [140] JA Newton Bishop. Lentigos, melanocytic naevi and melanoma. *Rook’s Textbook of Dermatology, Eighth Edition*, pages 1–57, 2010.

- [141] N.H. Nguyen, T.K. Lee, and M.S. Atkins. Segmentation of light and dark hair in dermoscopic images: a hybrid approach using a universal kernel. In Benoit M. Dawant and David R. Haynor, editors, *Proc. SPIE*, volume 7623 of *Medical Imaging: Image Processing*, San Diego, CA, 2010. SPIE.
- [142] T. Ojala, M. Pietikäinen, and D. Harwood. A comparative study of texture measures with classification based on featured distributions. *Pattern Recognition*, 29(1):51–59, 1996.
- [143] T. Ojala, M. Pietikainen, and T. Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 24(7):971–987, 2002.
- [144] N. Otsu. A threshold selection method from gray-level histograms. *IEEE Trans. Systems, Man and Cybernetics*, 11(285-296):23–27, 1975.
- [145] F. Peng, H. and Long and C. Ding. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 27(8):1226–1238, 2005.
- [146] F. Peruch, F. Bogo, M. Bonazza, Vincenzo-Maria Cappelleri, and E. Peserico. Simpler, faster, more accurate melanocytic lesion segmentation through meds. *IEEE Trans. Biomedical Engineering*, 61(2):557–565, 2014.
- [147] C. Phua, D. Alahakoon, and V. Lee. Minority report in fraud detection: classification of skewed data. *ACM SIGKDD Explorations Newsletter*, 6(1):50–59, 2004.
- [148] S.B. Powell and V. Gruev. Calibration methods for division-of-focal-plane polarimeters. *Optics Express*, 21(18):21039–21055, 2013.
- [149] R.C. Prati, G. EAPA Batista, and M.C. Monard. Data mining with imbalanced class distributions: concepts and methods. In *IICAI*, pages 359–376, 2009.
- [150] J. Premaladha and K.S. Ravichandran. Novel approaches for diagnosing melanoma skin lesions through supervised and deep learning algorithms. *Journal of medical systems*, 40(4):1–12, 2016.
- [151] J. Quintana, R. García, and L. Neumann. A novel method for color correction in epiluminescence microscopy. *Computerized Medical Imaging and Graphics*, 35(7-8):646–652, 2011.
- [152] P. Radivojac, N.V. Chawla, A.K. Dunker, and Z. Obradovic. Classification and knowledge discovery in protein databases. *Journal of Biomedical Informatics*, 37(4):224–239, 2004.

- [153] M.I. Rajab, M.S. Woolfson, and S.P. Morgan. Application of region-based segmentation and neural network edge detection to skin lesions. *Computerized Medical Imaging and Graphics*, 28(1-2):61–68, 2004.
- [154] J.C Ramella-Roman, D. Duncan, and T.A. Germer. Out-of-plane polarimetric imaging of skin: surface and subsurface effects. In *Biomedical Optics 2005*, pages 142–153. International Society for Optics and Photonics, 2005.
- [155] M. Rastgoo, R. Garcia, O. Morel, and F. Marzani. Automated classification of melanoma lesions using polarimetric images. *to be submitted*.
- [156] M. Rastgoo, R. Garcia, O. Morel, and F. Marzani. Pigmented lesions screening using polarimetric dermoscope. *to be submitted*.
- [157] M. Rastgoo, R. Garcia, O. Morel, and F. Marzani. Automatic differentiation of melanoma from dysplastic nevi. *Computerized Medical Imaging and Graphics*, 43:44–52, 2015.
- [158] M. Rastgoo, G. Lemaitre, J. Massich, O. Morel, F. Marzani, R. Garcia, and F. Mériauveau. Tackling the problem of data imbalancing for melanoma classification. In *Bioimaging*, 2016.
- [159] M. Rastgoo, G. Lemaitre, O. Morel, J. Massich, R. Garcia, F. Mériauveau, F. Marzani, and D. Sidibé. Classification of melanoma lesions using sparse coded features and random forests. In *SPIE Medical Imaging*, 2016.
- [160] M. Rastgoo, G. Lemaitre, X.R. Palou, F. Miralles, and P. Casale. Pruning adaboost for continuous sensors mining applications. In *Workshop on Ubiquitous Data Mining*, page 53, 2012.
- [161] M. Rastgoo, O. Morel, F. Marzani, and R. Garcia. Ensemble approach for differentiation of malignant melanoma. In *The International Conference on Quality Control by Artificial Vision 2015*, pages 953415–953415. International Society for Optics and Photonics, 2015.
- [162] D.S. Rigel, J. Russak, and R. Friedman. The evolution of melanoma diagnosis: 25 years beyond the abcds. *CA: a cancer journal for clinicians*, 60(5):301–316, 2010.
- [163] Lior Rokach. Ensemble-based classifiers. *Artificial Intelligence Review*, 33(1-2):1–39, 2010.
- [164] Pietro Rubegni, Gabriele Cevenini, Marco Burroni, Roberto Perotti, Giordana Dell’Eva, Paolo Sbano, Clelia Miracco, Pietro Luzi, Piero Tosi, Paolo Barbini, et al. Automated diagnosis of pigmented skin lesions. *International Journal of Cancer*, 101(6):576–580, 2002.

- [165] R. Rubinstein, M. Zibulevsky, and M. Elad. Efficient implementation of the k-svd algorithm using batch orthogonal matching pursuit. *CS Technion*, 40(8):1–15, 2008.
- [166] M. Ruela, C. Barata, and J.S. Marques. What is the role of color symmetry in the detection of melanomas? In *Advances in Visual Computing*, pages 1–10. Springer, 2013.
- [167] M. Ruela, C. Barata, T. Mendonca, and J.S. Marques. On the role of shape in the detection of melanomas. In *8th International Symposium on Image and Signal Processing and Analysis (ISPA)*, pages 268–273. IEEE, 2013.
- [168] D. Ruiz, V. Berenguer, A. Soriano, and B. Sánchez. A decision support system for the diagnosis of melanoma: A comparative approach. *Expert Systems with Applications*, 38(12):15217–23, 2011.
- [169] V. Sankaran, K. Schönenberger, J. Walsh, and D.J. Maitland. Polarization discrimination of coherently propagating light in turbid media. *Applied Optics*, 38(19):4252–4261, 1999.
- [170] V. Sankaran, J. Walsh, and D.J. Maitland. Comparative study of polarized light propagation in biologic tissues. *Journal of biomedical optics*, 7(3):300–306, 2002.
- [171] A. Sboner, C. Eccher, E. Blanzieri, P. Bauer, M. Cristofolini, G. Zumiani, and S. Forti. A multiple classifier system for early melanoma diagnosis. *Artificial Intelligence in Medicine*, 27(1):29–44, 2003.
- [172] G. Schaefer, B. Krawczyk, M.E. Celebi, and H. Iyatomi. An ensemble classification approach for melanoma diagnosis. *Memetic Computing*, 6(4):233–240, 2014.
- [173] G. Schaefer, M.I. Rajab, M.E. Celebi, and H. Iyatomi. Colour and contrast enhancement for improved skin lesion segmentation. *Computerized Medical Imaging and Graphics*, 35(2):99–104, 2011.
- [174] P. Schmid. Segmentation of digitized dermatoscopic images by two-dimensional color clustering. *IEEE Trans. Medical Imaging*, 18(2):164–171, 1999.
- [175] P. Schmid-Saugeon, J. Guillo, and J.P. Thiran. Towards a computer-aided diagnosis system for pigmented skin lesions. *Computerized Medical Imaging and Graphics*, 27(1):65–78, 2003.
- [176] K. Shimizu, H. Iyatomi, M.E. Celebi, K. Norton, and M. Tanaka. Four-class classification of skin lesions with task decomposition strategy. *IEEE Trans. Biomedical Engineering*, 62(1):274–283, Jan 2015.

- [177] D. Sidibé, I. Sadek, and F. Mériauadeau. Discrimination of retinal images containing bright lesions using sparse coded features and svm. *Computers in biology and medicine*, 62:175–184, 2015.
- [178] M. Silveira, Jacinto C Nascimento, Jorge S Marques, André RS Marçal, Teresa Mendonça, Syogo Yamauchi, Junji Maeda, and Jorge Rozeira. Comparison of segmentation methods for melanoma diagnosis in dermoscopy images. *IEEE Journal of Selected Topics in Signal Processing*, 3(1):35–45, 2009.
- [179] M. Silveira, J.C. Nascimento, J.S. Marques, A.R.S. Marçal, T. Mendonça, S. Yamauchi, et al. Comparison of segmentation methods for melanoma diagnosis in dermoscopy images. *IEEE Journal of Selected Topics in Signal Processing*, 3(1):35–45, 2009.
- [180] Ning Situ, T Wadhawan, Xiaojing Yuan, and G Zouridakis. Modeling spatial relation in skin lesion images by the graph walk kernel. In *International Conference on the IEEE Engineering in Medicine and Biology Society*, pages 6130–6133. IEEE, 2010.
- [181] Ning Situ, Xiaojing Yuan, Ji Chen, and George Zouridakis. Malignant melanoma detection by bag-of-features classification. In *30th Annual International Conference of the IEEE, Engineering in Medicine and Biology Society*, pages 3110–3113. IEEE, 2008.
- [182] Matthew H Smith. Optimization of a dual-rotating-retarder mueller matrix polarimeter. *Applied Optics*, 41(13):2488–2493, 2002.
- [183] American Cancer Society. Cancer facts & figures 2014, 2014.
- [184] L-K Soh and Costas Tsatsoulis. Texture analysis of sar sea ice imagery using gray level co-occurrence matrices. *IEEE Trans. Geoscience and Remote Sensing*, 37(2):780–795, 1999.
- [185] Sotiris K Tasoulis, Charalampos N Doukas, Ilias Maglogiannis, and Vassilis P Plagianakos. Skin lesions characterisation utilising clustering algorithms. In *Artificial Intelligence: Theories, Models and Applications*, pages 243–253. Springer, 2010.
- [186] Lioudmila Tchvialeva, Gurbir Dhadwal, Harvey Lui, Sunil Kalia, Haishan Zeng, David I McLean, and Tim K Lee. Polarization speckle imaging as a potential technique for in vivo skin cancer detection. *Journal of Biomedical Optics*, 18(6):061211–061211, 2013.
- [187] Sergios Theodoridis and Konstantinos Koutroumbas. Chapter 7 - feature generation {II}. In Sergios TheodoridisKonstantinos Koutroumbas, editor, *Pattern Recognition (Third Edition)*, pages 327 – 396. Academic Press, San Diego, third edition edition, 2006.

- [188] Ivan Tomek. Two modifications of cnn. *IEEE Trans. Systems, Man and Cybernetics*, 6:769–772, 1976.
- [189] Elisabetta La Torre, Barbara Caputo, and Tatiana Tommasi. Learning methods for melanoma recognition. *International Journal of Imaging Systems and Technology*, 20(4):316–322, 2010.
- [190] V. V. Tuchin, L. V. Wang, and D. A. Zimnyakov. *Optical Polarization in Biomedical Applications*. 2006.
- [191] KM Twietmeyer, RA Chipman, Ann E Elsner, Y Zhao, and D VanNasdale. Mueller matrix retinal imager with optimized polarization conditions. *Optics Express*, 16(26):21339–21354, 2008.
- [192] Joost Van De Weijer and Cordelia Schmid. Coloring local feature extraction. In *Computer Vision–ECCV 2006*, pages 334–348. Springer, 2006.
- [193] V Vapnik and AJ Lerner. Generalized portrait method for pattern recognition. *Automation and Remote Control*, 24(6):774–780, 1963.
- [194] M.E. Vestergaard and S.W. Menzies. Automated diagnostic instruments for cutaneous melanoma. *Seminars in cutaneous medicine and surgery*, 27(1):32–36, 2008.
- [195] Luc Vincent. Morphological grayscale reconstruction in image analysis: applications and efficient algorithms. *IEEE Trans. Image Processing*, 2(2):176–201, 1993.
- [196] Steven Q Wang and Pantea Hashemi. Noninvasive imaging technologies in the diagnosis of melanoma. In *Seminars in cutaneous medicine and surgery*, volume 29, pages 174–184. WB Saunders, 2010.
- [197] P. Wighton, T. Lee, H. Lui, D. McLean, and M. Atkins. Generalizing common tasks in automated skin lesion diagnosis. *IEEE Trans. Information Technology in Biomedicine*, 4:622–629, 2011.
- [198] P. Wighton, T.K. Lee, and M.S. Atkins. Dermoscopic hair disocclusion using inpainting. In Joseph M. Reinhardt and Josien P. W. Pluim, editors, *Proc. SPIE*, volume 6914 of *Medical Imaging: Image Processing*, San Diego, CA, 2008. SPIE.
- [199] P. Wighton, T.K. Lee, G. Mori, H. Lui, D.I. McLean, and M.S. Atkins. Conditional random fields and supervised learning in automated skin lesion diagnosis. *International Journal of Biomedical Imaging*, 2011, 10 pages, 2011.

- [200] Paul Wighton, Tim K. Lee, Harvey Lui, David McLean, and M. Stella Atkins. Chromatic aberration correction: an enhancement to the calibration of low-cost digital dermoscopes. *Skin Research Technology*, 17(3):339–347, 2011.
- [201] Dennis L Wilson. Asymptotic properties of nearest neighbor rules using edited data. *IEEE Trans. Systems, Man and Cybernetics*, (3):408–421, 1972.
- [202] David H Wolpert. Stacked generalization. *Neural networks*, 5(2):241–259, 1992.
- [203] Michael FG Wood, Nirmalya Ghosh, Xinxin Guo, and I Alex Vitkin. Towards noninvasive glucose sensing using polarization analysis of multiply scattered light, 2008.
- [204] John Wright, Allen Y Yang, Arvind Ganesh, Shankar S Sastry, and Yi Ma. Robust face recognition via sparse representation. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 31(2):210–227, 2009.
- [205] Feng-Ying Xie, Shi-Yin Qin, Zhi-Guo Jiang, and Ru-Song Meng. PDE-based unsupervised repair of hair-occluded information in dermoscopy images of melanoma. *Computerized Medical Imaging and Graphics*, 33(4):275–282, 2009.
- [206] Xiaojing Yuan, Ning Situ, and George Zouridakis. A narrow band graph partitioning method for skin lesion segmentation. *Pattern Recognition*, 42(6):1017–28, 2009.
- [207] M. Emin Yüksel and Murat Borlu. Accurate segmentation of dermoscopic images by image thresholding based on type-2 fuzzy logic. *IEEE Trans. Fuzzy Systems*, 17:976–982, August 2009.
- [208] Frederic Zana and Jean-Claude Klein. Segmentation of vessel-like patterns using mathematical morphology and curvature evaluation. *IEEE Trans. Image Processing*, 10(7):1010–1019, 2001.
- [209] Y. Zhao, L. Zhang, and Q. Pan. Spectropolarimetric imaging for pathological analysis of skin. *Applied Optics*, 48(10):D236–D246, 2009.
- [210] Zhaohui Zheng, Hongyuan Zha, Tong Zhang, Olivier Chapelle, Keke Chen, and Gordon Sun. A general boosting method and its application to learning ranking functions for web search. In *Advances in neural information processing systems*, pages 1697–1704, 2007.
- [211] H. Zhou, M. Chen, R. Gass, J.M. Rehg, L. Ferris, J. Ho, and L. Drogowski. Feature-preserving artifact removal from dermoscopy images. In Joseph M. Reinhardt and Josien P. W. Pluim, editors, *Proc. SPIE*, volume 6914 of *Medical Imaging: Image Processing*, San Diego, CA, 2008. SPIE.

- [212] H. Zhou, G. Schaefer, A.H. Sadka, and M.E. Celebi. Anisotropic mean shift based fuzzy c-means segmentation of dermoscopy images. *IEEE Journal of Selected Topics in Signal Processing*, 3(1):26–34, 2009.
- [213] Huiyu Zhou, Gerald Schaefer, M. Emre Celebi, Faquan Lin, and Tangwei Liu. Gradient vector flow with mean shift for skin lesion segmentation. *Computerized Medical Imaging and Graphics*, 35(2):121–127, 2011.
- [214] Barbara Zitova and Jan Flusser. Image registration methods: a survey. *Image and Vision Computing*, 21(11):977–1000, 2003.
- [215] Maciel Zortea, Thomas R Schopf, Kevin Thon, Marc Geilhufe, Kristian Hindberg, Herbert Kirchesch, Kajsa Møllersen, Jörn Schulz, Stein Olav Skrøvseth, and Fred Godtliebsen. Performance of a dermoscopy-based computer vision system for the diagnosis of pigmented skin lesions compared with visual evaluation by experienced dermatologists. *Artificial Intelligence in Medicine*, 60(1):13–26, 2014.
- [216] Maciel Zortea, SO Skrovseth, and Fred Godtliebsen. Automatic learning of spatial patterns for diagnosis of skin lesions. In *Annual International Conference of the IEEE, Engineering in Medicine and Biology Society*, pages 5601–5604. IEEE, 2010.
- [217] Maciel Zortea, Stein Olav Skrøvseth, Thomas Schopf, Herbert Kirchesch, and Fred Godtliebsen. Automatic segmentation of dermoscopic images by iterative classification, 2011.
- [218] Mark H Zweig and Gregory Campbell. Receiver-operating characteristic (roc) plots: a fundamental evaluation tool in clinical medicine. *Clinical Chemistry*, 39(4):561–577, 1993.