



Universidad
Carlos III de Madrid

Bachelor Thesis

Computer aided diagnosis system using dermatoscopical image

Javier López Labraca

Bachelor degree in audiovisual systems engineering

Tutor:

Fernando Díaz de María

ABSTRACT

Computer Aided Diagnosis (CAD) systems for melanoma detection aim to mirror the expert dermatologist decision when watching a dermoscopic or clinical image. Computer Vision techniques, which can be based on expert knowledge or not, are used to characterize the lesion image. This information is delivered to a machine learning algorithm, which gives a diagnosis suggestion as an output.

This research is included into this field, and addresses the objective of implementing a complete CAD system using 'state of the art' descriptors and dermoscopy images as input. Some of them are based on expert knowledge and others are typical in a wide variety of problems. Images are initially transformed into oRGB, a perceptual color space, looking for both enhancing the information that images provide and giving human perception to machine algorithms. Feature selection is also performed to find features that really contribute to discriminate between benign and malignant pigmented skin lesions (PSL). The problem of robust model fitting versus statistically significant system evaluation is critical when working with small datasets, which is indeed the case. This topic is not generally considered in works related to PSLs. Consequently, a method that optimizes the compromise between these two goals is proposed, giving non-overfitted models and statistically significant measures of performance. In this manner, different systems can be compared in a fairer way. A database which enjoys wide international acceptance among dermatologists is used for the experiments.

ACKNOWLEDGEMENTS

I would like to thank my tutor, Dr. Fernando Díaz, for his permanent support and valuable advices throughout this thesis.

I would also like to thank all the members of the Multimedia Processing Group, for their helpful assistance and for making my learning so pleasant with their explanations.

Last, but definitely not the least, I would like to thank my family and friends for their unconditional encouragement and empathy.

INDEX

INDEX OF FIGURES	VI
INDEX OF TABLES	VII
INDEX OF ALGORITHMS	VIII
INDEXES OF EQUATIONS	VIII
1. INTRODUCTION AND PROJECT OBJECTIVES	1
2. REGULATORY FRAMEWORK	2
3. PLANNING	3
4. ECONOMIC ENVIRONMENT	5
4.1. FINANCES	5
5. SOME NOTIONS OF MELANOMA DIAGNOSIS AND COMPUTER AIDED DIAGNOSIS SYSTEMS.....	7
5.1. THE HUMAN SKIN.....	7
5.2. PIGMENTED SKIN LESIONS.....	8
5.3. MELANOMA AND OTHER TYPES OF SKIN CANCER	9
5.4. MELANOMA SCREENING AND IMAGING TECHNIQUES.....	10
5.4.1. Clinical images.....	10
5.4.2. Dermoscopy	10
5.5. METHODS OF MELANOMA DIAGNOSIS	11
5.6. COMPUTER AIDED DIAGNOSIS (CAD) SYSTEMS FOR MELANOMA DETECTION	13
6. STATE OF THE ART OF COMPUTER-AIDED MELANOMA DIAGNOSIS	15
6.1. LESION SEGMENTATION AND IMAGE PREPROCESSING	15
6.2. FEATURE EXTRACTION	16
6.2.1. Features for lesion description	16
6.2.2. Pattern-oriented features	17
6.3. CLASSIFICATION	17
6.4. COMPLETE SYSTEMS	19
6.4.1. Automatic versions of dermoscopic diagnosis methods.....	19
6.4.2. Commercial CAD systems.....	19
6.5. IMAGE REGISTRATION SYSTEMS	20
7. SYSTEM DESCRIPTION	22
7.1. SEGMENTATION.....	22
7.2. FEATURE EXTRACTION.....	22
7.2.1. Color space description and justification	23
7.2.2. Entropy.....	24
7.2.3. Variance of the contour gradient.....	25
7.2.4. Area	26
7.2.5. Compacity	27
7.2.6. Asymmetry	27
7.2.7. Ellipsoidality	28
7.2.8. Histogram modeling features	29
7.3. FEATURE SELECTION AND CLASSIFICATION	29
7.3.1. Support Vector Machine	29
7.3.2. Sequential Forward Selection	33

7.4.	DIAGNOSIS SUGGESTION	34
8.	SYSTEM EVALUATION TECHNIQUES FOR SMALL DATASETS.....	35
8.1.	SMALL DATASETS PROBLEM INSIGHTS	35
8.2.	PROPOSED TRAINING AND EVALUATION PROCEDURES	36
8.3.	EVALUATION MEASURE FOR THE PERFORMANCE OF THE CLASSIFIER	38
8.3.1.	Useful ratio definitions.....	38
8.3.2.	Receiving Operating Characteristic curve	39
9.	EXPERIMENTAL RESULTS	41
9.1.	SCOPE OF THE STUDY	41
9.2.	DATABASE DESCRIPTION AND PREPARATION	41
9.3.	SYSTEM PERFORMANCE EVALUATION	42
9.4.	FEATURE SELECTION ANALYSIS	44
10.	CONCLUSIONS AND FUTURE WORK.....	47
	REFERENCES.....	48
	APPENDIX I: MELANOMA DIAGNOSIS ALGORITHMS.....	56
	APPENDIX II: TASKS BREAKDOWN	61

Index of figures

Figure 3.1: Gantt chart.	3
Figure 5.1: The human skin.	7
Figure 5.2: Layers of the epidermis.	8
Figure 5.3: Squamous cell carcinoma, basal cell carcinoma, melanoma.	10
Figure 5.4: Clinical image (a), non-polarized light contact dermoscopy (b), polarized light contact dermoscopy (c), polarized light non-contact dermoscopy (d).	11
Figure 7.2: Manual segmentation of a nevus.	22
Figure 7.1: Block diagram of a generic CAD system.	22
Figure 7.3: Color division according to their suggested temperature. A color is getting temperate when it approximates the drawn diameter.	23
Figure 7.4: of the gamut or oRGB chromatic components for a certain luminance value, before (left) and after (right) uniform rotation to obtain orthogonal axes of opponent colors.	24
Figure 7.5: Color component of a nevus image (left) and its histogram (right).	24
Figure 7.6: Segments modeling the direction of lesion-skin gradient (left). Pixel samples from a segment and estimated line to model transition sharpness (right).	25
Figure 7.8: Compacity descriptor is based on the common area between the ROI of the lesion and the equivalent circle.	27
Figure 7.7: Comparison of lesion to skin transition histograms between a nevus with regular border (left) and another with more irregular contour (right).	26
Figure 7.9: On the left, a lesion divided into two halves by one rotating axis. On the right, the half below overlapped over the half above, with the intersection highlighted as the green-colored region.	28
Figure 7.10: Ellipsoidality. Area inside the equivalent ellipse contributes to ellipsoidality.	28
Figure 7.11: Upper left: Blue-Yellow chromatic component. Upper right: Red-Green chromatic component. Bottom: Intensity component.	29
Figure 7.12: Influence of outliers over global margin when there is no regularization term.	30
Figure 7.13: Space transformation using a convenient kernel.	31
Figure 7.14: Decision boundary for a fixed gamma (0.1, 10 and 100, respectively) and its corresponding C term at its optimal value. The region for class decision is red or blue colored.	32
Figure 7.15: Performance of the classifier as a function of C and γ	33
Figure 8.1: Proposed cross-validation method.	37
Figure 8.2: ROC curve registers all possible TPR-FPR for all possible values of the threshold θ . The equal error rate diagonal provides an operating point in which FPR and FNR are equal.	39
Figure 9.2: Frequency histogram (left) and weighted frequency histogram(right) for feature set 1.	45
Figure 9.3: Frequency histogram (left) and weighted frequency histogram(right) for feature set 2.	45

Index of tables

Table 3.1: Planning summary	4
Table 4.1: Staff costs for each project phase.	5
Table 4.2: Costs of amortizable material.....	5
Table 4.3: Costs of perishable goods.	6
Table 4.4: Summary of costs.	6
Table 4.5: Final budget.	6
Table 5.1: Common benign lesions description.	9
Table 5.2: Classification of diagnosis algorithms according to the type of images considered.	13
Table 5.3: Comparison between the ABCD rules for clinical and dermoscopic images.	13
Table 5.4: Comparison between the 7-point checklist methods for clinical and dermoscopic images.	13
Table 6.1: References for different descriptor classes.	17
Table 6.2: Types of classifiers and examples of use.	18
Table 6.3: Summary of commercial CAD systems.	20
Table 9.1: Initial feature sets.	43
Table 9.2: Occurrence probabilities of some interesting groups of features.	46

Index of algorithms

Algorithm 7.1: C and γ grid search validation algorithm.	32
Algorithm 7.2: Training and test algorithm to use a SVM RBF classifier.	33
Algorithm 7.3: Sequential Forward Selection.	34
Algorithm 8.1: Double K-fold cross-validation nested loop algorithm.	38

Indes of equations

Equation 7.1: Transformation matrix of the $R'G'B'$ device dependent cube into the $L'C1'C2'$	24
Equation 7.2: Shannon's classic Entropy formula.....	25
Equation 7.3: Lineal model to fit.	26
Equation 7.4: Slope formula.	26
Equation 7.5: Independent term formula.	26
Equation 7.6: Centroid computation.	27
Equation 7.7: Compacity formula.....	27
Equation 7.8: Symmetry formula.	28
Equation 7.9: Asymmetry computation.	28
Equation 7.10: Ellipsoidality formula.....	28
Equation 7.11: Expression to minimize on SVMs.	30
Equation 7.12: Restrictions based on data.	30
Equation 7.13: Restriction for regularization.	30

1. Introduction and project objectives

Melanoma is the leading cause of death among skin diseases, being responsible for the 75% of deaths associated with this organ. According to the World Health Organization (WHO), there are about 48.000 deaths related to this disease each year. The number of diagnosed cases is also increasing in white people living in places with sunny climate. Its aggressive nature creates the need of detecting it at its early stages, when there are high expectations to cure it. Dermoscopy has revealed as a useful imaging technique to accomplish this objective when a dermatologist is well trained.

The introduction of digital systems has provided an easy way of getting and store images, making easier the routine of the dermatologist. This also increases the availability of digital images, which leads Computer Vision techniques to appear in dermatology, as in some other medical fields.

At first, there was an initial tendency of developing image segmentation techniques for dermoscopy images. Later, common recommendations from dermatologists started to be replicated into automated systems, using common and dermoscopic images. This started the developing of Computer-Aided Diagnosis (CAD) systems, which tried to objectify rules used by dermatologists and to apply classical imaging techniques to help the diagnosis of the clinician. Today, there is a large set of commercial systems that offers this possibility and mobile applications like Dermaskin have started to diffuse among people.

This project aims to design a 'state of the art' CAD system which uses dermoscopy images as input. Descriptors will be developed, in order to get the relevant information from images and turning them into feature vectors. This strategy allows the use of Machine Learning techniques to allow the implementation of an automatic system that makes diagnosis suggestions to the images it receives.

Small databases are usual in this field, so the problem of experimentation with small datasets will be deeply addressed. Concretely, a new cross-validation procedure which leads getting correct evaluation measures and the training of robust systems is described. A technique of feature selection is also implemented, in order to get information about what features are really significant among all that can be tried.

2. Regulatory framework

When working with medical images, it is mandatory to respect the privacy rights of the patients. The construction of a skin lesions database requires a signed consent for its publication. This issue is not treated in this project, because the database was commercial, so coming with these problems solved.

The other fact that needs to be clarified is the use of software. Almost all tools used in this project are free software, so they can be obtained by any user through the internet. However, all the expensive computations needed by experiments required computation software and Matlab was chosen. Microsoft Office is used for the making of this memory. So, these two tools are commercial and require the use of a license for its use.

3. Planning

The tasks involved in the project were gathered in three phases. The temporary distribution of tasks needed to make a whole system is shown in [FIGURE 3.1](#) (See [APPENDIX II: Tasks breakdown](#) for a connection between ID numbers of the figure and tasks). Due to the nature of research projects, some schedule variations are expected. A workday of 4 daily hours from Monday to Friday, so as 2 daily hours at the weekend, have been established.

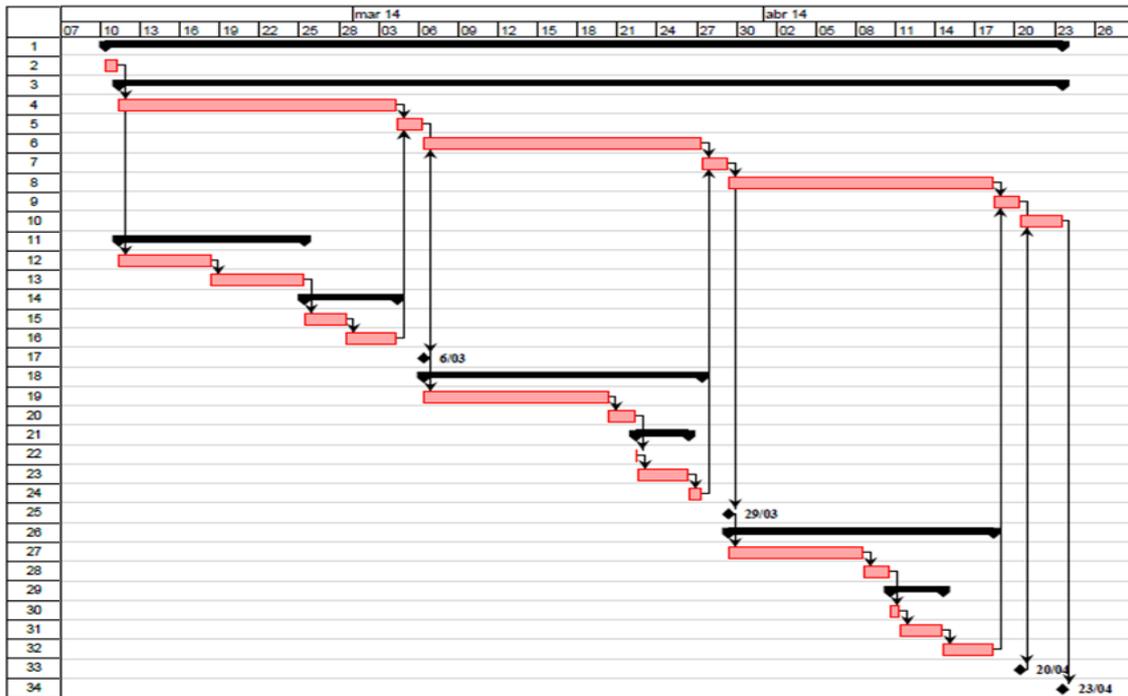


Figure 3.1: Gantt chart.

The first phase includes two conceptually different groups of tasks: the previous study and the subsequent preparation of the database for the later experimentation phase. The former starts with some background knowledge on dermatology of skin lesions and dermoscopy with the aim of understanding the medical viewpoint and getting insights about what makes benign lesions dermoscopically different from malignant lesions. It is followed by an initial contact with the state of the art of similar systems, including the analysis of overview articles and a first group of proposed systems and descriptors. The latter group starts with the selection of the images decided to be included in the study, the extraction of manual masks and the analysis of all useful metadata included with them. This phase ends with the definition of the architecture of the prototype system, at *Milestone 1*, estimated on March, the 6th.

The second phase covers the implementation of each of the blocks of the prototype system and the first experiments. This involves coding an initial set of descriptors, a classifier and algorithms for feature selection and parameter validation. After an evaluation to check the correct operation of the prototype, the first set of experiments is launched. The analysis of its performance gives the needed feedback to deal with the decision of the components of the final system, programmed to be made on March, the 29th.

3. Planning

The final system implementation is addressed at the third phase. The analysis of a new group of state of the art articles and personal reflection itself should have offered a new set of image descriptors at this point. New improvement proposals of the implemented algorithms are coded and validated. Experimentation and results analysis end on April, the 20th. This milestone is set as an estimation of the date to finish the first whole system. Future improvements of the system follow the same sequence of this phase, with a conclusive experimentation finishing on June, the 10th, when the final revision of the project report is programmed.

The documentation tasks not included above occur in parallel with each of the described phases and contain the three scheduled phases of project report composition and revision, so as further extensions to the study of dermoscopy, image descriptors and similar systems. They converge with implementation tasks in a revision process which ends at the milestone and starts the next phase.

The summary of deadlines and work hours can be viewed in [TABLE 3.1](#). See [APPENDIX II](#): Tasks breakdown for a detailed description of the hours that involved each task.

Start date	10-February
End date	10-June
Work hours	364

Table 3.1: Planning summary

4. Economic environment

Computer systems are gaining importance in the dermatological field. At present, its influence can be found on mobile applications which analyze the potential malignancy of a mole through common images, software tools to help the routine tasks of a dermatologist and even commercial automated diagnosis systems to help non-experts (See section 5.6). As our system aims to help the clinician with a diagnosis suggestion, its potential usefulness leads to develop the budget of the project, in order to value the study.

4.1. Finances

The staff costs have been calculated based on the internship period in which the project was developed. Assuming a salary of 543.91 €/month with 80 work hours per month, the expense per hour is fixed at 6.80 €. The costs per phase are presented in [TABLE 4.1](#).

Phase	Work hours	Cost (€)
Planning	8	54,39 €
Initial study	36	244,76 €
Prototype development	54	367,14 €
Complete system development	52,5	356,94 €
System improvement	136	924,65 €
Documentation	77,5	526,91 €
Total	364	2.474,79 €

Table 4.1: Staff costs for each project phase.

The amortizable materials needed for the execution of the project were classified in hardware and software categories. The considered amortization period is 5 years for hardware and 10 years for software. See [TABLE 4.2](#) for details.

Type	Element	Cost (€)	Amortization period (months)	Amortization (€/month)	Use period (months)	Cost for the project (€)
Hardware	HP Pavilion dv6 Notebook PC	699,00 €	60	11,65 €	4	46,60 €
	Computer at workspace	1.200,00 €	60	20,00 €	4	80,00 €
Software	Windows 8	- €	120	- €	4	- €
	Office 2013 Professional	539,99 €	120	4,50 €	4	18,00 €
	Open Project	- €	120	- €	4	- €
	LiBSVM	- €	120	- €	4	- €
	Fedora	- €	120	- €	4	- €
	Matlab 7.9	2.000,00 €	120	16,67 €	4	66,67 €
Total						211,27 €

Table 4.2: Costs of amortizable material.

The last direct costs taken into account have been perishable goods, which are specified in [TABLE 4.3](#).

Element	Units	Prize/unit	Cost
Paper 500 sheets	1	4,99 €	4,99 €
Pen BIC Crystal Blue	10	0,25 €	2,45 €
Total			7,44 €

Table 4.3: Costs of perishable goods.

The workspace rent covers light, office furniture, maintenance... It costs 1000 € per month, which is shared among 7 people and used during 4 months for the project, leading to 572 €. This analysis is assembled to give a summary of direct and indirect costs. See [TABLE 4.4](#):

Type	Concept	Cost
Direct costs	Staff	2.474,79 €
	Amortizable equipment	211,27 €
	Perishable equipment	7,44 €
	Total	2.693,50 €
Indirect costs	ADSL line	34,29 €
	Workspace rent	572,00 €
	Total	606,29 €

Table 4.4: Summary of costs.

After addressing the costs for the project, the budget is completed with the inclusion of earnings, considering a profit margin of 10%, and the VAT tax (See [TABLE 4.5](#)). The value of the project is then finally established on **3.604,38€**.

Concept	Cost
Direct costs	2.693,50 €
Indirect costs	606,29 €
Total costs	3.299,79 €
Earnings	263,98 €
Total budget	3.563,77 €
Total budget + VAT	4.312,16 €

Table 4.5: Final budget.

5. Some notions of melanoma diagnosis and computer aided diagnosis systems

In this section, an introduction to the clinical view of the problem of the melanoma diagnosis along with an explanation of different image acquisition techniques and the role of Computer Aided Diagnosis (CAD) systems are made. It starts with an introduction to the human skin, followed by a description of the skin lesions covered in this study (types of skin cancer and some others). Then, it follows a description of the different image data available. Also, some diagnosis algorithms used by dermatologists are commented, because of the valuable insights that they provide for feature extraction. Finally, this section ends with an overview of the role of CAD systems and the evaluation of it by clinicians.

5.1. The human skin

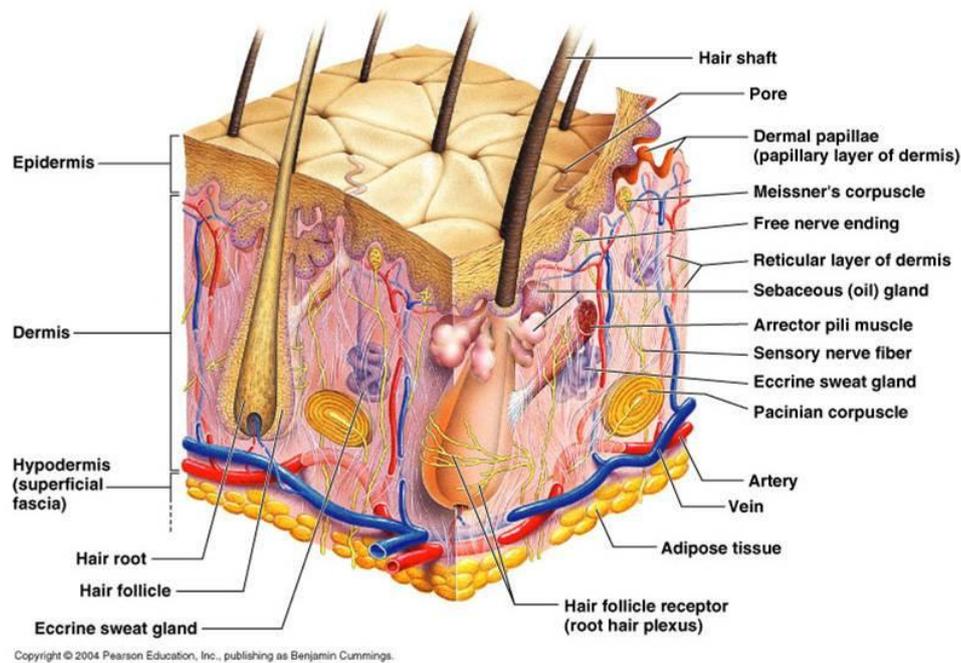


Figure 5.1: The human skin. [113].

The human skin is the outer covering of the body and its largest organ. The skin manages the contact with the environment, so it is responsible of protection against external damage, sensations, heat regulation and water loss control. It is globally divided into two layers: epidermis and dermis. See [FIGURE 5.1](#) to check each part described next.

The epidermis is made up of stratified squamous epithelium, a scale-like tissue arranged in layers, which carries out protection functions against infections, injuries and ultraviolet radiation. Four types of cells can be recognized: keratinocytes, melanocytes, Langerhans cells, and Merkel cells.

Keratinocytes are present at all stratum of the epidermis and are filled with keratin. In their lifecycle, they undertake a journey that lasts around 30 days, changing their morphology and

5. Some notions of melanoma diagnosis and computer aided diagnosis systems

biochemical properties on each epidermis layer through a cell process called differentiation. This trip starts at the basal layer, where are named basal cells, and ends at the stratum corneum, the inner and outer epidermis layers, respectively. Due to this differentiation process the cells are flattened and lose their nuclei when they reach the stratum corneum, named corneocytes, and finally lose their cohesion to the epidermis in the desquamation process.

Melanocytes are found at the basal layer of the epidermis and they are responsible for distributing melanin to keratinocytes, which gives its color to skin and hair.

Langerhans cells are found in all layers of the epidermis and are in charge of capturing antigens (foreign bodies) and delivering them to the lymph nodes.

Merkel cells are associated with the sense of touch and act as sensors.

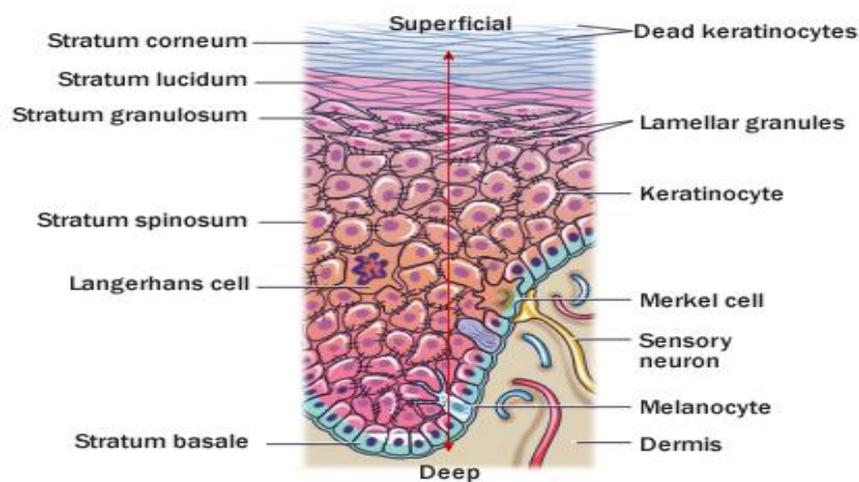


Figure 5.2: Layers of the epidermis. [114].

The dermis is the inner layer of skin, arranged between the epidermis and subcutaneous tissue. It consists of collagen and elastic fibers, and cushions the body from stress and strain. This skin stratum contains two sub-layers: the papillary dermis and the reticular dermis. The first is the union with the epidermis. The latter contains blood, lymph vessels, nerve endings, sweat glands and hair follicles. It provides energy and nutrition and has important functions in thermoregulation, healing and the sense of touch.

5.2. Pigmented skin lesions

Pigmented skin lesions (PSL) materialize when melanocytes grow in clusters among other normal surrounding skin cells. They are also known as moles or nevi. In a very high percent they are a normal part of the skin, however, they are also closely related to melanomas. A general description of the most common benign lesions follows. For completeness, [1] can be consulted.

5. Some notions of melanoma diagnosis and computer aided diagnosis systems

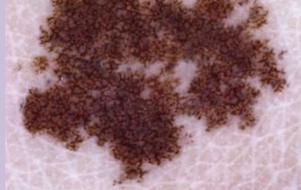
<p>Clark nevi: these are the most common nevi and they encompass all flattened common moles. These nevi contain various brown shades and they are often found in skin that has been exposed to sunlight, but not always. Atypical ones should be excised or followed-up because of their potential risk of developing melanoma.</p>	
<p>Blue nevi: these nevi present homogeneous brownish-blue pigmentation, sometimes gray-blue or grey-black. They are also relatively regular and their transition to normal skin is sharp. These features make them easy to detect clinically. Malignancy is rare, but difficult to separate from benign cases.</p>	
<p>Congenital nevi: these pigmented lesions are present at birth or rise during the first months of life. They are precursor of melanoma with an estimated risk from 5% to 10%. Pigmentation varies between light-brown and dark-brown and they may be elevated or not. The variety of sizes is huge.</p>	
<p>Dysplastic nevi: they are acquired pigmented lesions of the skin. These lesions are usually bigger than common moles and have greater shape irregularity, heterogeneous coloration and blurry borders may appear. The previous features are also characteristic of melanoma, theoretically less noticeable, but they frequently make the lesion at least suspicious.</p>	
<p>Spitz nevi: these nevi are more common in the first two decades of life. Variable distinctive patterns can be found, such as well defined borders, small black to brown or reddish elevations or verrucous plaques. For this reason they are often difficult to distinguish from melanoma only with information from the naked eye.</p>	

Table 5.1: Common benign lesions description.

It can be seen that although there are pigmented lesions more potentially malignant than others, only a careful examination helps the detection of malignant moles which are perceptually very similar to benign ones. Apart from this, melanoma may appear either in a previously existing skin lesion or as a new one. Moreover, changes in size and color are very informative to detect early melanoma [2]. This makes regular skin screening a valuable tool for an early detection of melanoma.

5.3. Melanoma and other types of skin cancer

All cells have a little possibility to develop cancer, but there are types more prone to than others. Specifically, the most frequent skin cancer is developed by keratinocytes from the stratum basale (basal cells) and the stratum corneum (squamous cells), called basal cell carcinoma and squamous cell carcinoma, respectively. See [FIGURE 5.3](#).

Melanoma, developed by an abnormal growth of melanocytes, is the less frequent but its aggressive malignancy turns it into the most deadly skin cancer [3] [4, 5]. That is why it is more likely to fast spread out of its origin and start metastasis. In these stages it is incurable and the

5. Some notions of melanoma diagnosis and computer aided diagnosis systems

treatment, only palliative, includes immunotherapy, chemotherapy, radiation therapy or palliative surgery [6]. Nevertheless, there are high survival rates when melanoma is detected in its primary stages, when it is locally confined to its lesion of origin (98% 5-year survival rate was obtained in the USA between 2002 and 2008 [6]). Consequently, the main objective for physicians is to detect melanoma in its early stages. These facts have placed computerized analysis of clinical and dermoscopic images as the main field of research in skin diseases.

Two measures help to prognosticate the malignancy of melanoma: Clark's levels and Breslow's depth of thickness. The first counts the sublayers of epidermis and dermis that the invasion has reached [7]. The latter consists of an evaluation of the depth of the tumor through a biopsy of the entire lesion [3] and it is the current best prognosis method of malignancy.



Figure 5.3: Squamous cell carcinoma, basal cell carcinoma, melanoma. [Fuente](#)

5.4. Melanoma screening and imaging techniques

When a dermatologist considers a lesion as suspicious, it is excised by a surgeon and histologically analyzed by a pathologist. The diagnosis of the pathologist is considered as the ground truth about the malignancy of the lesion when it was biopsied, because the tissue of the lesion itself is evaluated with a microscope. Applying this kind of definitive analysis to all atypical lesions is not appropriate: esthetically, a scar is developed for every excised lesion; medical resources to practice every biopsy are limited and shouldn't be used indiscriminately.

This environment creates the need for non-invasive techniques that allow the dermatologists to detect malignancy without practicing any unneeded excision. Different imaging techniques have been developed after the traditional clinical images (ordinary photographs of the lesions) such as Dermoscopy, Confocal Scanning Laser Microscopy (CSLM), Optical Coherence Tomography (OCT), Positron Emission Tomography (PET), Magnetic Resonance Imaging (MRI) and ultrasound techniques [5]. At present, computer vision techniques focus their efforts on the analysis of clinical and dermatoscopic images.

5.4.1. Clinical images

These are ordinary photographs taken by dermatologists to map the location of the skin lesion in the whole body or to perform lesion tracking over time. Clinical images show what a dermatologist sees with the naked eye.

5.4.2. Dermoscopy

Dermoscopy is a non-invasive imaging technique for the observation of pigmented skin lesions that allows appreciation of skin subsurface structures that are not visible with the naked eye. It

5. Some notions of melanoma diagnosis and computer aided diagnosis systems

is based on illuminating the skin directly by a device that allows magnification, typically from 6x to 40x or even 100x that can be dynamically adjusted. The optical device uses a method to reduce light reflections to visualize structures below the stratum corneum.

This technique improves melanoma detection with the naked eye, as it has been shown by many studies. For instance, Kittler et al. show that the improvement is near 49% after analyzing 27 studies from 1987 to 2002 [8]. However, it is needed to point out that this accuracy can only be obtained by experts in dermoscopy. Indeed, melanoma detection for dermatologists not familiar with this technique is not better than clinical examination without dermoscopy.

Two imaging techniques coexist: non-polarized light dermoscopy and cross-polarized light dermoscopy. The former needs to place an immersion fluid on the lesion to place a handheld scope or digital imaging system used to evaluate the lesion. This gel is responsible for reducing the high scattering of light at the stratum corneum to force deeper light reflections, revealing details from inner skin layers. The latter has the ability to use polarized light and polarizing filters to isolate useful reflections with or without immersion liquid, obtaining very similar images [9]. Subtle differences exist between these two techniques that cause selective enhancement of different dermoscopic structures.

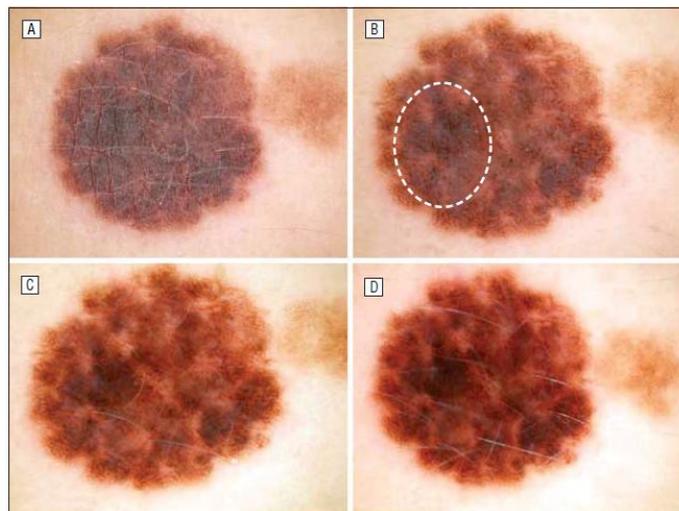


Figure 5.4: Clinical image (a), non-polarized light contact dermoscopy (b), polarized light contact dermoscopy (c), polarized light non-contact dermoscopy (d). [9].

5.5. Methods of melanoma diagnosis

Dermatologists have developed some algorithms to be used by themselves or even by clinicians to determine if a pigmented lesion is suspicious to be malignant. They are based only on visual information. It should be noted that a clinical diagnosis also goes with complementary information, such as the place of the body where the mole has developed, the family history of dangerous lesions, changes reported by the patient, the age, and some others. Also, it is helpful to remember that a dermatologist selects suspicious lesions to be excised, but it is the pathologist who makes the final decision on the malignancy of the lesion analyzing histologically the excised tissue. The methods of melanoma diagnosis can be classified into two types, according to the kind of image they are applied to either clinical or

5. Some notions of melanoma diagnosis and computer aided diagnosis systems

dermoscopic images. [TABLE 5.2](#) shows a classification of diagnosis algorithms according to the image type; [TABLE 5.3](#) and [TABLE 5.4](#) contain disambiguation of two methods adapted to both input images.

Among the existing methods for clinical images, the ABCD criteria and the Glasgow 7-point checklist methods should be mentioned. The first was born for clinical practice and it is more generally accepted because of its simplicity [10] [11]. This mnemonic stands for Asymmetry, Border irregularity, Color variegation and Diameter. It was proposed in 1985 by Friedman et al. [12]. Some years later, in 2004, the rule was expanded by Abbasi et al [11] to include evolution, which added changes in size, shape, texture and color to the previous list, to turn it into ABCDE criteria. These changes over time are pointed out by dermatologists to be meaningful [13] [14]. The Glasgow 7-point checklist [15] defines changes in size, shape and color as major criteria, while diameter, inflammation, crusting or bleeding and sensory change are minor criteria.

A larger number of diagnosis algorithms have been developed to analyze dermoscopic images [16], some of them are redefinitions of the previously discussed methods for this type of images. This is due to different visual information between clinical and dermoscopic images. These methods, with the exception of pattern analysis, need to determine before if the lesion is really a nevus, a melanocytic lesion. In other case, their conclusions are meaningless. Dermatologists have procedures to get this information and they are based on the fact that certain dermoscopic patterns are characteristic of an abnormal proliferation of melanocytes, which is the definition of nevus. This is the key that allows distinguishing between melanocytic lesions and the rest of brown colored lesions. Stolz et al [17] published in 1994 the method that is more generally followed at present. A detailed description of this method can be found in [APPENDIX I](#): Melanoma diagnosis algorithms.

The ABCD rule of dermoscopy was originally introduced by Stolz et al [17] to provide a reproducible algorithm to non experts in dermoscopy. Lesions are scored using the modified mnemonic ABDC, 'asymmetry, border sharpness, color and differential structure', leading to a final score that is compared with a threshold to decide if they are malignant.

Pattern analysis is considered by experts as the most accurate in melanoma diagnosis, but in contrast to the previous method, its application is not a rule of thumb and doesn't provide a quantitative result. Its application is only effective for highly experienced dermatologists. This is because it takes into account the whole list of dermoscopic patterns. It was originally developed by Pehamberger et al. in 1987 [18]. This method is based on the thought that each diagnostic category is characterized by a particular combination of specific dermoscopic patterns. However, the exceptions and wrong pattern identification can only be reduced by experience.

5. Some notions of melanoma diagnosis and computer aided diagnosis systems

Clinical image	Dermoscopic image
ABCD rule	ABCD rule
ABCDE rule	ABCDE rule
Glasgow 7-point checklist	7 point checklist
-	Menzies method
-	Pattern analysis

Table 5.2: Classification of diagnosis algorithms according to the type of images considered.

ABCD rule of clinical images	ABCD rule of dermoscopic images
Asymmetry: shape	Asymmetry: shape, colors and structures
Border: soft skin-lesion transition	Border: abrupt skin-lesion transition in 8 divisions
Color: uniformity of present colors	Color: presence of 6 defined colors
Diameter: threshold of 6 mm	Differential structures: presence of 5 dermoscopic structures

Table 5.3: Comparison between the ABCD rules for clinical and dermoscopic images.

Glasgow 7-point checklist	7 point checklist
1. Changes in size	1. Atypical pigmented network
2. Changes in shape	2. Blue whitish veil
3. Changes in color	3. Atypical vascular pattern
4. Diameter greater than 7mm	4. Irregular streaks
5. Inflammation	5. Irregular dots and globules
6. Crusting or bleeding	6. Irregular blotches
7. Sensory change	7. Regression structures

Table 5.4: Comparison between the 7-point checklist methods for clinical and dermoscopic images.

5.6. Computer Aided Diagnosis (CAD) systems for melanoma detection

Systems for automated diagnosis of melanoma try to mirror the expert dermatologist decision when watching the same dermoscopic image. Their primary objective is to help non expert dermatologists to improve their sensitivity and reduce benign lesions excision. Although computer vision techniques can be applied on every type of image data, clinical and dermoscopic images are preferred. These automated systems follow a very stable scheme: locating the lesion on the image, preprocessing when necessary, extracting features through descriptors, and making a diagnosis suggestion.

As revealed in some studies, the performance of these systems is satisfactory under experimental conditions [19]. A few systems showed enough true positive detection rate, but a false positive tendency greater than that of human diagnosis was detected, so automated diagnosis is not recommended as the only judgment to be taken into account. Day and Barbour [20] suggested a general flaw of the usual approach to the problem: an automated system is intended to reflect the decision of a pathologist using the same information than a dermatologist, who only decides on the suspiciousness of the lesion to be malignant; histopathological data may be not available for some examples because only suspicious lesions

5. Some notions of melanoma diagnosis and computer aided diagnosis systems

considered by dermatologists are excised. According to the study made by Dreiseitl et al [21], CAD systems should provide a kind of suspiciousness index to really resemble the decision of a dermatologist. This also would be helpful information for the clinician.

It may be added that the design of these systems should explore more alternatives to really support clinicians. Features extracted by a machine are objective and may find out information that is hidden to the eye. So, after selecting features through proper selection techniques, an analysis of significance could be made. The results would be explained to the clinician in order to be used as an information complement. This approach would let the dermatologist to learn from the numbers extracted by the automatic system, which may be a useful contribution.

6. State of the art of computer-aided melanoma diagnosis

Research on this field includes a great variety of applications, although the number of publications is not evenly distributed among all of them. A lot of effort has been devoted to segmentation of either the whole skin lesions or sub-regions corresponding to relevant dermoscopic local patterns. Feature extraction approaches follow two different paradigms: 1) extracting a large list of potential features followed by a feature selection process; and 2) relying on expert-based feature selection by exploiting the experience of dermatologists to design specific descriptors that account for similar information to that sought by dermatologists. In addition, some commercial CAD systems have been developed, mainly based on dermoscopy, likewise systems have been proposed for alerting patients about suspicious lesions, using either clinical images or even multispectral dermoscopy. The general review by Konstantin Korotkov and Rafael García [22] or the one by Magliogianis and Doukas [23] about feature extraction, selection, and classification techniques are highly recommended.

6.1. Lesion segmentation and image preprocessing

This is one of the classical topics of research in the field. First, the lesion segmentation consists in isolating the pigmented skin lesion from the normal skin as a singular region. Second, the image preprocessing covers techniques to highlight patterns and details which are informative for dermatologists through the application of certain image processing algorithms on the original image (image enhancement), or techniques to perform artifact removal, such as gel bubbles or hair, that could make more difficult subsequent computer vision-based analysis. Celebi et al. [24] make an interesting review of image preprocessing techniques.

Studies about lesion segmentation usually rely on manual outlines of the lesions performed by dermatologists as reference, which are used as ground truth to evaluate the performance of the method. However, it has to be noted that human observers are not very good at detecting soft changes in contrast or blur [25]. Particularly, dermatologists tend to overestimate lesion borders, usually including some healthy skin as a security margin when they outline the lesion contour to surgeons [20]. Furthermore, the morphology of some structures of pigmented skin lesions, such as regressions, or the variability of lesion-to-skin gradient also contribute to different segmentation criteria in practice. This problem is usually addressed by means of the fusion of manual segmentations of a certain number of experts.

Clinical images. The earliest works that can be found, published in 1989, used a spherical coordinate system for color space representation [26] and reported a comparison of color spaces for segmentation [27]. Jackowski et al [28] proposed a thresholding method followed by a refinement using border detection, using CIE Lab color space. A combination of grey level intensity and textural information was the choice of Dhawan and Sicsu [29]. Other approaches can be found, such as active contours [30] or gradient vector flow [31].

Dermoscopic images. Research using this imaging technique continued the trend started by clinical image. Fleming et al [32] performed a comparison of some segmentation and hair and

6. State of the art and computer-aided melanoma diagnosis

bubble removal methods, and developed a global segmentation method based on inverse diffusion equations. Thresholding approaches evolved from the previous simple ones [32] to threshold fusion [33] or hybrid thresholding [34]. Another set of techniques can be mentioned, such as evolution strategies [35], dynamic programming [36] or supervised learning [37]. It should be noted that, apart from the advantages and drawbacks offered by each method, algorithms were usually tested on small datasets coming from different sources, which makes difficult the evaluation of the results reported by the authors.

When a clinical or dermoscopic image is acquired, some noisy artifacts could be present on the image, which adds difficulties to subsequent image analysis. The objective of artifact rejection techniques is to remove all these spurious elements, keeping the rest of the image as intact as possible. These methods can be classified according to the artifact they aim to clear: Dullrazor [38] is a widely used method for hair removal. A procedure that tries to remove hair and interpolate the skin pixels below the hair along with a recent overview of the technique on dermoscopic images can be found in the work of Abbas et al. [39]. Gel bubbles are treated in [40]. Ruler markings are sometimes made by dermatologists to have a reference concerning the size of the lesion are considered in [41]. Image enhancement includes: color calibration techniques, to suppress undesired effects of the capture device can be found on [42] [37]; illumination correction is useful to normalize images [43] [44]; contrast enhancement procedures could help automatic segmentation and improve the perception of some dermoscopic structures [37] [45].

6.2. Feature extraction

There is a wide variety of features used to describe pigmented skin lesions. Some of them have been developed directly from expert knowledge, specifically created for this discrimination problem. The other approach is to adapt classic Computer Vision features to this matter. Both ideas are detailed below.

6.2.1. Features for lesion description

This category includes features intended to describe some global aspect of the lesion. They are usually implemented relying on a dermoscopic insight; however, other descriptors either coming from another medical field or even others classical descriptors in image analysis are sometimes tried also. Umbaugh et al [46] performed a review of CAD systems to introduce their feature extraction system, which has a wide variety of typical visual features (properly described in the paper). In particular, this paper describes features extracted from the binary mask of regions of interest, histogram-based descriptors, color transformations or Fourier descriptors for spectral analysis. The study reported by Magliogianis and Doukas [23] is a good review of features typically used for describing pigmented skin lesions. Descriptors based on the dermoscopic ABCD rule can be found, likewise other more general geometric features (area, perimeter, circularity, symmetry distance ...). Histogram features and perceptual color spaces used are also described. Also, wavelet analysis and Gray Level Co-occurrence Matrix-based features for texture modeling are described and referenced.

6. State of the art and computer-aided melanoma diagnosis

[TABLE 6.1](#) provides a list of descriptors. The first four rows cover ABCD rule based descriptors, which are the typical expert concepts used for feature developing. The rest of rows list more general image analysis techniques which have been adapted in some way to work on this field.

Asymmetry	[47]
Border	[48] [49]
Color variation	[47] [50] [51] [52] [53]
Differential structures	[51]
Fractal geometry	[50] [54]
Local Binary Pattern	[55]
Gabor filters	[56] [57]
Wavelets	[58] [59]
Bag of features	[60] [61]
Geometric	[47] [50] [52]

Table 6.1: References for different descriptor classes.

6.2.2. Pattern-oriented features

As described in section [5.5](#), abnormal local patterns can raise certain alerts. Indeed, the detection of these patterns is what allows expert dermatologists to make accurate and early diagnosis of melanoma. Some of the most common are pigmented network, globules, blue whitish veil, or homogeneous pattern (See [APPENDIX I](#): Melanoma diagnosis algorithms for a complete description). Additionally, the detection of this local pattern by clinicians becomes a difficult task for non-experts in dermatoscopy. These two facts make automated pattern detection an important matter of research in this field.

Articles on this field are essentially focused on the detection of one or two of the dermoscopic patterns described in [APPENDIX I](#): Melanoma diagnosis algorithms. In contrast, Di Leo et al [62] developed a complete system that allows the detection of all the dermoscopic patterns needed by the application of the 7-Point Checklist algorithm [63]. Specifically, a combination of one or more of these techniques was suggested to perform the detection of each local pattern: spectral analysis, Principal Component Analysis, structural analysis, and 2D thresholding. The work by Sáez, Acha and Serrano [64] performs an exhaustive review of these dermoscopic pattern detection research works. It starts with a discussion about dermoscopic characteristics of the patterns, followed by an explanation about the 7-Point Checklist [63] and pattern analysis-based [18] diagnosis methods. Finally, a wide variety of algorithmic methods for pattern detection and classification are described, and others referenced.

6.3. Classification

The last block of the CAD systems is the classification stage, whose output is the diagnosis/suspiciousness suggestion regarding the analyzed lesion. Some lesion distinctions can be made such as benign/malignant, common nevus/dysplastic nevus/melanoma, or even a finer classification that distinguishes other classes known by dermatologists. The classifier is fed with features extracted from the image, which try to obtain meaningful values for discrimination (instead of using directly the image pixels). See [TABLE 6.2](#). The paper by Magliogianis and Doukas [23] provides a review of the different classifiers used in literature

6. State of the art and computer-aided melanoma diagnosis

and reports their performance. Besides discussing the classifiers, the paper also proposes feature extraction and selection processes. A subset of the general-purpose image analysis features, such as border modeling, asymmetry, statistical moment-based texture features, or Grey Level Co-occurrence Matrix-based features are proposed. Then, three feature selection methods are evaluated together with the 11 most-frequently used classifiers. The comparison of such different system configurations using the same dataset is considered as highly useful among all the studies proposing different systems evaluated on proprietary databases.

When the number of features extracted is large, feature selection algorithms are often used. This is due to the fact that the performance evaluation of the proposed systems is typically performed using small datasets. Feature selection algorithms are then a solution that prevents from noisy features or an undesired number of the parameters to estimate for a limited dataset. An interesting example is the work by Rahil Garnavi [65], in which images are filtered by a 4-level wavelet decomposition and then some statistics are obtained from each filtered image as descriptors. Then, Gain Ratio Feature Selection (GRFS) and Correlation-based Feature Selection (CFS) are compared to find an optimal subset of features with reduced dimensionality. Finally, the performance of Support Vector Machines (SVM), Random Forest (RF), Logistic Model Tree and Hidden Naïve Bayes classifiers is checked, leading to a maximum Area Under Curve (AUC) (See section [8.3.2](#) for details of this measure) of 0.923 obtained with RF.

Classification algorithms	References
SVM	[47] [55] [66] [65]
Neural networks	[67] [68] [51] [69]
Discriminant analysis	[70]
Bayesian classifier	[61]
Decision trees	[62] [65] [71] [60]
Multiple classifiers	[72] [73]

Table 6.2: Types of classifiers and examples of use.

It should be noted that the most general case is that of presenting a research work evaluated on proprietary databases obtained through the collaboration with a determined hospital or dermatologist. The number of samples is usually small, because of the added efforts needed from dermatologists for segmenting lesions and local patterns besides their routine clinical work. These databases are not published and so, fair same-database comparisons are not possible with these studies.

On the other hand, there is another set of articles that use commercial databases to evaluate system performance. However, the whole set of images is never used because of reasons such as the presence of noisy artifacts, the lesion is partially out of the image, heterogeneity of image sizes, types the lesions not considered in the study, etc. In a few words, a subset of images is selected from the database according to some criteria determined by the authors of each work. Consequently, again, fair same-database comparisons are difficult to perform, even when the compared works using the same commercial database. Moreover, sometimes these subsets of commercial databases are combined with other sets obtained again through collaboration with a hospital or dermatologist, which increases the difficulty of making proper

comparisons. 'Interactive Atlas of Dermoscopy' [74], 'Interactive CD-ROM of Dermoscopy' [75] or 'An Atlas of Surface Microscopy of Pigmented Skin Lesions: Dermoscopy' [76] are often found in these works.

6.4. Complete systems

In this section we review the most important CAD systems found in the literature. This covers the implementation of diagnosis methods used by dermatologists, with the aim of putting the results in objective terms and making them repeatable, and complete commercial systems with hardware for image acquisition and storage that also provide image analysis software.

6.4.1. Automatic versions of dermoscopic diagnosis methods

Among all contributions whose features are based on previous expert knowledge, some of them try to follow in a more strict way the diagnosis methods used by dermatologists. The aim is that automated implementations of these procedures can make them repeatable because of deleting subjectivity variations inner to humans.

The works by Gola, García and Méndez [77] develop a system which implements an automatic ABCD rule calculation method taking a dermoscopic image as input, combining it with local pattern detection. The latter consists of making a detection of 8 different dermoscopic patterns, with an average accuracy above 85%. First, the lesion is automatically segmented, so the mask taken can be used to compute four features representing each of the steps of the rule and then, punctuation is fused with the results of automated pattern segmentation to perform the final punctuation.

An automatic implementation of the 7-Point Checklist is developed by Di Leo et al [62]. An image processing procedure is used to detect each dermoscopic structure taken into account in this algorithm. These automatic decisions are used as input to the original algorithm developed by dermatologist to decide its malignancy on a dataset of 300 images. For the detection of each dermoscopic pattern a training and a test set are taken from the primal dataset, keeping the same percentage of occurrences and absences of the corresponding dermoscopic criterion in both sets. Feeding the automatic detection of dermoscopic patterns to the original algorithm used by dermatologists, they get sensitivity and specificity of 0.83 and 0.76, respectively.

6.4.2. Commercial CAD systems

Here follows an exposition about commercial automated diagnosis systems. Reported evaluations of these commercial systems are commented, in order to explain the actual acceptance of these systems by the community of dermatologists.

Clinical support decision systems have been the center of many publications and their commercial availability is a fact. These image analysis systems use in general dermoscopic images as input data. Most of them consist of a complete set of devices, covering image

6. State of the art and computer-aided melanoma diagnosis

acquisition, analysis and storage tasks as a unit. Among the expert systems that consist purely on image analysis, DANAOS and MoleMax should be highlighted (See [TABLE 6.3](#) for references).

System	Imaging technique	References
DB-Mips	Dermoscopy	[78] [79] [80] [81] [67]
DANAOS	Dermoscopy	[82] [83] [84]
MEDS	Dermoscopy	[50] [72]
MoleMax	Dermoscopy	[85] [86]
MelaFind	Multispectral dermoscopy	[87] [88]
SpectroShade	Spectrophotometry	[89]

Table 6.3: Summary of commercial CAD systems.

The idea of supporting the dermatologist is focused on providing a diagnosis suggestion as output. For this reason, a large amount of medical articles places them as methods that work in parallel rather than in support of physicians. Vestergaard and Menzies [90] made a review of evaluation studies of these systems, assessing their methods and test datasets to find really meaningful works. Their revision shows three systems (DB-MIPS, Telespectrometry and Electrical Impedance) that provide similar performances to that of the experts in terms of sensitivity. However, human specificity is always found higher, so they have not gained the confidence of the experts. Dreiseitl and Binder [86] tried to find out insights about CAD systems interactions with dermatologists, in the sense of checking if they changed the diagnosis of a physician when different decisions occurred. It was found that in a 24% of cases clinicians changed their decision accepting the automatic system suggestion. This variation arose at interactions with non-expert dermatologists or dichotomous situation, revealing that clinicians are prone to accept CAD system decisions. An alternative view about the idea of CAD systems may be needed on this field. An interesting research area could focus on how automatically extracted features should be displayed to the dermatologist, in order to give them complementary information.

It is also worth mentioning that CAD systems are usually focused on visual information, while clinicians manage other data related to medical history of the patient, which is also useful and (usually) unknown by automated systems because of privacy reasons.

6.5. Image registration systems

Image registration is usually applied to make body maps of lesions, in order to detect changes which need more careful examination. This initial selection is critical when the patient exhibits a high number of pigmented skin lesions, which makes impractical more rigorous approaches. Other interesting field is the study of the evolution of the lesion during follow-up examination, removing disturbing transformations such as illumination changes or rotations, in order to appreciate variations in a fair way or even modeling them. However, the absence of databases with these characteristics directs the research efforts towards fields with available data.

The work by Rönig and Riech [91] described a semi automatic method for registration of skin lesions in subsequent body images, called baseline algorithm. The physician can provide the initial matches or they can be selected by the algorithm, finding the most likely initial matches

6. State of the art and computer-aided melanoma diagnosis

according to the images provided. The evaluation shows independence of the number of lesions and a matching accuracy of 99%. McGregor [92] aimed to remove noise caused by changes of camera angle or patient position through an automatic registration system that initially creates a lesion map from each image and then provides the identification of initial lesion in the following images using 'local neighborhood graphs'. The four pairs that provide lower matching distance are used to estimate an appropriate transformation between images. Huang and Bergstresset [93] proposed a graph matching technique, validated by means of the addition of noisy points after the lesion segmentation step, leading to a reduction in accuracy of 3% when 10% of points are noisy. Mirzaalian et al [94] added proximity-based regularization, used of anatomical information in the matching process, angular consistency checking between pairs, and normalized coordinates to compare graphs. The method was then evaluated over 56 pairs of real cases and hundreds of synthetic data pairs.

7. System description

The block diagram of the proposed system follows the standard arrangement of CAD systems and it is shown on [FIGURE 7.1](#). As it can be observed, the input to the system is a dermoscopic image and the output is a diagnosis suggestion (expressed as a soft decision: a real number between in the interval $\{-1, 1\}$). The classes to be distinguished are benign lesion and melanoma. The process starts with the segmentation of the region of interest, which in this case has been done manually because this study has focused on finding out features that are useful to discriminate the considered classes. Then, the feature vector is obtained from the region of interest (ROI). Finally, the classifier provides a soft decision. Each block of the system is explained in detail in the following subsections.

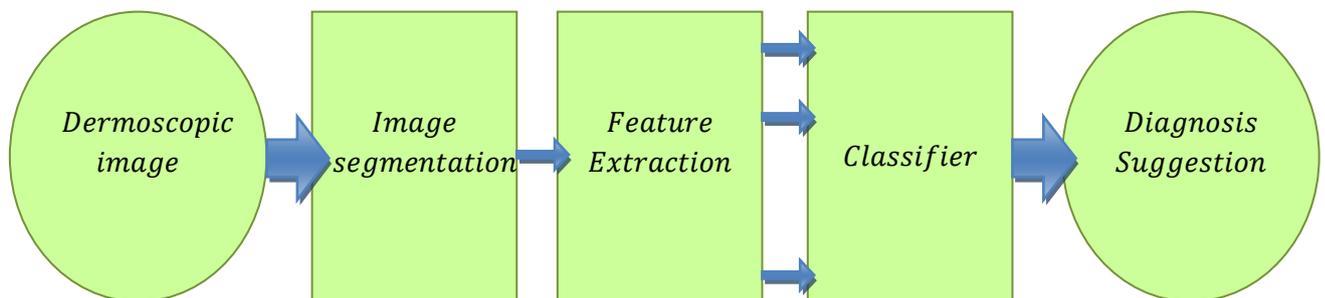


Figure 7.1: Block diagram of a generic CAD system.

7.1. Segmentation

Focusing the feature extraction process on the region of interest prevents from some noise sources, such as chromatic differences of the skin in different people or the differences between skin pixels and lesion pixels. As already mentioned, all the images of the database considered in our study have been segmented manually, in order to focus on feature extraction and selection and classification.



Figure 7.2: Manual segmentation of a nevus.

7.2. Feature extraction

The following features have been extracted and their relevance to the considered classification problem has been studied.

7.2.1. Color space description and justification

Opponent RGB [95] (oRGB) is part of the group of color spaces which consider a separation between luminance and chrominance information, such as CIE Lab, YCbCr or HSV. The model retains the ideas of hue and saturation present in HSV but using a non linear perceptual brightness, similar to HSB, but focusing on colors perceived as opposed. The oRGB color space proposes a different spacing between hues, setting the red-green opponence as the perpendicular chroma axis to yellow-blue opponence, rather than the magenta/red-cyan/green opponence. It covers the subjective fact that yellow does not seem to be composed by primaries as orange (mix of red and yellow) or purple (mix of red and blue). This adds the advantage of characterizing the warmth or coldness of a color using this confrontation (See [FIGURE 7.3](#)). These ideas make this model useful for color adjustment, color transformation or color transfer [95].

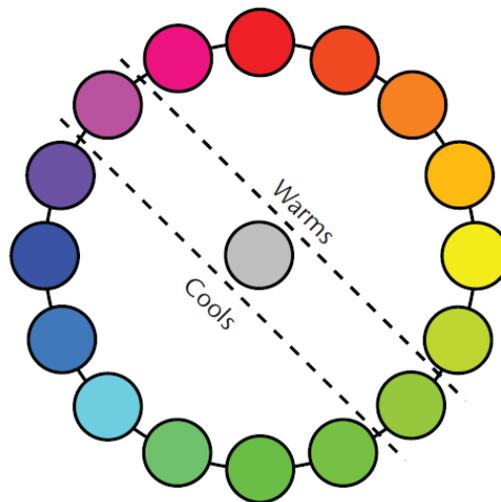


Figure 7.3: Color division according to their suggested temperature. A color is getting temperate when it approximates the drawn diameter. [95].

After the first theory of the perceptual base of color opponency, advocated by Hering [95], some psychological and physiological discoveries have evidenced it. Leo Hurvich and Dorothea Jameson suggested that the organization of color perception is partially based on opponent axes [96]. Cones are known to be the cells of the visual system that are responsible for the perception of colors, acting as photon counters. Each of the three types is associated with the detection of each of the three primary colors (red, green and blue), but the information about the wavelengths of individual photons is lost [97] and the information about wavelength and intensity for an individual cone, are both lost. The visual system performs at the retina a processing stage of the cones information to create three channels that are transmitted to the brain [98]. These three channels are related to the intensity, the red-green component and the blue-yellow component of a stimulus.

It is important to note that oRGB is merely a simplification of the real perception of colors that is focused on a particular aspect of human vision, as all perceptual color models.

The oRGB computation starts from perceptually corrected RGB coefficients as some models used for video encoding and televisions systems such as YIQ and YCbCr. A brightness-chroma separation followed by a non uniform color transformation leads to oRGB coefficients. The

simplest implementation approach described below leads to some out-of-gamut problems performing the inverse transform [95]. However, this process is chosen because of computational reasons and knowing that the problem does not include any inverse transform.

Linear transform of RGB to a transition color space: the initial RGB cube is turned into the LCC parallelepiped through a linear transformation expressed by the matrix shown in [EQUATION 7.1](#). This intermediate color space has a brightness axis, a yellow-blue axis and a magenta/red-cyan/green axis.

$$\begin{bmatrix} L' \\ C'_1 \\ C'_2 \end{bmatrix} = \begin{bmatrix} 0.2990 & 0.5870 & 0.1140 \\ 0.5000 & 0.5000 & -1.0000 \\ 0.8660 & -0.8660 & 0.0000 \end{bmatrix} \begin{bmatrix} R' \\ G' \\ B' \end{bmatrix}$$

Equation 7.1: Transformation matrix of the $R'G'B'$ device dependent cube into the $L'C'_1C'_2$. [95].

Non uniform rotation around brightness axis: as red and green are not positioned perpendicular to the yellow-blue axis, a non uniform transformation to stretch the space from yellow to red and green, compressing at the same time the space from blue to red and green (see [FIGURE 7.4](#)). It consists of a different constant scaling of angle for each of the two regions.

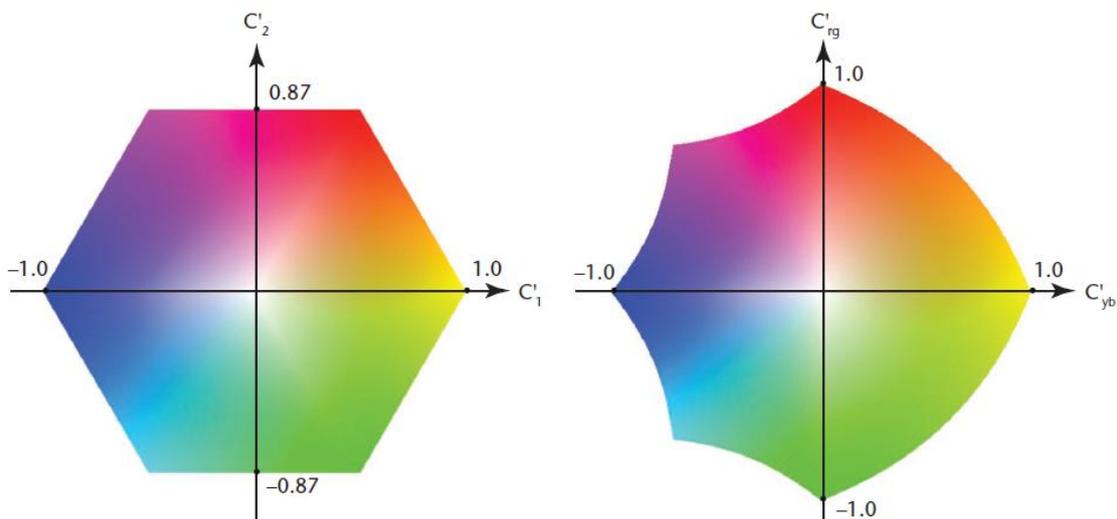


Figure 7.4: of the gamut or oRGB chromatic components for a certain luminance value, before (left) and after (right) uniform rotation to obtain orthogonal axes of opponent colors. [95].

7.2.2. Entropy

This feature measures the degree of heterogeneity of the pixel levels and can be computed for any of the color components. The entropy concept is taken from the classic Shannon's theory and applied to digital image processing [99]. It is a statistical measure of randomness that is computed from the histogram of the image (See [FIGURE 7.5](#)).

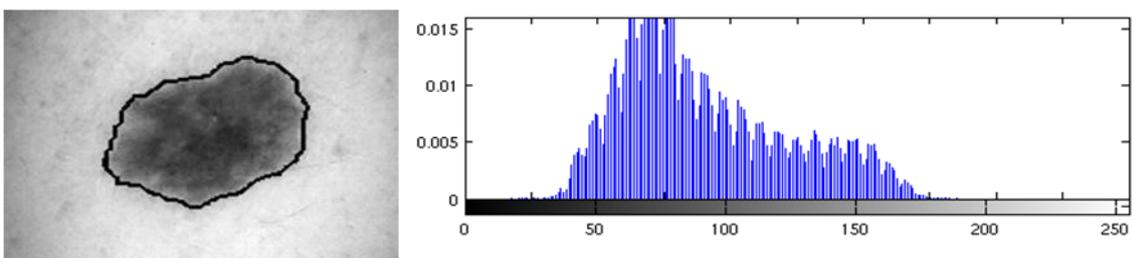


Figure 7.5: Color component of a nevus image (left) and its histogram (right).

Relying on the entropy, it is easy to separate images with uniform histograms from those of sparse histograms. Consequently, this descriptor can be used to describe the texture of an image focusing on intensity proportions, without taking into account any spatial relation between pixels. The computation over the histogram of a color component ranging from 0 to 255, where $p(x = i)$ is the probability of each value, is as follows:

$$Entropy = \sum_{i=0}^{255} p(x = i) \log_2 \frac{1}{p(x = i)}$$

Equation 7.2: Shannon's classic Entropy formula. [99].

7.2.3. Variance of the contour gradient

This descriptor measures the variability of the skin-to-lesion transition along the lesion contour. It allows us to distinguish between a mole with a uniformly defined border from a mole with alternating steep and soft transitions. As the entropy, it can be computed for any of the color components. The feature is based on the works by Grana et al [49]. The calculation procedure is detailed below.

The contour of the lesion is approximated as the boundary of the (manually segmented) region of interest. This contour is densely covered by short segments, which are centered on the contour line and perpendicular to it. The steepness of skin-lesion transition is then estimated on each segment using the pixel samples that each one covers. The length of the gradient segments, which will be denoted as L , is selected by visual inspection through the images of the database. To prevent from segmentation differences caused by human intervention, the length of gradient segments is initially duplicated. This brings a longer line of length $2L$ (see black lines on [FIGURE 7.6](#)) where a sliding window technique, with window size L , can be performed from one end to the other. A gradient slope is estimated at each sliding window step, and then the highest slope is selected to represent the steepness of the skin-lesion transition at each particular segment.

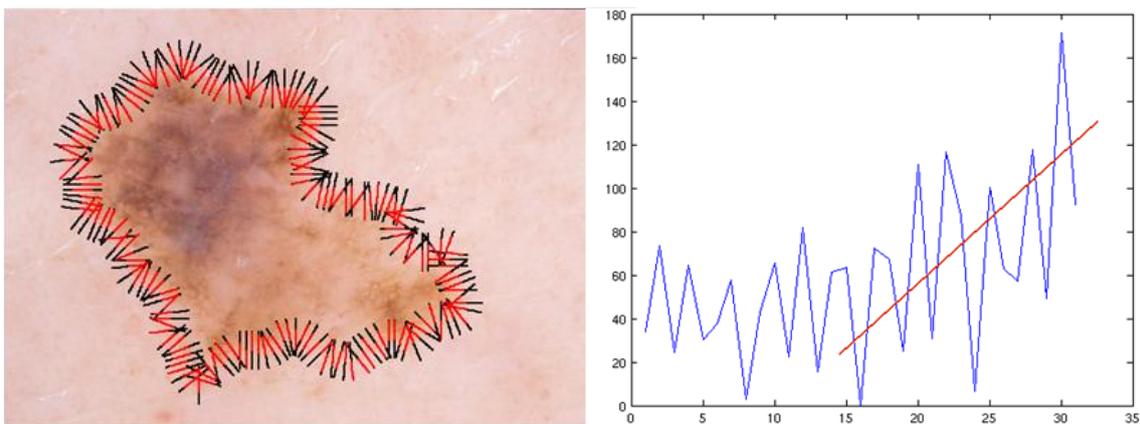


Figure 7.6: Segments modeling the direction of lesion-skin gradient (left). Pixel samples from a segment and estimated line to model transition sharpness (right).

Then, for each segment the slope of the skin-to-lesion transition is estimated through a linear regression. This method aims to get the line that minimizes the mean square error from a particular set of samples ([EQUATION 7.3](#) to [EQUATION 7.5](#)), each of them consisting on a pair of

coordinates (x_i, y_i) . In this case, x_i is the index of the pixel value taken from one line segment, ranging from 1 to the length of the segment and incremented in the sense lesion-skin. This rule gives a common x axis which gives sense to the comparison among slopes of different segments. The y_i values are the pixel values corresponding to the x_i index of a particular gradient segment.

$$y = ax + b$$

Equation 7.3: Lineal model to fit.

$$a = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2}$$

Equation 7.4: Slope formula.

$$b = \frac{1}{n} \left(\sum_{i=1}^n y_i - a \sum_{i=1}^n x_i \right)$$

Equation 7.5: Independent term formula.

Where n is the number of samples taken by each gradient segment.

The slopes measure how steep is the skin-lesion transition. The variance over all of them finally gives the feature.

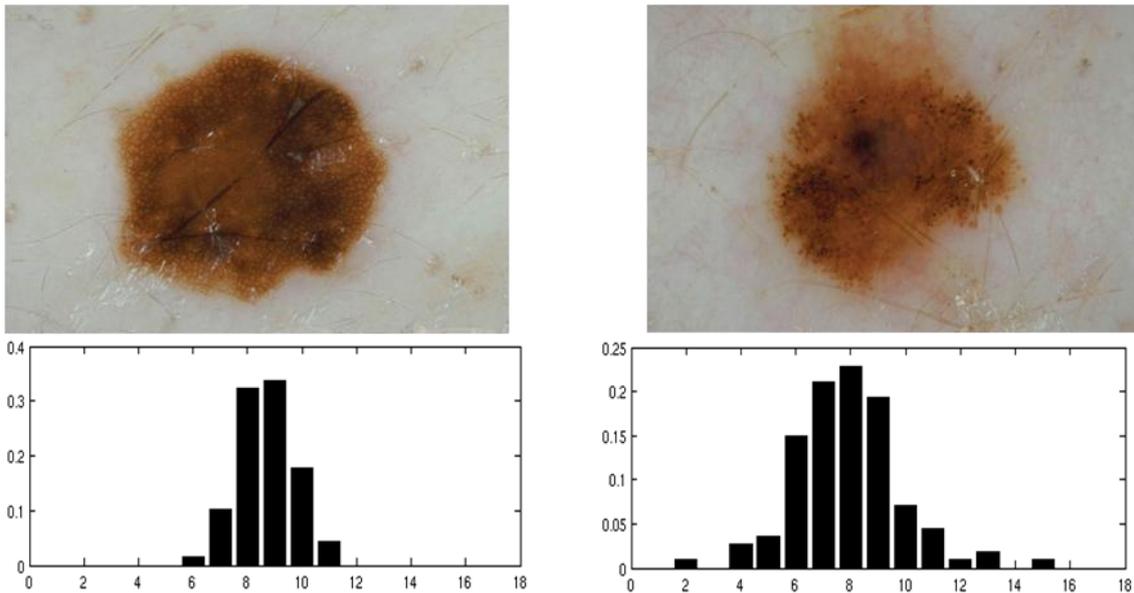


Figure 7.7: Comparison of lesion to skin transition histograms between a nevus with regular border (left) and another with more irregular contour (right).

7.2.4. Area

The area is computed as the sum of the pixels that make up the segmented region of interest. Melanomas tend to be bigger than benign lesions, but this difference becomes smaller when malignancy is more incipient.

7.2.5. Compacity

It describes how similar is the shape of the lesion to that of a circle. Malignant lesions tend to have more irregular shapes and, therefore, they would be less similar to a circle.

The procedure to compute this feature is as follows: first, the center of mass of the lesion is computed as the average of all lesion pixel coordinates; then, a circle of the same area is superimposed on the center of mass; the overlapping area that the circle and the lesion share constitutes the degree of similarity between them, which is divided by the area of the lesion to normalize with respect to all different lesion sizes.

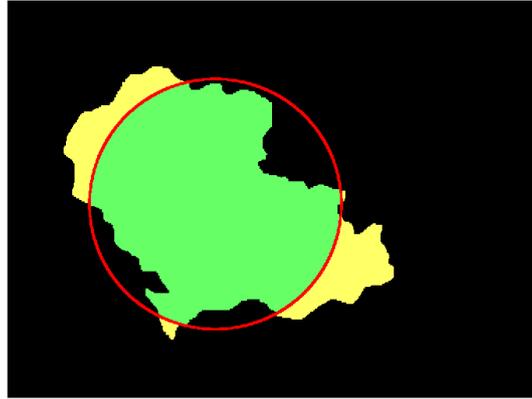


Figure 7.8: Compacity descriptor is based on the common area between the ROI of the lesion and the equivalent circle.

$$C(x_c, y_c) = \begin{cases} x_c = \frac{1}{n} \sum_i x_i | P \in M \\ y_c = \frac{1}{n} \sum_i y_i | P \in M \end{cases}$$

Equation 7.6: Centroid computation.

$$Compacity = \frac{M \cap E}{n}$$

Equation 7.7: Compacity formula.

$M = Lesion\ mask$

$E = Equivalent\ circle$

$n = number\ of\ pixels\ in\ S$

$P = (x, y)$

7.2.6. Asymmetry

This feature measures the lack of symmetry of a lesion with respect to its center. Melanomas tend to grow differently on each direction, becoming more asymmetric than benign lesions.

The computation of this features is performed as follows: first, the center of mass of the lesion is computed; an axis passing through this computed center is created to divide the lesion into two halves; taking this estimated symmetry axis, one half is superimposed on the other and the non-overlapping region is obtained; the process is repeated for several potential symmetry axis; the mean non-overlapping area divided by the total area of the lesion becomes a normalized asymmetry descriptor (to make it invariant to lesion size).

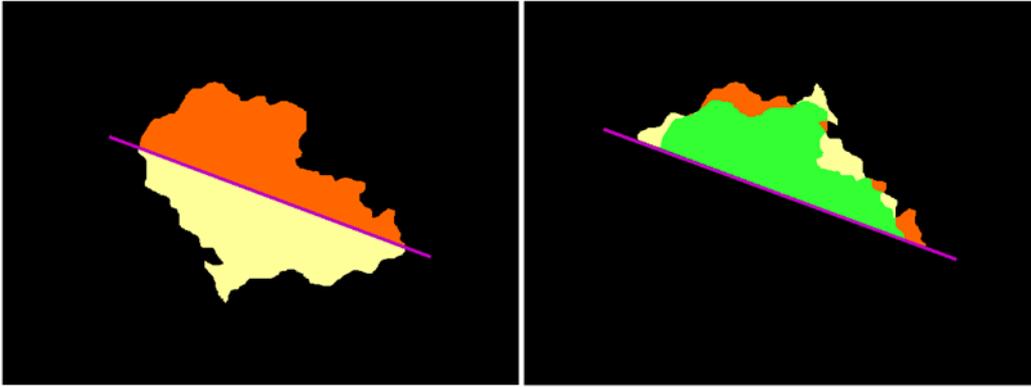


Figure 7.9: On the left, a lesion divided into two halves by one rotating axis. On the right, the half below overlapped over the half above, with the intersection highlighted as the green-colored region.

$$Sim_i = \frac{Half_i^1 \cap Half_i^2 (overlapped)}{Lesion\ area}$$

Equation 7.8: Symmetry formula.

$$Asymmetry = 1 - \frac{1}{n} \sum_{i=1}^n Sim_i$$

Equation 7.9: Asymmetry computation.

n = number of half division steps

7.2.7. Ellipsoidality

This descriptor measures how much the shape of the lesion resembles that of an ellipse with the same area. It is a generalization of circularity that encompasses lesions which are benign but have ellipsoidal shape rather than circular.

The computing procedure is as follows: the center of mass of the region of interest is computed as the average of all lesion pixel coordinates (See [EQUATION 7.6](#)); then, an ellipse with the same area of the ROI is centered at the computed center of mass, matching its major axis direction with the main axis of the lesion; finally, the overlapping area between the ellipse and the lesion region is divided by its total area to make the descriptor invariant to lesion size.

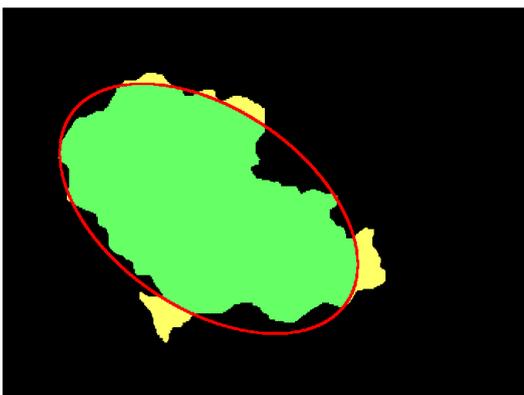


Figure 7.10: Ellipsoidality. Area inside the equivalent ellipse contributes to ellipsoidality.

$$Ellipsoidality = \frac{M \cap E}{n}$$

Equation 7.10: Ellipsoidality formula.

M = Lesion mask

E = Equivalent ellipse

n = number of pixels in M

7.2.8. Histogram modeling features

Every pixel of an image is represented as a three-component color vector. Consequently, three color histograms can be extracted. The histogram of a color component is a graphical representation of the relative frequency of every possible intensity value (typically ranging from 0 to 255 in discrete values). After some preliminary experiments dealing with several color spaces (RGB, HSV, Lab, Opponent RGB), the chosen color space was opponent RGB. Furthermore, we observed that the histograms can be approximately modeled as Gaussian (see [FIGURE 7.11](#)). For this reason, the histograms were decided to be modeled through their means and standard deviations.

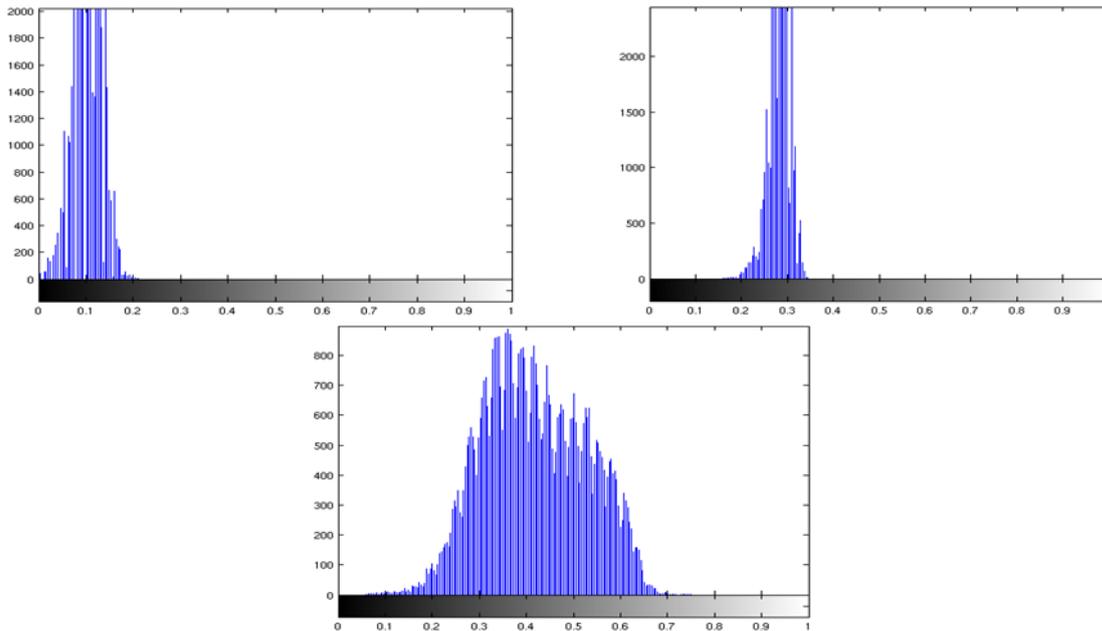


Figure 7.11: Upper left: Blue-Yellow chromatic component. Upper right: Red-Green chromatic component. Bottom: Intensity component.

7.3. Feature selection and classification

Since the feature selection process that we propose involves the use of the classifier, we first describe the basic of the proposed classifier, the Support Vector Machine, and then the feature selection algorithm, the Sequential Forward Selection algorithm.

7.3.1. Support Vector Machine

Support Vector Machines (SVM) [100] [101] [102] are a supervised learning algorithm, which means that they need a training set made up by example-label pairs to infer a model able to predict the class of unknown labeled data. Currently, they are among the most used learning algorithms because of both simplicity of use and high generalization ability for small datasets, which is indeed our case.

Given a training set, it is reasonable to try to find a decision boundary which looks for the largest separation gap between the samples of the two classes. This would bring confident predictions for the examples of the training set besides a model that correctly fits the data. In the following explanation, the training examples will be denoted as $x^{(i)}$, representing a feature

vector of n observations, with $y^{(i)} \in \{-1,1\}$ being their associated labels, forming a training set of size m . The maximum margin classification idea, developed by the SVM, goes after this sensible objective, trying to find the parameter values of a separating hyperplane that follows the expression $w^T \phi(x^{(i)}) + b = 0$ and maximizes the distance between the two classes, which is indeed the margin. This idea leads to the following optimization problem:

$$\min_{\gamma, w, b} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^m \xi_i$$

Equation 7.11: Expression to minimize on SVMs.

$$s. t. \quad y^{(i)}(w^T \phi(x^{(i)}) + b) \geq 1 - \xi_i, \quad i = 1, \dots, m$$

Equation 7.12: Restrictions based on data.

$$\xi_i \geq 0, \quad i = 1, \dots, m$$

Equation 7.13: Restriction for regularization.

Here, w and b represent the separating hyperplane parameters, result of the optimization process. The second term of [EQUATION 7.11](#) is the regularization term, which arises to solve the non-separable case. It occurs when the data is not linearly separable, that is, a perfect separation of the data cannot be done by a linear function. This is a realistic assumption; furthermore, finding a separating hyperplane could not be strictly desired in some cases. In [FIGURE 7.12](#) it can be seen that the addition of only one sample has dramatically rotated the decision boundary, leading to a margin that is strictly a good fit to the training data. In this case, the margin is far from being as large as initially, in a global sense. Regularization reduces the influence of outliers and noisy samples in order to get a more robust margin, but paying a cost at the objective function that can be tuned by parameter C . This parameter has the mission of adjusting the relative weighting between fitting rigorously the training set and making the margin large globally, and should be selected empirically by some validation process.

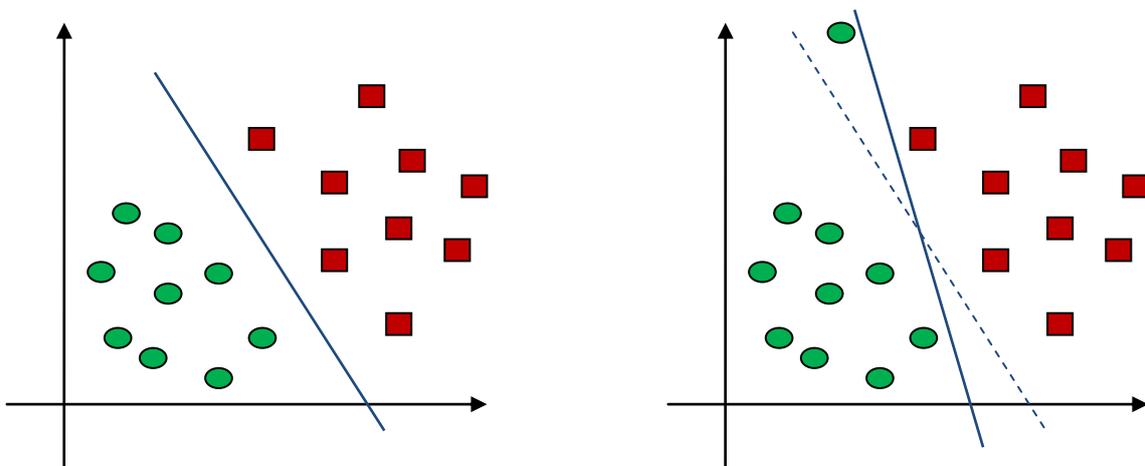


Figure 7.12: Influence of outliers over global margin when there is no regularization term.

The natural solution of a SVM results in a linear decision boundary. The use of kernels extends the SVM definition to cover non-linearity. The technique consists of performing a feature

mapping applying a particular function $\phi(x^{(i)})$ to the feature vectors. Carefully choosing this transformation, they can be turned into non-linear classifiers. Through the ϕ function the data is carried to a space with larger dimension, where is expected to find a separating hyperplane between the classes (see [FIGURE 7.13](#)).

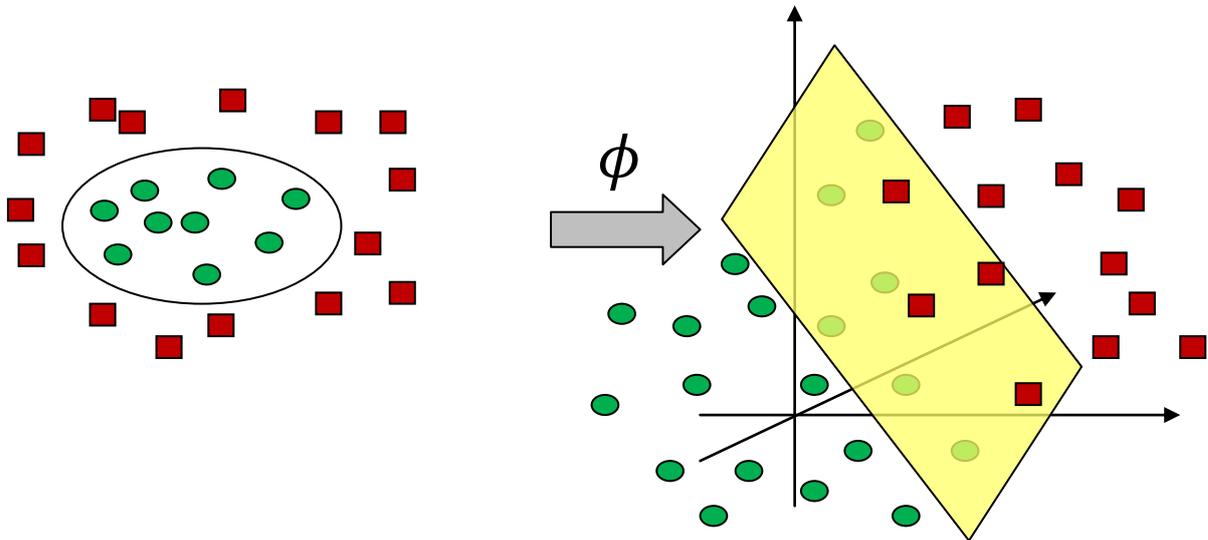


Figure 7.13: Space transformation using a convenient kernel. [102].

The kernel function is then defined as $K(x_i, x_j) = \phi(x^{(i)})^T \phi(x^{(j)})$. The most common are listed next:

- Linear: $K(x_i, x_j) = x_i^T x_j$
- Polynomial: $K(x_i, x_j) = (\gamma x_i^T x_j + r)^d, \gamma > 0$
- Radial Basis Function (RBF): $K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2), \gamma > 0$
- Sigmoid: $K(x_i, x_j) = \tanh(\gamma x_i^T x_j + r)$

Where γ , r and d are kernel parameters that need to be adjusted empirically.

The RBF kernel was chosen because of its ability in making robust models when the size of the dataset is small. Its characteristic parameter is γ , which models the range of influence of each sample in feature space. The RBF kernel is similar to a Gaussian, so γ can be seen as the inverse of the variance, which models the width of the Gaussian (See [FIGURE 7.14](#)).

The LIBSVM [103] library was selected as the implementation of SVMs to perform the experiments of this project.

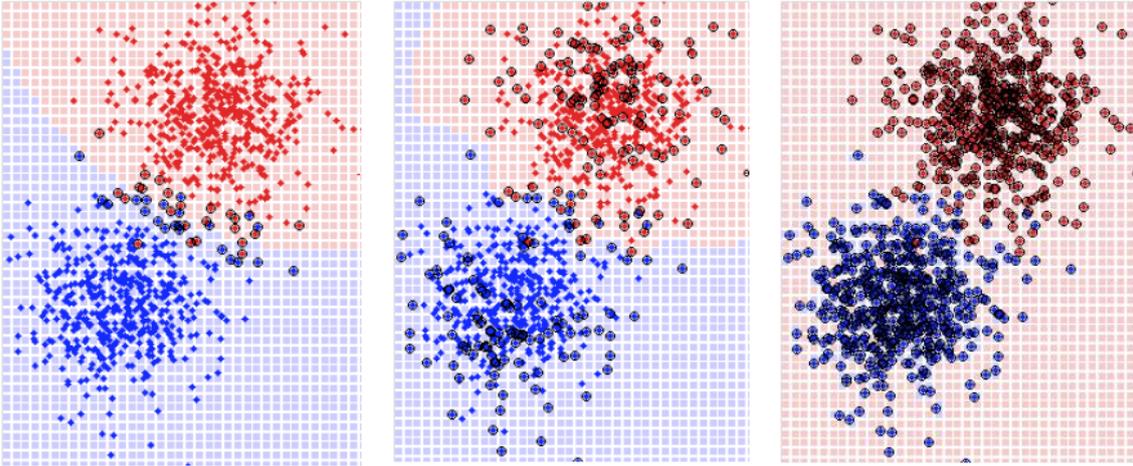


Figure 7.14: Decision boundary for a fixed gamma (0.1, 10 and 100, respectively) and its corresponding C term at its optimal value. The region for class decision is red or blue colored. [115].

To perform the parameter selection, the method proposed in [104] was followed, which uses RBF kernel and performs cross-validation to find the optimal value of the parameters: C and γ , in this case. It lies in sweeping C and γ values and pairing them to build a grid where the performance is checked for each (C, γ) pair. The process is detailed in [ALGORITHM 7.1](#). [FIGURE 7.13](#) is an example of how the performance of the classifier could improve when its parameters are tuned.

Set of training examples X is partitioned into M subsets.

$$C = 2^{-10}, 2^{-9}, \dots, 2^{10}$$

$$\gamma = 2^{-10}, 2^{-9}, \dots, 2^{10}$$

For each (C, γ) pair

For $i = 1$ to M

- $X_{train}^i = \bar{X}^i$
- $X_{test}^i = X^i$
- Use X_{train}^i to fit a model using the particular (C, γ) pair.
- Evaluate the performance of the (C, γ) pair with X_{test}^i as test set.

End

- Average the M performance measures of the (C, γ) pair to obtain its global cross-validation performance.

End

- Choose the (C, γ) pair with the highest cross-validation performance as the optimal one.

Algorithm 7.1: C and γ grid search validation algorithm.

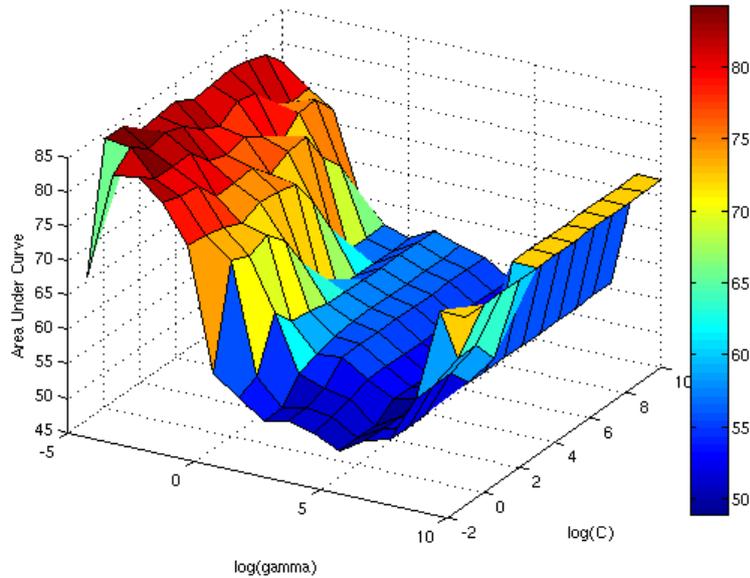


Figure 7.15: Performance of the classifier as a function of C and γ .

Scaling of the data before starting the work with the SVM is very important, because of the Vector nature of the algorithm. This is mainly to avoid that features with higher numeric ranges do not dominate over features with smaller numeric ranges. Another benefit is to prevent from numerical problems with the inner products of some kernel computations. See [ALGORITHM 7.2](#) to check the training and test procedure for the SVM.

1. Normalization of the samples. The mean and standard deviation of each feature along all the examples is computed. Then, the values are transformed by the following formula:

$$x_{norm}^{(i)} = \frac{x_i - \mu}{\sigma}$$
2. Perform cross validation through the training set to find the optimal combination of C and γ according to a particular efficiency measure (See [ALGORITHM 7.1](#)).
3. Using the best C and γ found to train the SVM using the whole training set.
4. Test the classifier onto unseen examples to evaluate its performance.

Algorithm 7.2: Training and test algorithm to use a SVM RBF classifier.

7.3.2. Sequential Forward Selection

Once the classifier has been described, the problem of making a proper feature selection is addressed. Extract all possible features and concatenating them as the input vector to the classifier would not be a good approach because feature vector would have a very high dimensionality compared with the number of examples in the database. Therefore, the vector that represents each lesion should ideally include only those features that really contribute to improve the classifier discrimination power. Otherwise, there would not be enough data to fit the model because the classifier would have too many free parameters. This task is part of the design of the system and tries to select only those features that are useful for discrimination between the two managed classes (melanoma vs. benign lesion).

The chosen algorithm for feature selection is based on ‘Sequential Forward Selection’ (SFS). The procedure starts from an initial feature set, which can contain only one feature. Then, a new feature is added among the ones that have not been added yet and the performance of the resulting feature set is evaluated. The operation is repeated adding each of the remaining features to finally select the one that carried a higher performance measure. The process is repeated until a certain stop criterion is met. In our case, the algorithm stops when the performance of the new set does not improve the performance of the previous set. See [ALGORITHM 7.3](#) for details.

```

It starts with an empty optimal feature set  $\mathbf{O}$  and a full remaining
feature set  $\mathbf{R}$  with all the extracted features.

Set of training examples  $\mathbf{X}$ .

For  $i = 1$  to  $M$  (maximum number of iterations)

    For  $j = 1$  to  $N$  ( $N$  is the number of remaining features in  $\mathbf{R}$ )
        ▪ A temporary feature set  $\mathbf{P}^{(j)}$  is created adding feature  $\mathbf{R}_j$ 
          to  $\mathbf{O}$ .
        ▪ Using feature set  $\mathbf{P}^{(j)}$ : the parameters of the classifier
          are cross validated using  $\mathbf{K}$ -fold cross validation.
        ▪ The optimal parameters for the feature set are saved and
          an efficiency measure is computed to compare feature set
          performances.
    End

    ▪  $\mathbf{P}^{(jmax)}$  is obtained as the temporary feature set with the highest
      performance measure.

    If ( $AUC[\mathbf{P}^{(jmax)}]$  is greater than  $AUC[\mathbf{O}]$ )
        ▪ The feature that was added to  $\mathbf{O}$  to form  $\mathbf{P}^{(jmax)}$  is added
          to  $\mathbf{O}$ , and the rest of them come back to  $\mathbf{R}$ .
    End

End

```

Algorithm 7.3: Sequential Forward Selection.

7.4. Diagnosis suggestion

Margin distance provided by the SVM classifier is used as the diagnosis suggestion. This magnitude is related to the distance that is between the sample and the boundary of the model fitted by the SVM. Its absolute value is related to the confidence of the prediction because large absolute values correspond to samples easy to classify because they are far from the margin. The sign of this value gives the information of the class that has been predicted, giving one sign to each class. This output is also known as soft output and its usefulness will be addressed in section [8.3.2](#).

8. System evaluation techniques for small datasets

This section covers the design decisions of validation and evaluation of the system. At first, the problem of small databases is analyzed to clarify its consequences. Then, the decided technique for training, validation and evaluation steps is presented. It gives a statistically correct solution to exploit the data in each of these three actions. Finally, the measure chosen to quantify the quality of the system is explained.

8.1. Small datasets problem insights

When the number of data examples is large enough, they can be separated into a training set, a validation set (for parameter selection), and a test set. If the quantity of data is large enough, these sets capture the variability of the data even though they are subsets of the total set of examples. In this case, training examples are enough to fit a robust data model and test examples are enough to provide a precise estimation of the performance of the system on unseen examples.

When the number of data examples is not large enough, a tradeoff appears: on the one hand, a large number of samples is needed for the training process, to fit a robust model of the data and to reduce the influence of outliers; on the other hand, a large number of samples is needed for the test process, to solve the problem of biased performance evaluation measures, which occurs when test data do not capture properly the variability of the problem.

When data samples are scarce, a simple division of the dataset in training and test sets (in any proportion) does not solve this problem. Dividing the available data into two halves may be fair but not enough for any of the two processes. Giving more samples to the training set will let us obtain the best model possible, but evaluation measures estimated from the test set will be biased. Consequently, the chosen test samples will have strong influence on evaluation measures, leading to high performance variability according to the samples chosen for test. In the opposite case, fewer samples for training and more for test, would cause poorly trained models highly dependent of the chosen training samples, which would end up in biased parameter estimation, high outlier sensibility, low system performance, and so on.

The proposed system also includes a validation process to optimize the C and γ parameters of the SVM classifier and a feature selection process, which reduces the dimensionality of the input feature vector. This introduces a need to add a validation partition from the whole set of available samples, which makes the data even more scarce.

To solve all these scarcity of samples problems, a more intense use of the data is required in order to build robust data models for classification, to make a feature selection process that finds out a statistically significant structure among features, and to use enough test samples that guarantee confident evaluation measures.

K-fold cross-validation is a statistical process to perform data exploitation; particularly useful when there are not enough available examples to perform an initial partition of the data that guarantees a good fit of the data model and statistically significant evaluation measures. The

8. System evaluation techniques for small datasets

algorithm is focused on assessing how the results of a statistical procedure generalize to a set that has not been seen, namely a set of independent samples, but exploiting available examples in a more intensive way than dividing into a training set and a test set. The process consists of partitioning the data set into K disjoint sets. One of them is used as the validation set and the others are joined to be the training set. This process is repeated with one of the sets being the validation set at a time. The best classifier parameters obtained for each validation set are averaged to find the optimal global ones and the performance measures for each set are also averaged to obtain a global averaged performance measure. The number of folds is a free parameter. In the stratified version, all folds are built so that they keep the same class proportion as the original non-partitioned dataset. Leave-one-out cross-validation is the K -fold variation that gives the highest priority to the number of examples used to training, using 1 sample for testing, i.e., the number of folds equals the number of observations.

The purpose of the method can be clarified in general terms as follows. A problem is defined through a set of examples and the goal is to adjust a model to this examples. The fitting process has to optimize the model so that it captures the structure of the data. If the same data is fed to the parameter validation process, the resulting model would be very likely overfitted to the data, losing generalization capability. Using cross-validation, independent samples are selected for training and validation steps, assuring that the model is not only focused on improving performance on its training examples. The process is repeated with different partitions and then averaged to reduce variability.

8.2. Proposed training and evaluation procedures

The database used in our study is not large enough to properly capture the problem variability of the data when dividing it into a training set and a test set. In addition, classifier parameters need to be optimized and a feature selection process has to be performed. A single cross validation process would let us perform parameter validation and feature selection. However, there would not be any way of knowing the validated system performance on previously unseen samples, which is an important drawback for this classic method and provides no way of drawing conclusions about how the system would work in real conditions.

The proposed technique enjoys general acceptance in the world of genomic experiments [105] [106], where the number of samples is very scarce and a statistically right method of evaluating the performance of a system is strongly needed. As far as we know, the method satisfies the requirements of this problem better than any other.

In particular, two nested K -fold cross validation procedures are performed. The inner one is responsible for validating the parameters of the classifier and selecting an optimal feature set. The outer one is given the job of evaluating the predicting power of the validated system against examples that have not been seen before by the system.

8. System evaluation techniques for small datasets

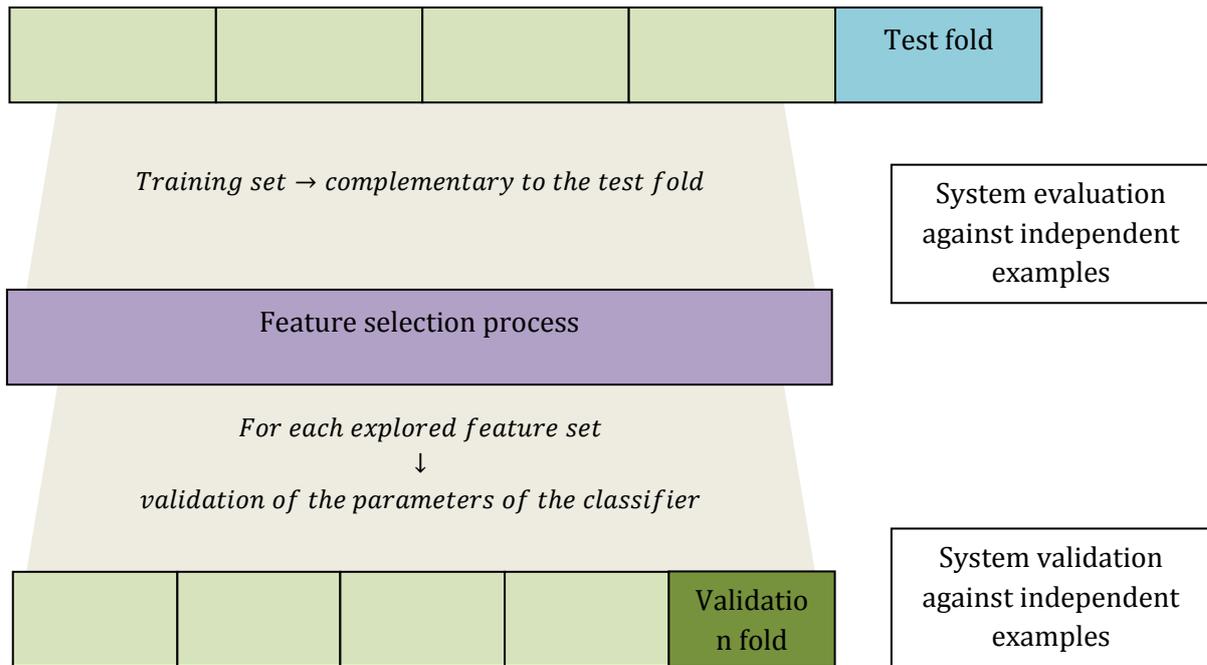


Figure 8.1: Proposed cross-validation method.

On each iteration of the outer layer, the training set is divided into folds to perform a K-fold cross validation process to get the optimized parameters of the classifier and an optimal set of selected features. The system is evaluated against the test samples, the fold that was not used in training, to assure independent evaluation measures. The process is repeated with another fold of the outer loop as test set, and the training set is divided to perform the nested K-fold cross validation process. Finally, evaluation statistics are averaged to reduce the variance of the estimation (See [ALGORITHM 8.1](#): Double K-fold cross-validation nested loop algorithm.).

It can be argued that this procedure brings no characterized system, in other words, there is not a final set of optimized parameters and a final set of selected features. Indeed, there are so many systems as folds are in the outer loop. Then, it is possible that different features selected and different optimized parameters of the classifier come from each of the partitioned training sets. Nevertheless, this method brings more advantages than drawbacks. It makes possible to perform a cross validation process to optimize parameters and select features where both tasks are validated with independent data. This validated system can be evaluated using again samples that have not been seen by the system before. The K-fold iterations let evaluation measures to be averaged, so the variance of the performance estimation is decreased.

Thus, this procedure provides an evaluation of the procedure that is statistically correct and fair, but it requires a more rigorous analysis of selected features and it does not offer a fixed system as output. These disadvantages are not critical because it is the scarce number of examples itself which prevents from building a unique system that can be properly assessed.

8. System evaluation techniques for small datasets

```
▪ Number of folds of the outer cross validation loop  $K$ .
▪ Number of folds of the inner cross validation loop  $L$ .

▪ Initial sample set  $X$  is  $X$  divided into  $K$  validation subsets
 $X = \{x_1, x_2, \dots, x_K\}$  with their complementary training subsets
 $X_{tr} = \{\bar{x}_1, \bar{x}_2, \dots, \bar{x}_K\}$ .

For  $i=1$  to  $K$ 
▪ The  $i$ -th training subset is partitioned into  $L$  validation subsets
 $\bar{x}_i = \{\bar{x}_i^1, \bar{x}_i^2, \dots, \bar{x}_i^L\}$  with their complementary training subsets.

    While the addition of a feature increments the previous performance

        For  $k=1$  to number of features out of the optimal feature set

            ▪ Temporarily add the  $k$ -th feature to the optimal feature set.

            ▪ Use  $\bar{x}_i$  (partitioned in  $L$  subsets) to perform the validation of the parameters of the classifier (See \[ALGORITHM 7.1\]).

        End

        If the best new feature set is better than the previous optimal set
            ▪ Add the new feature to the optimal feature set.
        End

    End

End

▪ Average all evaluation measures to obtain a global performance.
▪ Do feature selection analysis to try to find out some structure on the data.
```

Algorithm 8.1: Double K-fold cross-validation nested loop algorithm.

8.3. Evaluation measure for the performance of the classifier

A ratio of correctness is needed to guide model adjustment through validation processes and to evaluate the performance of the system. The Receiving Operating Characteristic (ROC) curve and the Area Under Curve (AUC) were respectively chosen for these two objectives. First, the required terminology is introduced, and then, both measures are explained in detail.

8.3.1. Useful ratio definitions

In binary (two classes) classification problems, four events can occur, according to the real class of a sample and the predicted one [107]:

- TP: *True Positives*. Correctly detected melanomas.
- FN: *False Negatives*. Not detected melanomas.
- FP: *False Positives*. Benign lesions incorrectly classified as melanomas.
- TN: *True Negatives*. Correctly classified benign lesions.

8. System evaluation techniques for small datasets

These absolute quantities can be turned into ratios using the definitions below [ROC intro]. Medical and machine learning domain use different terminologies for some identical concepts, which are paired to avoid confusions.

- TPR: *True Positive Rate*. Probability that a melanoma is detected by the system. It is also known as *detection probability, recall, or sensitivity*.

$$TPR = \frac{TP}{TP + FN}$$

- FPR: *True Positive Rate*. Relation between not detected melanomas and the whole bunch of these malignant lesions, and complementary to TPR. It is also known as *false alarm probability*.

$$FPR = \frac{FP}{TP + FN}$$

- TNR: *True Negative Rate*. Quotient of correctly classified benign lesions and all absent malignancy lesions. Known as *sensitivity* in medical fields.

$$TNR = \frac{TN}{TN + FP}$$

8.3.2. Receiving Operating Characteristic curve

This function is a graphical representation of the performance of a classifier which represents TPR as a function of FPR and represents the tradeoff between the two. It is also useful to compare different classifiers using a criterion of efficiency.

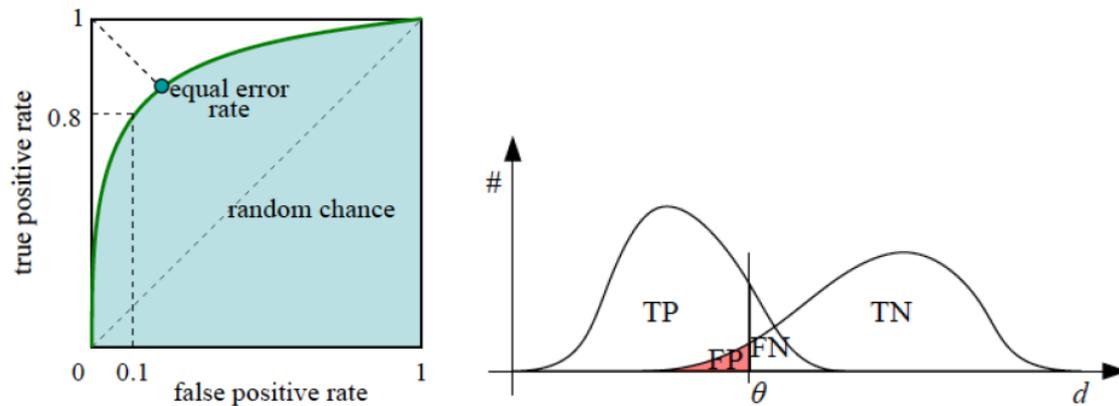


Figure 8.2: ROC curve registers all possible TPR-FPR for all possible values of the threshold θ . The equal error rate diagonal provides an operating point in which FPR and FNR are equal. [112].

The representation needs a classifier to provide soft outputs; this means a measure of the level of confidence of the decision given. The range of all outputs for the samples available to validate a classifier can be then separated by a shifting threshold which will have the decisions more likely to be positive at one side and the ones more likely to be negative at the other side. Example instances are then classified according to this threshold into one of the two classes (hard decision), which leads to a (FPR, TPR) pair and determines a particular working point for the ROC curve. With the move of the threshold, more (FPR, TPR) are taken to get the complete

8. System evaluation techniques for small datasets

trace of the ROC curve. A classifier that provides hard outputs is really working at a particular point of the ROC curve, so it can be associated only one (FPR, TPR) pair. The objective point of the ROC space is (0, 1), which represents the perfect classification: all positive examples detected and no false positives. On the other hand, the worst classifier can be found tracing the diagonal from (0, 0) to (1, 1) and represents the random chance classifier (See [FIGURE 8.2](#)).

The Area Under Curve (AUC) arrives as a simpler way of comparing classifier performances, which are reduced to a single scalar [107]. As the AUC of the perfect classifier is the unit square and the area under the random chance classifier is 0.5, it is deduced that a realistic classifier must lie in between these two limits.

9. Experimental results

This section covers the experimentation process that was carried out during the project. At first, there is an explanation about the questions which will be addressed and some relevant aspects of the problem. Then, the protocol followed in the preparation of the database for the experimentation phase is detailed. Finally, the results of performance and feature selection experiments will be analyzed, with a previous summary of the configuration of the system.

9.1. Scope of the study

The design decisions made in section 0 have also an impact on the information that can be obtained with the experiments. There are not one training set and one test set, both covering the variability of the problem, because the number of samples is limited. The training set has to be large because of model fitting, parameter validation and feature selection processes. In contrast, a correct evaluation of the performance of the system cannot be made without enough number of samples. This problem, detailed in 8.1, is solved with the method described in 8.2. However, the procedure does not provide one system as a solution. Feature selection and parameter validation are performed once for each fold of the outer layer, giving as many systems as folds has the outer layer. Therefore, we will take advantage of the fairness of the procedure to make a global analysis of the features selected by each system, in order to find out the most informative ones. It should be noted that this procedure is more statistically correct than any other seen on 'state of the art' papers, to the best of our knowledge. In addition, the ROC curve described in 8.3.2 will be used to provide an averaged performance along all folds. This will let the validation of the proposed procedure.

9.2. Database description and preparation

The database used in these experiments is the complement of [108], one sourcebook well known by dermatologists. Its images were gathered with the objective of providing a panoramic view of skin lesions diagnosis using dermoscopy. Consequently, the study is taken over a representative set of dermoscopic images.

All the pictures were saved in JPEG format. They were taken through the usual clinical practice in several hospitals, following no acquisition protocol. No kind of metadata can be found written on the image files, so both the device and the level of zoom used are unknown. The image size is almost uniform.

The database consists of a little more than 1000 images, all of them are labeled. However, it is remarked that each case, different images taken from the same lesion, is made up of 4 images on average: one clinical image, one dermoscopic images and two more pictures identical to the second but with annotations about interest points to make the diagnosis. Therefore, 1 out of 4 images can be used for our purposes. The evolution of lesions is only taken into account in 4 cases, not enough to make a study about follow-up information.

As a result of the facts commented before, the dermoscopy image free of annotations is chosen from each case. Non melanocytic images and lesions on palms or soles are discarded, because of being out of the scope of the study. Finally, lesions which are not entirely watched in the picture are rejected also, because geometric features calculation cannot be taken from them. 223 images come after this filtering process, 71 of them are melanomas and the other 152 are benign lesions.

Every image contained in the final set of images has some associated elements:

- A label, coming from the database.
- A mask, indicating the region of the lesion, obtained by manual segmentation.

In addition, a preprocessing step is applied to all images with the purpose of deleting the hair present on the image, which can distort feature computations on the region of the lesion. The Dullrazor software [38] was used to carry out this processing.

9.3. System performance evaluation

The experiments were all performed using almost the same system configuration. Fixed variables are specified below:

- SVM binary classifier, RBF kernel (7.3.1).
- Grid search for SVM C and γ parameters validation, both ranging $2^{-10}, \dots, 2^{10}$ (7.3.1).
- Feature normalization by mean and standard deviation (7.3.1).
- SFS for feature selection (7.3.2).
- 8 fold for the outer loop and 5 fold for the inner loop for the nested cross-validation algorithm (8.2).

ROC curve was chosen as the best evaluation measure for the problem, which needs test predictions from the soft output of the classifier to be calculated. It should be noted that the nested cross-validation method performs K training and test processes (one for each fold) giving predictions of the test samples on each step. This ends obtaining test predictions for all samples. However, all these test predictions cannot be jointly combined to compute a ROC curve directly. This is why each bunch of K predictions comes from a different system with potentially different features selected and optimal classifier parameters. The soft output for a SVM classifier is the margin distance and has different meanings on each of the K situations, so it does not make sense a combination of all margin distances.

Alternatively, a different ROC curve is obtained for each bunch of K predictions. Then, they are all averaged and the standard deviation is extracted. This provides an averaged evaluation measure of the performance of the proposed procedure with confidence intervals. At this step, the results will be biased because they are partitions-dependent. To solve this, partitions are made randomly and the experiment is repeated 50 times in a Montecarlo process. This leads to 50 different ROC curves showing the performance of the system on 50 different partitions, reducing the uncertainty of the evaluation measure. Their average gives the final estimated ROC curve for the global performance of the system. Supposing the Gaussian distribution, we

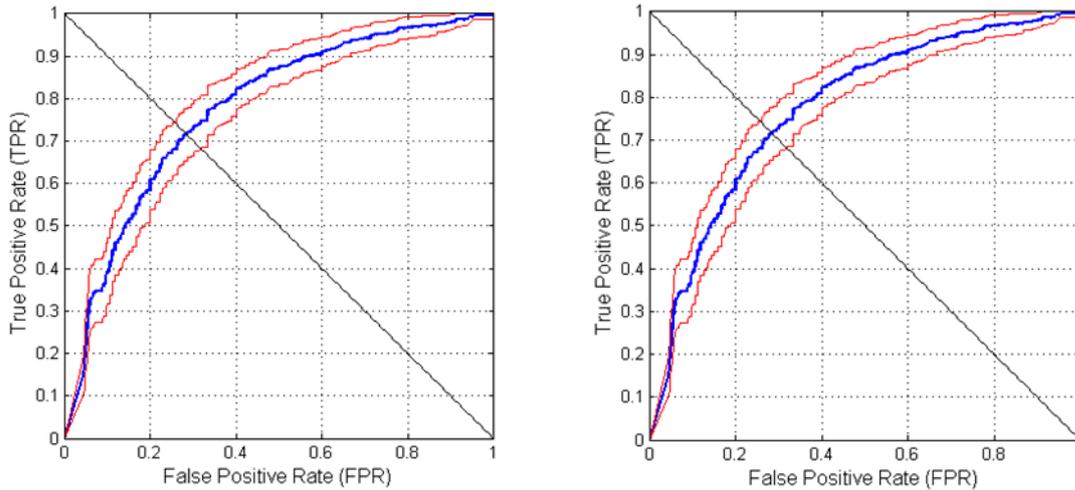
estimate a 95% confidence interval with 2 times the standard deviation divided by the square root of the number of iterations.

Two different initial sets of features are considered. The difference lies in the undesired effect of redundant information shared by features. This is applied to entropy and variance of border contour, which in *feature set 1* are extracted on the three color space components. The relevance of the feature itself is distributed among its components, hiding its real significance. Compacity is also removed because ellipsoidality is a generalization of this descriptor. These intuitions are verified by feature selection analysis in 9.4. The mapping between feature index and name for both sets is shown in [TABLE 9.1](#).

Initial feature sets		
Index	Feature set 1	Feature set 2
1	Lesion area	Lesion area
2	Entropy oRGB 1 st component	Entropy oRGB 3 rd component
3	Entropy oRGB 2 nd component	Ellipsoidality
4	Entropy oRGB 3 rd component	Variance of gradient contour oRGB 3 rd component
5	Ellipsoidality	Assymetry
6	Variance of gradient contour oRGB 1 st component	Mean oRGB 1 st component
7	Variance of gradient contour oRGB 2 nd component	Standard deviation oRGB 1 st component
8	Variance of gradient contour oRGB 3 rd component	Mean oRGB 2 nd component
9	Compacity	Standard deviation oRGB 2 nd component
10	Assymetry	Mean oRGB 3 rd component
11	Mean oRGB 1 st component	Standard deviation oRGB 3 rd component
12	Standard deviation oRGB 1 st component	–
13	Mean oRGB 2 nd component	–
14	Standard deviation oRGB 2 nd component	–
15	Mean oRGB 3 rd component	–
16	Standard deviation oRGB 3 rd component	–

Table 9.1: Initial feature sets.

The ROC curves obtained using the process described above for both feature sets are shown below. It can be seen that the proposed procedure gives good global performance even for two limited initial feature sets. It should be noted that *feature set 2* keeps the same performance of *feature set 1*, which supports the previous suspect of redundant information.



9-1: (Left) ROC curve for feature set 1, AUC = 0.7804. (Right) ROC curve for feature set 2, AUC = 0.7909. Both with 95% confidence intervals.

9.4. Feature selection analysis

The objective of this step is to know if the feature selection algorithm is finding out a structure inside the problem, or rather the procedure is overfitting the training set used to validate an optimal feature set.

The proposed nested cross validation algorithm (8.2) lets so many feature selection processes as folds has the outer cross-validation loop, which is 8 for all experiments. To minimize random variations, the whole experiment is also repeated 50 times.

A histogram which gathers the selection frequency of features is made. Each of them is obtained through the number of appearances among all feature selection processes, adding 1 to its corresponding histogram bar when it is present in a selection.

Another histogram reflects the step when each feature was chosen besides the number of times it was chosen. For each feature, a number, which will be denoted as M , ranging from the total number of features to 1 is added when it is present in a selection. This number is M when the feature was the first chosen and it is successively decremented by one as the step when it was selected moves forward. This lets the histogram to weight the selection frequency of a feature in particular with the order when it was chosen.

For the computation of both histograms, the feature selection algorithm is softly modified to stop at the best sixth feature, even when there is no performance improvement with respect to the previous step. This less restrictive criterion lets a more uniform analysis, not dependent on the number of features selected on each selection process. The histograms for *feature set 1* are shown in **¡Error! No se encuentra el origen de la referencia..** It can be seen that variance of border contour of the third oRGB component outperforms the other two components, so they can be removed. Ellipsoidality is a valuable feature, and compacity is a particular case, so it should be taken out. Entropy of the three oRGB components does not seem very significant but it would be possible that its significance was distributed along the three components, so

only the third one remains, as the most significant. These considerations build *feature set 2*. See [TABLE 9.1](#) for a mapping between histogram bars and features.

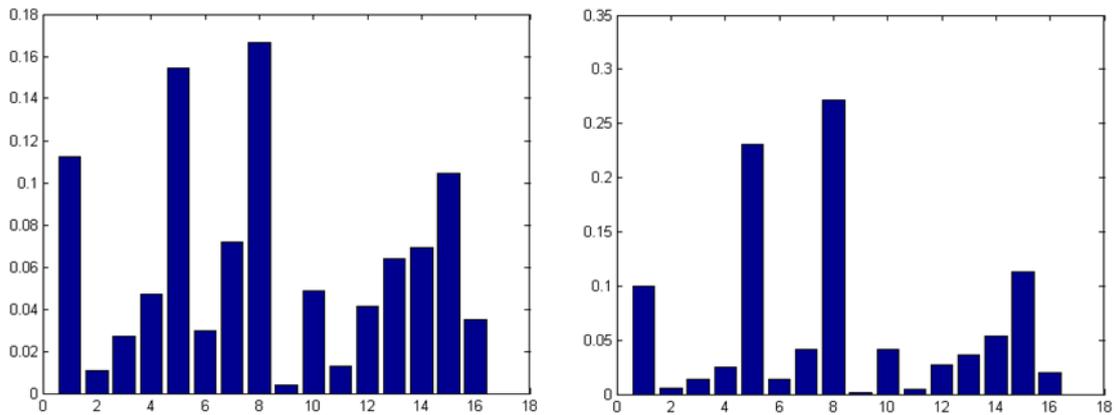


Figure 9.1: Frequency histogram (left) and weighted frequency histogram(right) for feature set 1.

The histograms for *feature set 2* are shown in [FIGURE 9.2](#). Variance of border contour and ellipsoidality continue being the most significant features. Entropy is indeed a low significant feature even when gained some importance due to maintaining only one color component. Lesion area and mean of the third oRGB component are again the next important features. This makes the third oRGB component highly informative, because it is present in this mean and in the computation of the variance of gradient contour kept.

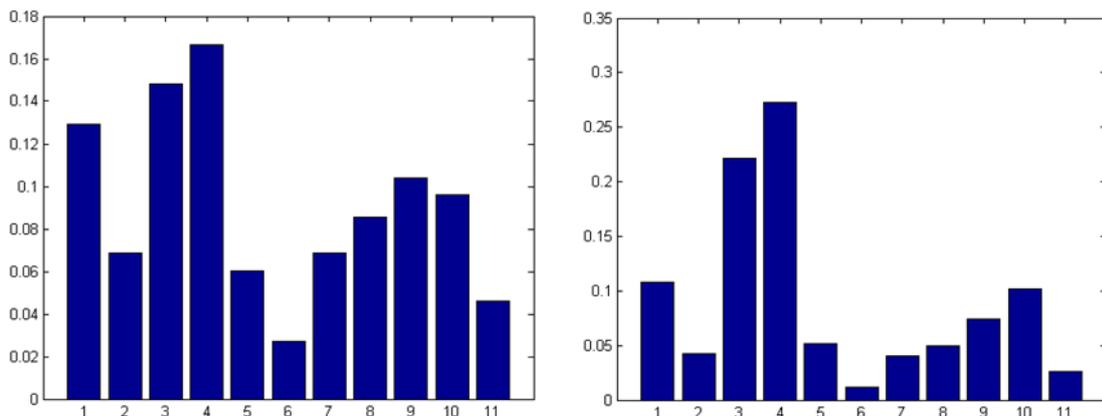


Figure 9.2: Frequency histogram (left) and weighted frequency histogram(right) for feature set 2.

This histogram analysis only reflects the importance feature by feature. All the feature selection sets obtained in different folds and iterations are then used to get statistics about the occurrence of some interesting combinations of features, which in this case involve features 1, 3, 4 and 10 (Lesion area, ellipsoidality, variance of contour gradient of third oRGB component and mean of the third oRGB component, respectively). See [TABLE 9.2](#). It can be seen that feature 4 is always selected, giving another argument in favor of its importance. Feature 3 is selected almost in 90% of times and when chosen, it is always joined to feature 4. This means that they are complementary, which is coherent with the fact that one is extracted from the shape and the other from pixel values. For the same reason, 3 and 10 are uncorrelated and the

table shows it, indeed. As it could be expected, 4 and 10 share some information because of coming from the same color component, as 4 almost doubles the importance of 10.

Set	Occurrence probability	Set	Occurrence probability
1	0,775	$3 \cap 4 \cap 1$	0,7375
3	0,8875	$3 \cap 4 \cap 1 \cap 10$	0,4875
4	1	$3 \cap 4 \cap 10$	0,575
10	0,575	$3 \cap 1$	0,7375
$3 \cap 4$	0,8875	$3 \cap 10$	0,575

Table 9.2: Occurrence probabilities of some interesting groups of features.

10. Conclusions and future work

A small set of features based on expert orientations have been tried in combination with basic Computer Vision features, represented by means and variances. The results are satisfactory, and, joined to the findings on 'state of the art' papers, reveal that image descriptors in this field are usually potentially better when based on some previous expert knowledge. In spite of this, the exploration of more elaborated descriptors will lead to better results.

CAD systems should also look friendlier to the clinician to favor their usage. It is difficult for a dermatologist to trust in a machine that only makes a final decision for what it is not responsible. Research should also explore finding Computer Vision features that could complement the information of the expert, to take them as another indication in diagnosis.

This work provides the needed tools for correct evaluation measures with small datasets, taking at the same moment the best exploitation of available data samples. The drawback of not providing a final system has been smoothed with correct evaluation measures and the extraction of informative features from different selection processes. It would be interesting that similar methods were tried, in order to make fair comparisons between different studies. It is also needed some database sharing to make comparisons possible, what is scarcely done right now. We think that this is the main issue that makes the advance in this field slower than that in other fields of Computer Vision.

References

- [1] DS Medica S.r.l., "dermoscopy.org," [Online]. Available: <http://dermoscopy.org>. [Accessed May 2014].
- [2] B. P. Negin, E. Riedel, S. A. Oliveria, M. Berwick, D. G. Coit and M. S. Brady, "Symptoms and Signs of Primary Melanoma," *Cancer, American Cancer Society*, vol. 98, no. 2, pp. 344-348, 2003.
- [3] J. A. J. J, S. C and C. T, "Early detection and treatment of skin cancer," *American Family Physician*, vol. 62, no. 2, pp. 357-386, 2000.
- [4] M. S, E. L, R. R, M. R, K. L and C. E, "Malignant melanoma in the 21st century. Part 1: Epidemiology, risk factors, screening, prevention, and diagnosis," *Mayo Clinic Proceedings*, vol. III, no. 82, pp. 364-380, 2007.
- [5] L. Smith and S. MacNeil, "State of the art in non invasive imaging of cutaneous melanoma," *Skin research and technology*, no. 17, pp. 257-269, 2011.
- [6] American Cancer Society, "Cancer facts & figures 2011," 2013.
- [7] B. S, C. N and G. C, Rook's textbook of dermatology, Oxford: Wiley-Blackwell, 2010.
- [8] K. H, P. H, W. K and B. M, "Diagnostic accuracy of dermoscopy," *Oncology The Lancet*, vol. III, pp. 159-165, 2002.
- [9] M. Cristiane Benvenuto-Andrade, M. Stephen W. Dusza, M. Anna Liza C. Agero, M. Alon Scope, P. Milind Rajadhyaksha, M. Allan C. Halpern and M. Ashfaq A. Marghoob, "Differences Between Polarized Light Dermoscopy and Immersion Contact Dermoscopy for the Evaluation of Skin Lesions," *Arch Dermatol*, no. 143, pp. 329-338, 2007.
- [10] D. Rigel, J. Russak and R. Friedman, "The evolution of melanoma diagnosis: 25 years beyond the ABCDs," *CA: A Cancer Journal for Clinicians*, vol. 5, no. 60, pp. 257-269, 2010.
- [11] N. Abbasi, H. Shaw, D. Rigel, R. Friedman, W. McCarthy, I. Osman and e. al, "Early diagnosis of cutaneous melanoma: revisiting the ABCD criteria," *Journal of the American Medical Association*, vol. 22, no. 292, pp. 2771-6, 2004.
- [12] R. Friedman, D. Rigel and A. Kopf, "Early detection of malignant melanoma: the role of physician examination and self-examination of the skin," *A Cancer Journal for Clinicians*, vol. III, no. 35, pp. 130-151, 1985.
- [13] S. Menzies, A. Gutenev, M. Avramidis, A. Batrac and W. McCarthy, "Short-term digital surface microscopic monitoring of atypical or changing melanocytic lesions," *Archives of Dermatology*, vol. 12, no. 137, pp. 1583-1589, 2001.
- [14] G. Salerni, T. Terán, S. Puig, J. Malvehy, I. Zalaudek, G. Argenziano and H. Kittler, "Meta-analysis of digital dermoscopy follow-up of melanocytic skin lesions: a study on behalf of the International Dermoscopy Society," *Journal of the European Academy of Dermatology and Venereology*, vol. 7, no. 27, pp. 805-814, 2013.
- [15] R. Mackie and V. Doherty, "Seven-point checklist for melanoma," *Clinical and Experimental Dermatology*, vol. II, no. 16, pp. 151-2, 1991.
- [16] G. Argenziano, H. Soyer, S. Chimenti, R. Talamini, R. Corona, F. Sera and e. al, "Dermoscopy of pigmented skin lesions: results of a consensus meeting via the Internet,"

- Journal of the American Academy of Dermatology*, vol. 5, no. 48, pp. 679-693, 2003.
- [17] W. Stolz, A. Riemann, A. Cagnetta and e. al, "ABCD rule of dermatoscopy: a new practical method for early recognition of malignant melanoma," *Eur J Dermatol*, no. 4, pp. 521-527, 1994.
- [18] H. Pehamberger, A. Steiner and K. Wolff, "In vivo epiluminescence microscopy of pigmented skin lesions. I. Pattern analysis of pigmented skin lesions," *American Acad Dermatol*, no. 17, pp. 571-583, 1987.
- [19] J. Frühauf, B. Leinweber, R. Fink-Puches, V. Ahlgrimm-Siess, E. Richtig, I. Wolf and A. Niederhorn, "Patient acceptance and diagnostic utility of automated digital image analysis of pigmented skin lesions," *Journal of the European Academy of Dermatology and Venereology*, vol. 26, pp. 368-372, 2012.
- [20] G. R. Day and R. H. Barbour, "Automated skin lesion screening - a new approach," *Melanoma Research*, no. 11, pp. 31-35, 2001.
- [21] D. Stephan and B. Michael, "Do physicians value decision support? A look at the effect of decision support systems on physician opinion," *Artificial Intelligence in Medicine*, no. 33, pp. 25-30, 2005.
- [22] R. G. Konstantin Korotkov, "Computerized analysis of pigmented skin lesions: A review," *Artificial Intelligence in Medicine*, no. 56, pp. 69-90, 2012.
- [23] S. Tasoulis, C. Doukas, I. Maglogiannis and V. Plagianakos, "Classification of Dermatological Images Using Advanced Clustering Techniques," in *32nd Annual International Conference of the IEEE EMBS*, Buenos Aires, Argentina, 2010.
- [24] M. Celebi, H. Iyatomi, G. Schaefer and W. Stoecker, "Lesion border detection in dermoscopic images," *Computerized Medical Imaging and Graphics*, vol. II, no. 33, pp. 148-153, 2009.
- [25] E. Claridge and A. Orun, "Modelling of edge profiles in pigmented skin lesions," in *Proc. medical image understanding and analysis*, Portsmouth, UK, 2002.
- [26] S. Umbaugh, R. Moss and W. Stoecker, "Automatic color segmentation of images with application to detection of variegated coloring in skin tumors," *IEEE Engineering in Medicine and Biology*, vol. 8, no. 4, pp. 43-50, 1989.
- [27] S. Umbaugh, R. Moss and W. Stoecker, "Automatic color segmentation algorithms-with application to skin tumor feature identification," *IEEE Engineering in Medicine and Biology*, vol. 12, no. 3, pp. 75-82, 1993.
- [28] L. Xu, M. Jackowski, A. Goshtasby, D. Roseman, S. Bines and C. Yu, "Segmentation of skin cancer images," *Image and Vision Computing*, vol. 17, no. 1, pp. 65-74, 1999.
- [29] S. A. Dhawan AP, "Segmentation of images of skin lesions using color and texture information of surface pigmentation," *Computerized Medical Imaging and Graphics*, vol. III, no. 16, pp. 163-177, 1992.
- [30] D. Chung and G. Sapiro, "Segmenting skin lesions with partial-differential-equations-based image processing algorithms," *IEEE Transactions on Medical Imaging*, vol. 7, no. 19, pp. 763-767, 2000.
- [31] J. Tang, "A multi-direction GVF snake for the segmentation of skin cancer images,"

- Pattern Recognition*, vol. 6, no. 42, pp. 1172-1179, 2009.
- [32] M. G. Fleming, C. Steger, J. Zhang, J. Gao and A. B. Cagnetta, "Techniques for a structural analysis of dermoscopic imagery," *Computerized Medical Imaging and Graphics*, vol. 5, no. 22, pp. 375-389, 1998.
- [33] M. Celebi, S. Hwang, H. Iyatomi and G. Schaefer, "Robust border detection in dermoscopy images using threshold fusion," in *Proc. IEEE Int. Conf. image processing (ICIP)*, Piscataway, NJ, 2010.
- [34] R. Garnavi, M. Aldeen, Celebi, G. M. Varigos and S. Finch, "Border detection in dermoscopy images using hybrid thresholding on optimized color channels," *Computerized Medical Imaging and Graphics*, vol. 11, no. 35, pp. 105-115, 2011.
- [35] X. Yuan, N. Situ and G. Zouridakis, "Automatic segmentation of skin lesion images using evolution strategies," *Biomedical Signal Processing and Control*, vol. 3, no. 3, pp. 220-228, 2008.
- [36] Q. Abbas, M. Celebi and I. García, "Skin tumor area extraction using an improved dynamic programming approach," *Skin Research and Technology*, vol. 18, no. 2, pp. 133-142, 2012.
- [37] G. Schaefer, M. Rajab, M. Celebi and H. Iyatomi, "Colour and contrast enhancement for improved skin lesion segmentation," *Computerized Medical Imaging and Graphics*, vol. 2, no. 35, pp. 99-104, 2011.
- [38] T. Lee, V. Ng, R. Gallagher, A. Coldman and D. McLean, "DullRazor®: a software approach for hair removal from images," *Computers in Biology and Medicine*, vol. 6, no. 27, pp. 533-543, 1997.
- [39] Q. Abbas, M. Celebi and I. Garcia, "Hair removal methods: a comparative study for dermoscopic images," *Biomedical Signal Processing and Control*, vol. 4, no. 6, pp. 395-404, 2011.
- [40] Q. Abbas, I. García and M. Rashid, "Automatic skin tumour border detection for digital dermoscopy using a new digital image analysis scheme," *British Journal of Biomedical Science*, vol. 4, no. 67, pp. 177-183, 2010.
- [41] H. Zhou, M. Chen, R. Gass, J. Rehg, L. Ferris and J. Ho, "Feature-preserving artifact removal from dermoscopy images," in *Proc. SPIE, vol. 6914 of medical imaging: image processing*, San Diego, CA, 2008.
- [42] J. Quintana, R. García and L. Neumann, "A novel method for color correction in epiluminiscence microscopy," *Computerized Medical Imaging and Graphics*, Vols. 7-8, no. 35, pp. 646-652, 2011.
- [43] J. Alcón, C. Ciuhu, W. ten Kate, A. Heinrich, N. Uzunbajakava and G. Krekels, "Automatic imaging system with decision support for inspection of pigmented skin lesions and melanoma diagnosis," *IEEE Journal of Selected Topics in Signal Processing*, vol. 1, no. 3, pp. 14-25, 2009.
- [44] K. Norton, H. Iyatomi, M. Celebi, S. Ishizaki, M. Sawada and R. Suzaki, "Three-phase general border detection method for dermoscopy images using non-uniform illumination correction," *Skin Research and Technology*, no. 18, pp. 290-300, 2012.
- [45] M. Celebi, H. Iyatomi and G. Schaefer, "Contrast enhancement in dermoscopy images by

- maximizing a histogram bimodality measure,” in *Proc. IEEE Int. Conf. image processing (ICIP)*, Piscataway, NJ, 2009.
- [46] S. Umbaugh, Y. Wei and M. Zuke, “Feature extraction in image analysis,” *IEEE Engineering in Medicine and Biology*, vol. 4, no. 16, pp. 62-73, 1997.
- [47] M. E. Celebi, H. A. Kingravi, [...] and R. H. Moss, “A methodological approach to the classification of dermoscopy images,” *Computer Medical Imaging and Graphics*, vol. 6, no. 31, pp. 362-373, 2007.
- [48] G. Day, “How blurry is that border? An investigation into algorithmic reproduction of skin lesion border cut-off,” *Computerized Medical Imaging and Graphics*, no. 24, pp. 69-72, 2000.
- [49] C. Grana, G. Pellacani, R. Cucchiara and S. Seidenari, “A new algorithm for border description of polarized light surface microscopic images of pigmented skin lesions,” *IEEE Transactions on Medical Imaging*, vol. 2, no. 8, pp. 959-964, 2003.
- [50] A. Sboner, P. Bauer, G. Zumiani, C. Eccher, E. Blanzieri and S. Forti, “Clinical validation of an automated system for supporting the early diagnosis of melanoma,” *Skin Research and Technology*, vol. 10, no. 3, pp. 184-192, 2004.
- [51] E. La Torre, B. Caputo and T. Tommasi, “Learning methods for melanoma recognition,” *International Journal of Imaging Systems and Technology*, vol. 4, no. 20, pp. 316-322, 2010.
- [52] G. Betta, G. Di Leo, G. Fabbrocini, P. A. and M. Scalvenzi, “Automated Application of the “7-point checklist” Diagnosis Method for Skin Lesions: Estimation of Chromatic and Shape Parameters,” in *Instrumentation and Measurement Technology Conference*, Ottawa, Canada, 2005.
- [53] 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,” *Systems Journal, IEEE*, vol. PP, pp. 1-15, 2013.
- [54] A. G. Manousakia, A. G. Maniosb and E. I. Tsompanakic, “Use of color texture in determining the nature of melanocytic skin lesions—a qualitative and quantitative approach,” *Computers in Biology and Medicine*, Vols. 419-427, no. 36, 2006.
- [55] X. Yuan, Z. Yang, G. Zouridakis and N. Mullani, “SVM-based texture classification and application to early melanoma detection,” in *Proc. 28th IEEE EMBS Annu. Int. Conf*, New York, 2006.
- [56] G. Capdehourat, A. Corez, A. Bazzano and P. Musé, “Pigmented Skin Lesions Classification Using Dermatoscopic Images,” in *14th Iberoamerican Conference on Pattern Recognition, CIARP*, Guadalajara, Jalisco, Mexico, 2009.
- [57] M. Sadeghi, T. K. Lee, D. McLean, H. Lui and M. S. Atkins, “Global Pattern Analysis and Classification of Dermoscopic Images Using Textons,” in *Proceedings- SPIE The International Society For Optical Engineering*, 2012.
- [58] R. Garnavi, M. Aldeen and J. Bailey, “Computer-Aided Diagnosis of Melanoma Using Border- and Wavelet-Based Texture Analysis,” *IEEE TRANSACTIONS ON INFORMATION TECHNOLOGY IN BIOMEDICINE*, vol. 16, no. 6, pp. 1239-1252, 2012.
- [59] R. Garnavi, M. Aldeen and J. Bailey, “Classification of Melanoma Lesions Using Wavelet-

- based Texture Analysis,” in *International Conference on Digital Image Computing: Techniques and Applications (DICTA)*, Sydney, NSW, 2010.
- [60] N. Situ, T. Wadhawan, X. Yuan and G. Zouridakis, “Modeling spatial relation in skin lesion images by the graph walk kernel,” in *Proc. Ann. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Piscataway, NJ, 2010.
- [61] N. Situ, X. Yuan, J. Chen and G. Zouridakis, “Malignant melanoma detection by bag-of-features classification,” in *International Conference IEEE Engineering Medical Biology Soc. (EMBC)*, Piscataway, NJ, 2008.
- [62] G. Di Leo, A. Paolillo, P. Sommella, G. Fabbrocini and O. Rescigno, “A software tool for the diagnosis of melanomas. Automatic Implementation of the 7-Point Check List Method,” in *Instrumentation and Measurement Technology Conference (I2MTC)*, Austin, TX, 2010.
- [63] G. Argenziano, G. Fabbrocini, P. Carli, V. De Giorgi, E. Sammarco and M. Delfino, “Epiluminescence microscopy for the diagnosis of doubtful melanocytic skin lesions. Comparison of the ABCD rule of dermoscopy and a new 7-point checklist based on pattern analysis,” *Arch Dermatol*, no. 134, pp. 1563-1570, 1998.
- [64] A. Sáez, B. Acha and C. Serrano, “Pattern analysis in dermoscopic images,” in *Computer Vision Techniques for the Diagnosis of Skin Cancer*, Berlin, Springer-Verlag, 2014, pp. 23-47.
- [65] R. Garnavi, “Computer-aided Diagnosis of Melanoma,” Australia, 2011.
- [66] F. Massimo and S. Ignazio, “Size functions for the morphological analysis of melanocytic lesions,” *International Journal of Biomedical Imaging*, 2010.
- [67] P. Rubegni, M. Burrioni, G. Cevenini, R. Perotti, G. Dell'Eva, P. Barbini, M. Fimiani and L. Andreassi, “Digital dermoscopy analysis and artificial neural network for the differentiation of clinically atypical pigmented skin lesions: a retrospective study,” *The Journal of Investigative Dermatology*, vol. 119, pp. 471-474, 2002.
- [68] J. Boldrick, C. Layton, J. Ngyuen and S. Swtter, “Evaluation of digital dermoscopy in a pigmented lesion clinic: Clinician versus computer assessment of malignancy risk,” *J. Amer. Acad. Dermatol.*, no. 3, pp. 417-421, 2007.
- [69] Y. Cheng, R. Swamisai, S. Umbaugh, R. Moss, W. Stoecker and S. Teegala, “Skin lesion classification using relative color features,” *Skin Research and Technology*, vol. 1, no. 14, pp. 53-64, 2008.
- [70] P. Rubegni, G. Cevenini, M. Burrioni, R. Bono, P. Sbanò and M. Biagioli, “Objective follow-up of atypical melanocytic skin lesions: a retrospective study,” *Arch Dermatol Res*, no. 302, pp. 551-560, 2010.
- [71] A. Kjoelen, M. Thompson, S. Umbaugh, R. Moss and W. Stoecker, “Performance of AI methods in detecting melanoma,” *IEEE Engineering in Medicine and Biology*, vol. 4, no. 14, pp. 411-416, 1995.
- [72] A. Sboner, C. Eccher, E. Blanzieri, P. Bauer, M. Cristofolini and G. Zumiani, “A multiple classifier system for early melanoma diagnosis,” *Artificial Intelligence in Medicine*, vol. 27, no. 1, pp. 29-44, 2003.
- [73] M. Ogorzałek, G. Surówka, L. Nowak and C. Merkwirth, “Computational intelligence and

- image processing methods for applications in skin cancer diagnosis," in *Biomedical engineering systems and technologies*, vol. 52 of *CCIS*, Berlin: Springer, 2010.
- [74] G. Argenziano, H. P. Soyer and V. De Giorgi, "Interactive Atlas of Dermoscopy," EDRA Medical Publishing and New Media, Milan, Italy, 2002.
- [75] A. A. Marghoob, R. P. Braun and A. W. Kopf, "Interactive CD-ROM of Dermoscopy," Informa Healthcare, London, 2007.
- [76] S. W. Menzies, K. A. Crotty, C. Ingvar and W. H. McCarthy, "An Atlas of Surface Microscopy of Pigmented Skin Lesions: Dermoscopy," McGraw-Hill Medical, 2003.
- [77] A. Gola Isasi, B. García Zapirain and A. Méndez Zorrilla, "Melanomas non-invasive diagnosis application based on the ABCD rule and pattern recognition image processing algorithms," *Computers in Biology and Medicine*, no. 41, pp. 742-755, 2011.
- [78] P. Rubegni, G. Cevenini, M. Burrioni, G. Dell'Eva, P. Sbrano and A. Cuccia, "Digital dermoscopy analysis of atypical pigmented skin lesions: a stepwise logistic discriminant analysis approach," *Skin Research and Technology*, vol. 4, no. 8, pp. 276-281, 2002.
- [79] M. Burrioni, U. Wollina, R. Torricelli, S. Gilardi, G. Dell'Eva and C. Helm, "Impact of digital dermoscopy analysis on the decision to follow up or to excise a pigmented skin lesion: a multicentre study," *Skin Research and Technology*, no. 17, pp. 451-460, 2011.
- [80] M. Burrioni, P. Sbrano, G. Cevenini, M. Risulo, G. Dell'Eva and P. Barbini, "Dysplastic naevus vs. in situ melanoma: digital dermoscopy analysis," *British Journal of Dermatology*, vol. 4, no. 152, pp. 679-684, 2005.
- [81] M. Burrioni, R. Corona, G. Dell'Eva, F. Sera, R. Bono and P. Puddu, "Melanoma computer-aided diagnosis," *Clinical Cancer Research*, vol. 6, no. 10, pp. 1881-1886, 2004.
- [82] P. A., O. Gaide, L. French, J. Saurat, A. Marghoob and R. Braun, "Can automated dermoscopy image analysis instruments provide added benefit for the dermatologist? A study comparing the results of three systems," *British Journal of Dermatology*, vol. 5, no. 157, pp. 926-933, 2007.
- [83] K. Hoffmann, T. Gambichler, A. Rick, M. Kreutz, M. Anschuetz and T. Grünendick, "Diagnostic and neural analysis of skin cancer (DANAOS). A multicentre study for collection and computer-aided analysis of data from pigmented skin lesions using digital dermoscopy," *British Journal of Dermatology*, vol. 4, no. 149, pp. 801-809, 2003.
- [84] M. Barzegari, H. Ghaninezhad, P. Mansoori, A. Taheri, Z. Naraghi and M. Asgari, "Computer-aided dermoscopy for diagnosis of melanoma," *BMC Dermatology*, vol. 5, no. 8, 2005.
- [85] 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 2001*, vol. 34, no. 1, pp. 28-36, 2001.
- [86] S. Dreiseitl, M. Binder, K. Hable and H. Kittler, "Computer versus human diagnosis of melanoma: evaluation of the feasibility of an automated diagnostic system in a prospective clinical trial," *Melanoma Research*, vol. 19, no. 3, pp. 180-184, 2009.
- [87] G. Monheit, A. Cognetta, L. Ferris, H. Rabinovitz, K. Gross and M. Martini, "The performance of MelaF ind: a prospective multicenter study," *Archives of Dermatology*, vol. 147, no. 2, pp. 188-194, 2011.

- [88] M. Elbaum, A. Kopf, H. Rabinovitz, R. Langley, H. Kamino and M. Mihm, "Automatic differentiation of melanoma from melanocytic nevi with multispectral digital dermoscopy: a feasibility study," *Journal of the American Academy of Dermatology*, vol. 44, no. 2, pp. 207-218, 2001.
- [89] S. Tomatis, M. Carrara, A. Bono, C. Bartoli, M. Lualdi and G. Tragni, "Automated melanoma detection with a novel multispectral imaging system: results of a prospective study," *Physics in Medicine and Biology*, vol. 50, pp. 1675-1687, 2005.
- [90] M. Vestergaard and S. Menzies, "Automated diagnostic instruments for cutaneous melanoma," *Seminars in Cutaneous Medicine and Surgery*, vol. 1, no. 27, pp. 32-36, 2008.
- [91] J. Röning and M. Riech, "Registration of nevi in successive skin images for early detection of melanom," in *Fourteenth International Conference of Pattern Recognition*, Brisbane, Qld., 1998.
- [92] B. McGregor, "Automatic registration of images of pigmented skin lesions," *Pattern Recognition*, vol. 31, no. 6, pp. 805-817, 1998.
- [93] H. Huang and P. Bergstresser, "A New Hybrid Technique for Dermatological Image Registration," in *Proceedings of the 7th IEEE International Conference on Bioinformatics and Bioengineering*, Boston, MA, 2007.
- [94] H. Mirzaalian, G. Hamarneh and T. K. Lee, "A Graph-based Approach to Skin Mole Matching Incorporating Template-Normalized Coordinates," in *IEEE Conference on Computer Vision and Pattern Recognition*, Miami, FL, 2009.
- [95] M. Bratkova, S. Boulos and P. Shirley, "oRGB: A practical opponent color space for computer graphics," *IEEE Computer Graphics and Applications*, vol. 1, no. 29, pp. 42-55, 2009.
- [96] L. Hurvich and D. Jameson, "An Opponent-Process Theory of Color Vision," *Psychology Rev.*, vol. 64, pp. 384-404, 1957.
- [97] W. Rushton, in *Handbook of Sensory Psychology, Vol. VII, Photochemistry of Vision*, New York, Springer, 1972, pp. 364-394.
- [98] K. R. Gegenfurtner, "Cortical mechanisms of colour vision," *Nature Reviews: Neuroscience*, vol. 4, pp. 563-572, 2003.
- [99] R. González and R. E. Woods, *Digital Image Processing*, Prentice Hall, 2008.
- [100] C. Cortes and V. N. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, pp. 273-297, 1995.
- [101] I. M. Guyon, B. E. Boser and V. N. Vapnik, "A training algorithm for optimal margin classifiers".
- [102] A. Ng, "<http://cs229.stanford.edu/>," 2013. [Online]. Available: <http://cs229.stanford.edu/notes/cs229-notes3.pdf>. [Accessed 5 29 2014].
- [103] C.-C. Chang and C.-J. Lin, "LIBSVM: a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, pp. 1-27, 2011.
- [104] C.-W. Hsu, C.-C. Chang and C.-J. Lin, "A practical guide to support vector classification," 2010. [Online]. Available: <http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf>.

[Accessed 30 May 2014].

- [105] W. Cao y K. Shimizu, «Identification of GPI-(like)-anchored Proteins by using SVM,» de *Proceedings of the First International Multi-Symposiums on Computer and Computational Sciences (IMSCCS'06)*, 2006.
- [106] H. Peng y J. Rui, «Integrating Multiple Gene Semantic Similarity Profiles to Infer Disease Genes,» de *Proceedings of the 31st Chinese Control Conference*, Hefei, China, 2012.
- [107] T. Fawcet, "An introduction to ROC analysis," *Pattern Recognition Letters*, vol. 27, pp. 861-874, 2006.
- [108] A. A. Marghoob, R. P. Braun, A. W. Kopf and J. Malvehy, *Atlas of Dermoscopy*, Abingdon, Oxon, UK: Taylor & Francis, 2004.
- [109] M. Yüksel and M. Borlu, "Accurate segmentation of dermoscopic images by image thresholding based on type-2 fuzzy logic," *IEEE Transactions on Fuzzy Systems*, no. 17, pp. 976-982, 2009.
- [110] M. Celebi, H. Iyatomi, G. Schaefer and W. Stoecker, "Approximate lesion localization in dermoscopy images," *Skin Research and Technology*, vol. III, no. 15, pp. 314-322, 2009.
- [111] M. Burrioni, L. Alparone and F. Argenti, "Comments on "A new algorithm for border description of polarized light surface microscopic images of pigmented skin lesions"," *IEEE Transactions on Medical Imaging*, vol. 25, no. 12, pp. 1655-1656, 2006.
- [112] *Computer Vision: Algorithms and Applications*, Springer, 2010.
- [113] B. Cummings, «osovo.com,» [En línea]. Available: <http://www.osovo.com/diagram/skindiagram.htm>. [Último acceso: 12 June 2014].
- [114] B. Publishing, «kreativestudios.com,» 2011. [En línea]. Available: <http://kreativestudios.com/Tooltip/05Integument/02epidermis.html>. [Último acceso: 12 June 2014].
- [115] F. de la Calle Silos, «Detección de eventos en secuencias con multitudes,» Universidad Carlos III de Madrid, DTSC., Madrid, 2012.

APPENDIX I: Melanoma diagnosis algorithms

In this section two algorithms used by dermatologists for the detection of melanoma are explained, as well as a method to check if a lesion is melanocytic, that is a previous step for the first one and many others. They are selected from a wide variety of them that exists nowadays, as the more related with our work and the more illustrative of dermatologists' procedures.

The ABCD rule

This dermoscopy diagnosis method was originally introduced by Stolz et al. [1] with the purpose of providing a more objective and reproducible diagnosis of melanoma for dermatologists who are not experienced in the use of dermoscopy. The algorithm was evaluated the same year of its publication by Nachbar et al. [2] in a prospective study. 172 melanocytic lesions, 69 melanomas and 103 melanocytic nevi, were used and they obtained 90.3% specificity and 90.8 sensibility.

Before applying the algorithm, the studied lesion must be classified as melanocytic. The results are meaningless for non melanocytic lesions. Melanocytic lesions are scored using the mnemonic ABCD, 'asymmetry, border, color and differential structure'. Each criterion gives a score that is multiplied by a given weight factor that yields a Total Dermoscopy Score (TDS). Values less than 4.75 correspond to benign melanocytic lesions, values between 4.8 and 5.45 point out suspicious lesions and values greater than 5.45 are strong candidates to melanoma. A detailed explanation of each criterion follows.

Asymmetry: the lesion is traversed by two perpendicular axes, each of them divides the lesion in two halves in such a way that the lowest asymmetry score is obtained. If there is asymmetry on one axis division, the score is one. If both axes show it, the score is two. When the lesion is symmetric according to both axes, the score is 0. Asymmetry encompasses differences of contour, color and dermoscopic structures between the two halves. It is a crucial criterion for the correct application of the method because of its high weight factor.

Border: the lesion is divided radially into eights. On each of them, sharp transition between the nevus and skin has a score of 1, and a gradual gradient has a score of 0. The maximum score is 8 and the minimum is 0. It is not very relevant because of its weight factor.

Color: it consists on analyzing the presence of six different significant colors in dermoscopy, namely, white, red, light brown, dark brown, blue gray, and black. A white area is only taken into account when it is lighter than the adjacent skin. Each color presence scores as 1. Melanomas are known to have a higher variety of colors than nevus.

Differential structures: It is related with the presence or absence of five selected dermoscopic patterns: pigment network, structureless or homogeneous areas, streaks, dots and globules. Structureless or homogeneous regions must be larger than 10% of the lesion. Streaks and dots are relevant only if more than two are clearly visible. Globules are considered even if only one is present. Each structure scores 1 if present, 0 if absent.

The tables below show a practical summary of the method.

ABCD rule of dermoscopy			
Criterion	Description	Score	Weight factor
Asymmetry	According to 0, 1 or 2 axes. Assess borders, colors and structures	0 – 2	X 1.3
Border	Abrupt skin lesion transition occurring at the border in 8 radial segments	0 – 8	X 0.1
Color	Presence of 6 different colors: black, dark brown, light brown, blue-grey, white, red	1 – 6	X 0.5
Differential structures	Presence of pigmented network, globules, dots, streaks, homogeneous areas or structureless areas	1 - 5	X 0.5

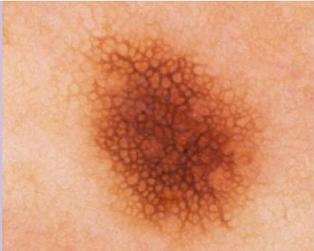
Summary table for ABCD rule of dermoscopy[17].

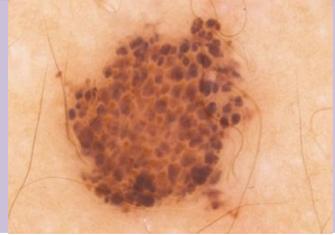
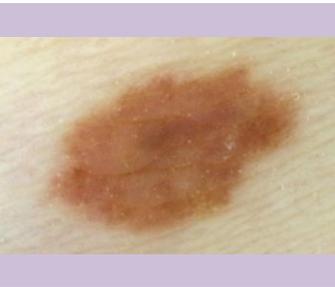
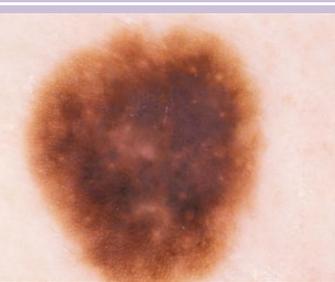
Total Dermoscopy Score (TDS)	Interpretation
Less than 4.75	Benign melanocytic lesion
From 4.8 to 5.45	Suspicious lesion: excision is recommended, or a close follow-up
Greater than 5.45	Lesion highly suspicious for melanoma

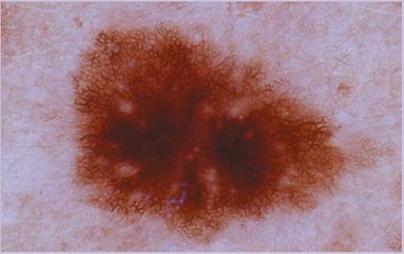
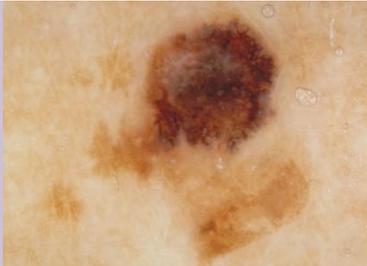
Matching between punctuations and diagnosis in ABCD rule of dermoscopy [17].

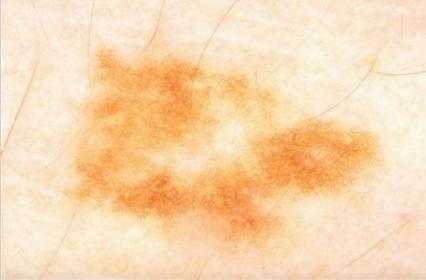
Pattern analysis

This method is considered by experts as the most accurate in melanoma diagnosis. Nevertheless, its application is only effective in highly experienced dermatologists. This is why it takes into account the whole list of dermoscopic patterns. It was originally developed by Pehamberger et al. in 1987 [1]. All of the rest diagnosis algorithms are simplifications of this one. It shows a correlation found out in different studies between a bunch of dermoscopic structures, ordered by how common is their appearance, and its corresponding type of skin lesion, no matter if it is melanocytic or not. However, the rules of decision depend on the dermoscopic patterns related with melanocytic skin lesions are explained in detail below, classified into global and local. Features of non melanocytic skin lesions are omitted because they are out of our object of study.

Global features	
<p>Reticular Pattern</p> <p>Pigmented network covering most parts of a given lesion. It appears as a grid of thin brown lines over a light brown background. It is a sign to detect melanocytic lesions and, if globally distributed, it is related to benign lesions. However, variations in size and form are one indicative of malignancy.</p>	

<p style="text-align: center;">Globular Pattern</p> <p>It consists on a certain number of round to oval structures, variously sized, with shades that can be brown and gray-black. It is a sign of lesion growing, very common in children. If present on the whole lesion, it is a sign of benignity.</p>	
<p style="text-align: center;">Cobblestone pattern</p> <p>This structure is similar to the globular pattern in the sense that it is made up by aggregated circular structures, but they are larger, more densely aggregated and somewhat angulated, resembling a cobblestone. It can be seen in congenital nevi, dermal nevi and sometimes in the dermal part of compound Clark nevi.</p>	
<p style="text-align: center;">Homogeneous pattern</p> <p>It is a diffuse structure, with brown, grey-black, grey-blue or reddish-black shade, where there is no other local feature that can be recognized. As a globally distributed pattern of bluish hue, it is the hallmark of the blue nevus. With other shades, it may be present in several types of lesions, such as Clark nevi, dermal nevi or nodular and metastatic melanomas.</p>	
<p style="text-align: center;">Starburst pattern</p> <p>It is a radial arrangement of pigmented streaks at the border of a given pigmented skin lesion. This feature is strongly related with Reed nevus, although malignant lesions may exhibit morphology very similar to this pattern.</p>	
<p style="text-align: center;">Parallel pattern</p> <p>This structure is exclusive of palms and soles skin, caused by the particular anatomic structure showed on this location. The pigmented pattern follows the direction of the parallel-furrow structure of the skin.</p>	
<p style="text-align: center;">Unspecific pattern</p> <p>This classification rises when a pigmented lesion cannot be categorized into one of the above global patterns, because there is not enough resemblance to any of them. No direct diagnosis implication can be extracted of it, nevertheless, it is more often related with melanoma. At least, it suggests that the lesion must be carefully explored.</p>	
<p style="text-align: center;">Multicomponent pattern</p> <p>It is a combination of three or more of the above global patterns joined to the following local patterns. This pattern brings to a lesion a high risk of being melanoma, but may also be present on several non melanocytic lesions. Consequently, the use of criteria to detect melanocytic lesions is needed in order to prevent from false positives.</p>	

Local features	
<p style="text-align: center;">Pigment Network</p> <p>Regular grid of brown lines over a light brown background. It is one dermoscopic indicator of melanocytic lesion. The assessment of this pattern is useful to help differentiating benign and malignant lesions, according to its uniformity. The typical pigmented network consists on a regular light to dark brown mesh with narrow spaces, dissipating smoothly at the border of the lesion. On the other hand, the atypical pigmented network is made of a black, brown or gray irregular mesh, which is also irregularly distributed through the lesion and ends suddenly at the periphery.</p>	
<p style="text-align: center;">Globules</p> <p>They are rounded or oval shaped structures with sharp border and usually aggregated. Sizes can be diverse and their color is limited to black, grey and light or dark brown. They are histologically related to aggregations of pigmented melanocytes, clumps of melanin or melanophages. Globules are placed in lesion regions that are growing. An even distribution of them with regular size and shape is associated with benignity; on the other hand various sizes and shapes, and irregular or located distribution occur in melanoma.</p>	
<p style="text-align: center;">Streaks</p> <p>They are black or light to dark brown longish structures of variable thickness, not clearly combined with pigmented networks, and easily observed when located at the periphery of the lesion but could be inside also. In the first case they tend to converge to the center or the lesion. An even, radial distribution of the streaks around the border of the lesion is characteristic of Reed nevus. However, an asymmetric or located distribution of streaks suggests malignancy.</p>	
<p style="text-align: center;">Blue-whitish veil</p> <p>It is a region of grey-blue to whitish-blue shade blurred pigmentation, correlated with pigmented network disorder, globules or streaks. The blue whitish veil is highly related to melanoma, although it can be present in Reed or Spitz nevi with no noticeable different appearance.</p>	

<p style="text-align: center;">Pigmentation</p> <p>This term encompasses pigmented diffuse areas with uniform color. As this definition allows a lot of variation, it can be found as blotches or irregular extensions too, its presence does not help discrimination. Nevertheless, an irregular distribution throughout the lesion is closer to malignancy.</p>	
<p style="text-align: center;">Hypopigmentation</p> <p>This term refers to diffuse areas with lighter pigmentation than it would correspond to an ordinary pigmented lesion. This localized pattern may appear focused in one area or more and its diagnostic power is poor, because only in rare occasions irregularly outlined hypopigmentation can be found in melanomas. In contrast, these areas are sometimes found in Clark nevi. It is important to remember that hypopigmented areas are distinct to melanomas with regression.</p>	
<p style="text-align: center;">Regression structures</p> <p>These structures are present when the immune system has attacked the lesion because of considering it potentially dangerous for some reason. It is dermoscopically revealed as white areas, blue areas and a combination of both. White areas are more or less well outlined and resemble a superficial scar. Blue areas may appear as diffuse blue-grey areas or peppering, which is an aggregation of blue-grey dots.</p>	

APPENDIX II: Tasks breakdown

ID	Task name	Work (hours)	Start	End	Previous tasks
1	Bachelor's Degree	364	10/02/2014	10/06/2014	–
2	Planning	8	Mon 10/02/2014	Tue 11/02/2014	–
3	Documentation	77,5	Wed 12/02/2014	23/04/2014	2
4	Project report composition - Milestone 1	12	Wed 12/02/2014	Tue 04/03/2014	2
5	Revision - Milestone 1	8	Wed 05/03/2014	Thu 06/03/2014	16;14;4
6	Project report composition - Milestone 2	18	Fri 07/03/2014	Thu 27/03/2014	17
7	Revision - Milestone 2	6	Fri 28/03/2014	Sat 29/03/2014	6;24
8	Project report composition - Milestone 3	17,5	Sun 30/03/2014	Fri 18/04/2014	7
9	Revision - Milestone 3	4	Sat 19/04/2014	Sun 20/04/2014	8;32
10	Project report global revision	12	Mon 21/04/2014	Wed 23/04/2014	33
11	Initial study	36	Wed 12/02/2014	Tue 25/02/2014	2
12	Dermoscopy and dermatology basis	18	Wed 12/02/2014	Tue 18/02/2014	2
13	Previous studies examination	18	Wed 19/02/2014	Tue 25/02/2014	12
14	Preparation of the DB for experimentation	18	Wed 26/02/2014	Tue 04/03/2014	13
15	Image selection	10,5	Wed 26/02/2014	Sat 1/03/2014	13
16	Mask production by manual segmentation	7,5	Sun 02/03/2014	Tue 04/03/2014	15
17	Milestone 1: Architecture of the prototype system and first phase of project report	0	Thu 06/03/2014	Thu 06/03/2014	5
18	Prototype development	54	Fri 07/03/2014	Thu 27/03/2014	17
19	Implementation of the system	36	Fri 07/03/2014	Thu 20/03/2014	17
20	Prototype validation	4,5	Fri 21/03/2014	Sat 22/03/2014	19
21	Experimentation using prototype system	10,5	Sun 23/03/2014	Wed 26/03/2014	20
22	Experimentation design	1,5	Sun 23/03/2014	Sun 23/03/2014	20
23	Experiments execution	9	Mon 24/03/2014	Wed 26/03/2014	22
24	Results analysis	3	Thu 27/03/2014	Thu 27/03/2014	23
25	Milestone 2: Second phase of project report	0	Sat 29/03/2014	Sat 29/03/2014	7
26	Complete system development	52,5	Sun30/03/2014	Fri 18/04/2014	25
27	Implementation of new components	22,5	29/03/2014	Mon 07/04/2014	25
28	System validation	6	Tue 08/04/2014	Wed 09/04/2014	27
29	Experimentation with complete system	15	Thu 10/04/2014	Tue 15/04/2014	28
30	Experimentation desing	3	Thu 10/04/2014	Thu 10/04/2014	28
31	Experiments execution	12	Fri 11/04/2014	Tue 15/04/2014	30
32	Results analysis	9	Wed 16/04/2014	Fri 18/04/2014	31
33	Milestone 3: Third phase of project report	0	Sun 20/04/2014	Sun 20/04/2014	9
34	Milestone 4: Complete project report	0	Wed 23/04/2014	Wed 23/04/2014	10
35	System improvement	136	Thu 24/04/2014	Tue 10/06/2014	–
36	Final Milestone: project report and system	0	Tue 10/06/2014	Tue 10/06/2014	–