Universitat Politècnica de Catalunya (UPC)
Universitat Autònoma de Barcelona (UAB)
Universitat de Barcelona (UB)
Institut de Ciències Fotòniques (ICFO)

http://www.photonicsbcn.eu

Master in Photonics

MASTER THESIS WORK

Morphological aspects in the diagnosis of skin lesions

Elisa Castañón García-Rovés

Supervised by Dr. Santiago Royo, (CD6, UPC)

Co-Supervised by Dr. Miguel Ares, (CD6, UPC)

Presented on date 8th September 2015

Registered at

Escola Tècnica Superior d'Enginyeria de Telecomunicació de Barcelona

Morphological aspects in the diagnosis of skin lesions

Elisa Castañón García-Rovés
Center of sensors, instruments and systems development (CD6), Rambla de Sant Nebridi, 10, Campus de Terrassa 08222, Barcelona

E-mail: elisa.castanon@upb.edu

Abstract. The ABCDE (Asymmetry, Border, Color, Diameter and Elevation) rule represents a commonly used clinical guide for the early identification of melanoma. Here we develop a methodology based on an Artificial Neural Network which is trained to stablish a clear differentiation between benign and malign lesions. This machine learning approach improves prognosis and diagnosis accuracy rates. In order to obtain the 6 morphological feature data set for each of the 69 lesions considered, a 3D handheld system is used for acquiring the skin images and an image processing algorithm is applied.

Keywords: Early melanoma diagnosis, data classification, feature selection, Artificial Neural Networks (ANN)

1. Introduction

The importance of recognizing early melanoma is generally accepted. The incidence of skin lesions, both non-melanocytic and malignant melanoma, has increased rapidly over the past two decades. Even though melanoma represents only 5% of skin tumours overall, it causes 91% of the deaths related with skin cancer. The aim of this project is to develop computer supported systems for melanoma diagnosis in order to increase diagnosis and prognosis accuracy.

The classic approach for the diagnosis of melanoma is based on simply visual inspection of skin surfaces by dermatologist, discriminating between health tissue and possible melanoma by using the ABCDE criteria, which stands for, Asymmetry, Border irregularity, Colour variegation, Diameter and Elevation [1]. This criterion relies almost completely on dermatologist and surgeons experience, so finding new methods that allow us to provide more accurate results is becoming a critical task.

In recent years, the introduction of imaging techniques in medical practice and the development of computers supported systems has changed the diagnostic approach, especially in dermatology, where the image is fundamental for the diagnosis. From 1995 on, the use of imaging techniques such as Epiluminescent microscopy (ELM) [2][3], Scanning Electron Microscopy (SEM) [4] or handheld 3D stereovision systems [5][6] has increased becoming one of the most relevant and promising research lines.

Imaging methods work in four typical steps: data acquisition of skin samples, lesion's segmentation, feature extraction and classification of the results. Features considered include specific patterns, colors, and intensities of pigmentation, as well as configuration, regularity, and other characteristics of both the margin and the surface of the pigmented skin lesions. Pattern analysis of these features by means of supervised learning methods allows to stablish a distinction between benign and malignant growth patterns. [7][8][9][10]

In this study, we use the images of different skin lesions, image processing techniques and an artificial neural network to distinguish melanoma from benign tumours. We first define the morphological features that are expected to provide a clear bias between both kind of lesions. Within the next step we implement a self-developed artificial neural network algorithm for data classification and discrimination purposes.

2. Theoretical background

2.1. Skin lesions

The skin is the largest organ present in the human body and its anatomy is quite complex.

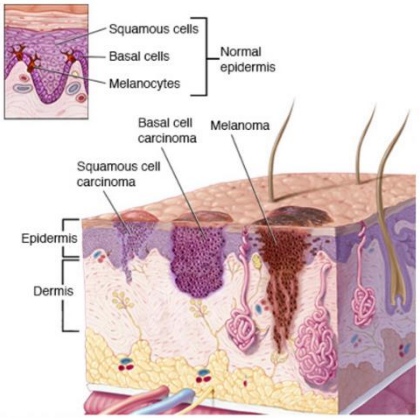

Figure 1. Different skin cells and malignant tumor evolution within the different skin layers.

Pigmented skin lesions are typically classified in malign and benign regarding if they are life-threatening or not, respectively.

2.1.1 Malignant lesions

There are three different types of malignant skin lesions: *Basal cell carcinoma (BCC)*, *Squamous Cell Carcinoma (SCC)* and *Melanoma (MM)*.

Basal cell carcinoma (BCC) is a malignant neoplasm or tumour derived from the basal cells which are small round cells located in the innermost layer of the skin. The incidence rate of this kind of lesion represents about the 80% of the non-melanoma skin cancer detected.

Squamous cell carcinoma (SCC) is a malignant tumour that appears due to the alteration of the epidermal keratinocytes which are located in the upper layer of the skin. This kind of lesion can appear in any part of the skin but they are most common to occur in the areas which have maximum sun exposure.

Malignant Melanoma (MM) is the most dangerous type of lesion. Although it only represents the 4% of total skin cancer cases detected, almost the 65% of all skin cancer-related deaths, are due to this kind of lesion. Melanomas often appear as an asymmetrical, irregularly bordered and coloured lesion that increases its size over time. Melanoma exists in two states, being in its first state localized occupying the uppermost layers of the skin, while in its second one beginning to become invasive penetrating more deeply into the skin and metastasize [11]

All the malignant lesions described above are not life-threatening if they are treated at an initial stage.[12]

2.1.2 Benign lesions

Within this work we consider two different kind of benign lesions: melanocytic nevus and dysplastic nevus. Melanocytic nevus, also known as birthmarks or moles, appear in the firsts decades of living. They are small brown spots that can be flat or elevated, and generally round and regularly shaped. These spots, although being no life-threatening, can be often confused with

melanoma. Actinik keratosis (AK) is produced by prolonged exposure to sunlight and is the most common form of pre-cancer. This kind of lesion tend to appear in face, lips, ears, neck, forearms and back. In its initial states is small in size but it grows gradually becoming large, and also presenting color variation.[12]

2.2. Measurement and selection of morphological features of lesions

Diagnostic applications require a selection of features that must be tailored separately for each problem. The most important point to consider is that the features selected should contain enough information to allow a clear differentiation between classes (benign vs malignant) while being insensitive to irrelevant variability in the inputs. In particular, 3D morphological information of the lesions can be obtained by using 3D measurement systems based on stereovision and fringe projection techniques, and from these, 3D morphological features can be extracted using the appropriate image processing algorithms.

2.3. Classification method: Neural Networks

Artificial Neural Networks (ANN) are a computational based learning algorithm inspired by networks of biological neurons and brain functionalities. ANNs are formed by many nodes, called neurons, ordered in a well-defined layered structure, as shown in Figure 2.

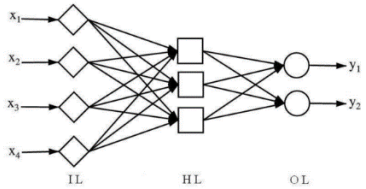

Figure 2. Multilayer feed-forward ANN. From the neurons in the input layer (IL) signals are propagated to the hidden layer (HL) getting finally to the output layer (OL) [10].

Each node is capable of operate simple calculations, both linear or non-linear, and they are interconnected by weighted functions.

During the training process, in which a set of data is given to the network to learn, the weight functions are set into a specific value. The latter performance of the network will completely depend on the configuration achieved during this first training step.

ANNS are a powerful tool since they need only one training, and show tolerance for discontinuity, accidental disturbances or even defects in the training data set. This allows for the usage of ANNs in solving problems which cannot be solved by other means effectively, offering a wide range of applications such as pattern recognition, object classification, medical diagnosis, forecast of economic risk, selection of employees, or approximation of functions [10][11][12]. For all of these reasons, ANNs have been studied in depth over the past 50 years by many different research groups [13], although its application to healthcare is relatively new.

2.3.1 Architecture of Neural Network and learning phase

In this work, we use a two-layer feed-forward neural network with sigmoid hidden and softmax output neurons. The network will be trained with scaled conjugate gradient backpropagation and its performance will be evaluated using cross-entropy and confusion matrices.

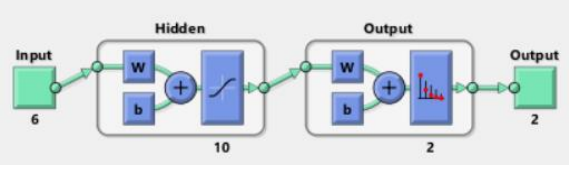

Figure 3. Two-layer feed-forward Neural Network with a sigmoid function in the hidden layer and a softmax function in the output layer.

The number of neurons in the input layer is specified by the type and amount of data which will be given as an input, normally the number of variables affecting the system. The number of output neurons corresponds to the type of answer of the network.

2.3.2 Measures for performance evaluation

Different measures can be applied in order to evaluate the performance of a classification method: *sensitivity*, *specificity*, *positive predictive value*, *negative predicted value*, *ROC curve and confusion matrix*, being all these measurements commonly applied independently of the classification algorithm used.

Table 1. Confusion matrix representation

Real results	Predicted results	
	Positive	Negative
Positive	True Positive (TP)	False Negative (FN)
Negative	False Positive (FP)	True Negative (TN)

A confusion matrix is a very simple, but very visual representation which contains information about actual and predicted classifications performed by a machine learning algorithm, as shown in Table 1. All the rest of measurements that can be extracted from it are specified below:

$$\text{Sensitivity: } (\%) = \frac{TP}{TP+FN} \cdot 100 \qquad (1)$$

$$\text{Specificity: } (\%) = \frac{TN}{TN+FP} \cdot 100 \qquad (2)$$

$$\text{Positive predicted value } (\%) = \frac{TP}{TP+FP} \cdot 100 \qquad (3)$$

$$\text{Negative Predicted value } (\%) = \frac{TN}{TN+FN} \cdot 100 \qquad (4)$$

ROC curves are built by representing the true positive rate or sensitivity versus the false positive rate (100-specificity), so they provide a visual trade-off, comparing model's predictions to true results. Specifically, for the evaluating the performance of a neural network one of the best parameters to consider is the *cross-entropy*, which measures the distance between the real expected results with the ones obtained by the neural network after transforming the initial data into probabilities by using the softmax function, as shown in the formula below:

$$Cross\ entropy = -\sum_i L_i \cdot \log(S_i) \qquad (5)$$

Where S_i are the distribution values obtained for each parameter after applying the softmax function and L_i are the hot-label parameters that we have in our problem, which in this case are 0 for melanoma and 1 for benign lesions.

3. Experimental Description

3.1. Data extraction

The skin lesion data used in this work has been provided by two medical institutions: *Hospital Clinic* (Barcelona, Spain) and *Azienda Hospedaliero Policlinico* (Módena, Italy). The 3D images of the lesions were taken by means of a handheld non-invasive scanning system based on stereophotogrammetry and structured light projection. Phase unwrapping algorithms are applied to the acquired images in order to obtain a dense 3D cloud of points (X, Y, Z). Light intensity is also recorded. The prototype was completely developed at CD6, and designed for imaging a 19x14 mm^2 field of view (FOV) of the skin surface at a working distance of 110 mm, capturing

images with a spatial resolution of 1280x960 pixels with 8-bit resolution, as shown in figures 4 and 5. From these images the 3D topography of the lesion is calculated [14].

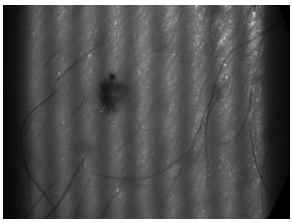

Figure 4. Fringes projected over the skin surface. With this information, after applying unwrapping algorithms, a cloud dense of 3D points is obtained.

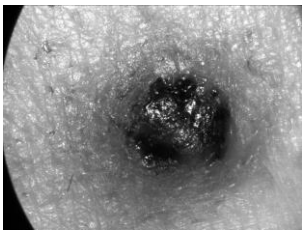

Figure 5. Monochromatic image of the lesion acquired directly with a camera.

The total set of 3D topographies analysed is formed by 69 samples corresponding to two different groups of lesions (BB and MM) as specified in Table 2.

Table 2. Summary of analyzed samples.

Lesion group	Number of samples
BB (benign)	42
MM (melanoma)	27

Regarding the features extracted which were used later on for skin lesion characterization, we have selected a first group related with morphological characteristics of the samples such as *maximum height, medium height, surface (area), volume, diameter and perimeter*.

Data segmentation and feature extraction was performed by using different software tools, including self-developed algorithms in C++ and a commercial software called Mountains Map. Each of the samples were analysed separately. The process involves a first step in which the data is treated, followed by a segmentation algorithm and finally feature extraction algorithms are applied.

3.2. Matlab: Software implementation

After the extraction of features indicated in the previous section, we use them as inputs of the neural network. In order to test the performance of our model different steps have been done:

1. Build the architecture of our network: In this study, the input layer is formed by 6 neurons, each of them corresponding to the selected features per lesion. The output layer is composed by two neurons, being these the two possible categories associated with each input vector. In our case, benign lesions are represented as (1, 0) and melanoma as (0, 1). The optimum value for the number of hidden neurons will be determined experimentally
2. Set-up the division of data: The total data set formed by 69 samples will be divided into a training set, a validation set and a testing set.
3. Training stage using a feed-forward propagation algorithm: During this step, the training set is provided to the neural network. When all the parameters with their responses (benign or malignant classes) are provided, the neural network searches a pattern and learns to differentiate which values of the parameters are more commonly related with melanoma and which ones with the benign lesions. After learning the pattern, the validation test is used by the neural network in order to test its ability to classify. The operation is repeated many times, called epoch, until the response given by the accuracy in recognising the pattern in the validation set reaches a minimum.
4. Network testing: In this step, the testing set is given to the neural network, that after learning, should be able to correctly classify new samples.

5. Extraction of results: Results are given in the form of a confusion matrix showing the sensitivity, specificity, positive predictive value and negative predictive value. Also, we obtain the corresponding ROC curve, and the error trial group, which is associated, this last one, with the performance of the network in the training state.

4. Results and validation

4.1. Pre-processing and image segmentation

For feature extraction the first step consists in applying an image pre-processing algorithm that allow us to substitute the z-component of the point cloud obtained for each sample by its intensity. As a result, we obtain a monochromatic image with the same values of xy. After that, by using this latter obtained image, segmentation is carried out manually for each sample, separating the damaged skin from the background.

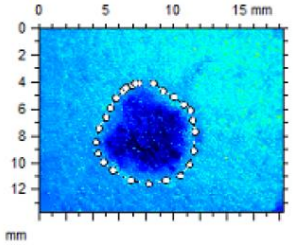

Figure 6. Segmentation performed on a typical monochromatic image.

4.2. Processing

After applying segmentation, the next step is to smooth the surface of the image in order to reduce artefacts resulting from the inherent noise added by the measuring device and by the shining's or hairs present in the patient's skin.

For this purpose, two algorithms are considered, depending in the amount of noise present. For samples with low noise a self-developed code for smoothing the topography based on surface fitting is applied. Samples with high amount of noise are smothered by applying a Fourier Transform Filtering.

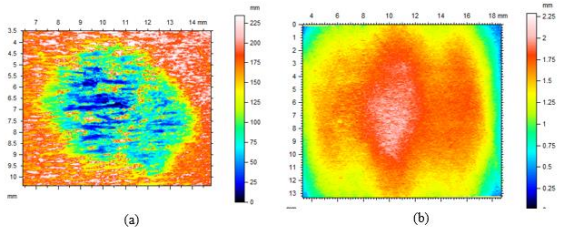

Figure 7. (a) Skin lesion image before the application of a smoothing algorithm and (b) same skin lesion image after applying a smoothing algorithm.

4.3. Feature Extraction

For our purpose, six different morphological features are going to be extracted by analysing the images obtained in the previous sections with a program called *Mountains Map*. Maximum height, medium height, surface, volume, perimeter and diameter of each sample are going to be extracted directly by using some special functions offered by the software

4.3.1 Perimeter and Width

These two features are calculated by using the intensity image. The contour is calculated by hand discriminating the lesion shape from the background. The diameter is also calculated this way searching for the longest distance existing within the lesion shape.

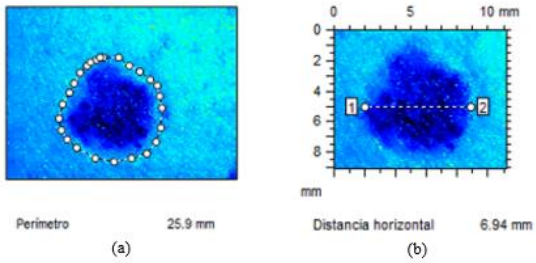

Figure 8. (a) Calculation of the lesion's perimeter and (b) Calculation of the lesion's diameter.

4.3.2 Maximum and mean height

In order to extract these parameters, a profile series conversion is applied. By using this tool, the lesion section and the mean profile are obtained. Maximum and medium height are calculated by applying a step measurement operation, which calculates the distance between peaks and discriminates

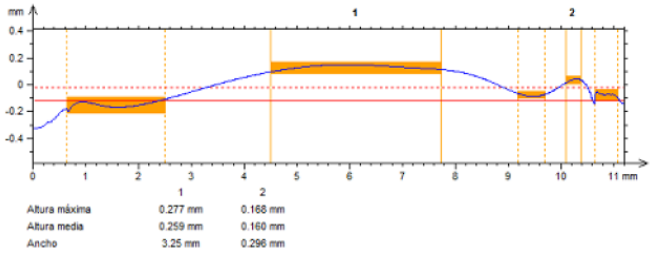

Figure 9. Maximum and Mean Height extraction algorithm.

4.3.3 Surface and Volume

Surface and volume parameters are calculated taking into account the perimeter obtained in the previous section. These two features are both automatically calculated by using the *Volume of peak/hole* operator. For each sample five different values of surface and volume are obtained corresponding to different calculation methods: vertical lines, horizontal lines, mean quadratic plane, polynomial surface of degree 2 and polynomial surface of degree 6. The final result corresponding to each feature is calculated by making the mean between all the values obtained, after excluding the major and minor results.

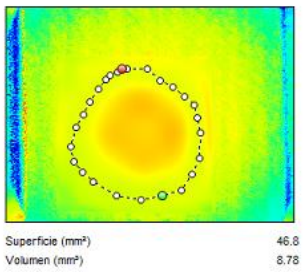

Figure 10. Surface and volume calculation.

4.4. Neural Network fitting and results evaluation

When working with neural networks, a problem of classification changes into a problem of numerical optimization. Finding the best configuration for a neural network that provides us the best results is not easy due to the huge number of parameters and functions that are involved in the calculations. This process of optimization is called fitting. The first typical approach is to determine the best configuration of the data sets and the number of hidden neurons as performed. In order to evaluate the performance of the neural network implemented and find the best parameters that will provide us the best classification results, 9 different models were constructed

by changing the number of hidden neurons between 3, 10 and 50. For each value of the number of hidden neurons, 3 different data set partitions were considered as specified in the table below.

Table 3. Different set partitions considered for a different number of hidden neurons

Number of hidden neurons	Partitions from the hole set		
	Training set (%)	Validation set (%)	Testing set (%)
3, 10, 50	35	35	30
	55	10	35
	70	5	25

Each time the algorithm is run values of the sensitivity, specificity, positive predicted value, negative predicted value, ROC curves and confusion matrix are obtained, as well as the value of the cross entropy.

In order to validate the robustness and repeatability of our method, for each of the models, i.e. when considering 3 hidden neurons and a partition of 55% training set, 10% validation set and 35% testing set, 5 different measurements were performed. So the final result for each parameter is going to be obtained by making the media between 3 of the 5 measurements after discriminating the minor and major values obtained.

Best results were obtained by choosing as a training set a 55% of the hole data set, while taking 10% for the validation set and 35% for the testing set. Results are summarized in table 4

Table 4. Results obtained by dividing the data set as 55% as training set, 10% as validation set and 35% as testing set, for three different number of hidden neurons: 3, 10 and 50.

	NUMBER OF HIDDEN NEURONS		
	3	10	50
Sensitivity (%)	81,2	95,6	97,4
Specificity (%)	37,5	36,7	20,3
PPV (%)	56,9	72,8	57,9
NPV (%)	44,7	66,1	80,0
Cross-entropy error	0,2976	0,2657	0,3237

As we can see in the table above, the values corresponding to the sensitivity are quite high, especially when the number of hidden neurons is increased (from 3 to 50) and the same occurs for the negative predicted value. On the other hand, the specificity shows the opposite behaviour, while the positive predicted value shows different tendencies in all the cases.
For our purposes what we expect is to achieve a good relationship between sensitivity and specificity in order to obtain the maximum area possible in the ROC curve. The best result for this curve is shown in the figure below. It was achieved when using 10 neurons in the hidden layer, and the data set division considered above.

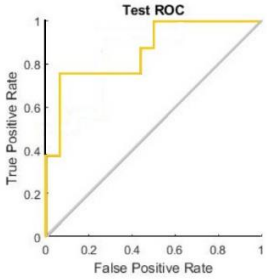

Figure 11. ROC curve obtained for 10 hidden layers, dividing the hole set taking 55% as a training set, 10% as validation set and 35% as testing set.

For the positive predicted value and the negative predicted value, the higher the values the better the performance, as shown in equations (3) and (4). The cross-entropy error obtained is similar in all the cases, the next step will be to continue minimizing it in order to obtain more accuracy and better performance of the neural network.
The obtained results are in perfect concordance with the expectations. Considering the partition set problem, if a huge training set is created and provided to the neural network, leaving small validation and testing sets, the model is not going to learn how to generalize results but it is going to memorize the data set. This phenomenon is called overfitting, and it derives in poor predictive performance, producing the neural network to overreact under minor fluctuations in the training data. The same happens when considering a huge number of neurons in the hidden layer because the number of degrees of freedom increases more than necessary for achieving the minimum error possible given by cross-entropy. [15]
Regarding the distributions of the validation and training set, the optimal configuration is given by a relatively large testing set, approximately the third part of the whole set, in order to obtain reliable results avoiding statistical uncertainties, and a medium size of the validation set. As this last set is devoted to study the performance of the algorithm at the end of each training within an epoch, if the validation set is too small, cross-validation results will not be reliable, and no information about the real performance of the neural network will be acquired.
In our case, no overfitting was detected for the configuration mentioned above, but it was clearly observed for almost all the results obtained when choosing a training set formed by the 70% of the whole data set. [16][17]
Regarding the number of hidden neurons, although it is expected to obtain better results with a small number of neurons in this layer, the results obtained for all the models showed up that the best performance is acquired for a neural network with 10 neurons in the hidden layer, in terms of sensitivity, specificity, positive predictive value, negative predictive value and cross-validation error.
From the results above, we conclude that feed-forward neural networks provide promising results in classifying the potential melanoma. This methodology can be very helpful to the physicians in order to perform faster and better diagnosis.

5. Conclusions

Malignant melanoma is a life-threatening dermatological disease, and its successful treatment relies heavily on early diagnosis. In this work, a new approach to skin cancer diagnosis is developed by combining image processing algorithms, features extraction, artificial neural networks and statistic data classification methods.
After image processing and feature extraction, 9 different neural network models have been tried, using as performance measurement for each one the values corresponding to the sensitivity, specificity, positive predictive value, negative predictive value and cross-entropy error Best results were obtained when dividing the hole data set of features assigning a 55% to the training set, a 10% to the validation set and a 35% to the testing set, and by having 10 neurons in the hidden layer.
Considering the results, the proposed methodology has shown to be very promising with future classification tasks. Neural networks are powerful tools that can be very helpful for both doctors and physicians in order to improve diagnosis accuracy, reducing the time lapse between prognosis and diagnosis.
The future work will be oriented to study the performance of the classification neural network algorithm when having more samples, also to consider new features, related not only with morphological parameters of the lesions and to introduce further improvements on the self-developed neural network algorithm that can yield to more accurate and interesting results.

6. References

[1] Y. Zhou, M. Smith, L. Smith, A. Farooq, and R. Warr, “Enhanced 3D curvature pattern and melanoma diagnosis,” *Comput. Med. Imaging Graph.*, vol. 35, no. 2, pp. 155–165, 2011.

[2] G. R. Day and R. H. Barbour, “Automated melanoma diagnosis: where are we at?,” *Skin Res. Technol.*, vol. 6, no. 5, pp. 1–5, 2000.
[3] S. Mcdonagh, R. B. Fisher, and J. Rees, “Using 3D information for classification of non-melanoma skin lesions,” pp. 9–13, 1995.
[4] V. Mazzarello, D. Soggiu, D. R. Masia, P. Ena, and C. Rubino, “Melanoma versus dysplastic naevi: microtopographic skin study with noninvasive method.,” *J. Plast. Reconstr. Aesthet. Surg.*, vol. 59, no. 7, pp. 700–5, 2006.
[5] H. Skvara, P. Burnett, J. Jones, N. Duschek, P. Plassmann, and J. P. Thirion, “Quantification of skin lesions with a 3D stereovision camera system: Validation and clinical applications,” *Ski. Res. Technol.*, vol. 19, no. 1, pp. 182–190, 2013.
[6] L. N. Smith, M. L. Smith, a. R. Farooq, J. Sun, Y. Ding, and R. Warr, “Machine vision 3D skin texture analysis for detection of melanoma,” *Sens. Rev.*, vol. 31, no. 2, pp. 111–119, 2011.
[7] 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.,” *J. Biomed. Inform.*, vol. 34, no. 1, pp. 28–36, 2001.
[8] Y. Ding, L. Smith, M. Smith, J. Sun, and R. Warr, “A computer assisted diagnosis system for malignant melanoma using 3D skin surface texture features and artificial neural network,” *Int. J. Model. Identif. Control*, vol. 9, no. 4, p. 370, 2010.
[9] A. Sboner, C. Eccher, E. Blanzieri, P. Bauer, M. Cristofolini, G. Zumiani, and S. Forti, “A multiple classifier system for early melanoma diagnosis,” *Artif. Intell. Med.*, vol. 27, no. 1, pp. 29–44, 2003.
[10] M. Antkowiak, “Artificial Neural Networks vs . Support Vector Machines for Skin Diseases Recognition,” *Neural Networks*, 2006.
[11] N. Shafie Pour, M. Saeedi, K. Morteza Semnani, and J. Akbari, “Sun Protection for Children: A Review,” *J. Pediatr. Rev.*, vol. 3, no. 1, pp. 1–7, 2015.
[12] F. Ercal, S. Member, I. A. Chawla, W. V Stoecker, H. Lee, and R. H. Moss, “Neural Network Diagnosis of Malignant Melanoma From Color Images,” vol. 41, no. 9, 1994.
[13] D. J.E. and D. J.M., “Artificial neural networks,” *Cancer*, vol. 91, no. S8, pp. 1615–1635, 2001.
[14] M. Ares, S. Royo, M. Vilaseca, J. A. Herrera-Ramírez, X. Delpueyo, and F. Sanabria, “Handheld 3D Scanning System for In-Vivo Imaging of Skin Cancer,” *Proc. 5th Int. Conf. 3D Body Scanning Technol. Lugano, Switzerland, 21-22 Oct. 2014*, no. October, pp. 231–236, 2014.
[15] S. Lawrence, C. L. Giles, and A. C. Tsoi, “Lessons in neural network training: overfitting may be harder than expected,” *14th Natl. Conf. Artificial Intell.*, pp. 540–545, 1997.
[16] K. K. . Dobbin and R. M. . Simon, “Sample size planning for developing classifiers using high-Dimensional dna microarray data,” *Biostatistics*, vol. 8, no. 1, pp. 101–117, 2006.
[17] A. Raad, A. Kalakech, and M. Ayache, “Breast Cancer Classification Using Neural Network Approach : Mlp and Rbf,” pp. 15–19, 2012.

Acknowledgments
I would like to express my gratitude to my advisor Dr. Santiago Royo and my co-advisor Dr. Miguel Ares for offering me this great learning opportunity and for being always there supervising my progress.
Also, I am deeply grateful to my laboratory colleagues, Sara Peña, Ferrán Sanabria, and Xana Delpueyo, for being always comprehensive, kind and patience, helping me during the whole process of development of the Master Thesis.