

# Statistical techniques applied to the automatic diagnosis of dermoscopic images

Valerio De Vita<sup>1</sup>, Giuseppe Di Leo<sup>2</sup>, Gabriella Fabbrocini<sup>1</sup>, Consolatina Liguori<sup>2</sup>, Alfredo Paolillo<sup>2</sup>, Paolo Sommella<sup>2</sup>

<sup>1</sup> Dermatology Research - University Federico II of Naples, Via S. Pansini 5 80131 Naples, Italy,

<sup>2</sup> Dep. of Industrial Engineering (DIIIn) - University of Salerno, Via Ponte Don Melillo, 84084 Fisciano (SA), Italy

An image based system implementing a well-known diagnostic method is disclosed for the automatic detection of melanomas as support to clinicians. The software procedure is able to recognize automatically the skin lesion within the digital image, measure morphological and chromatic parameters, carry out a suitable classification for the detection of structural dermoscopic criteria provided by the 7-Point Check. Original contribution is referred to advanced statistical techniques, which are introduced at different stages of the image processing, including the border detection, the extraction of low-level features and scoring of high order features (namely dermoscopic criteria). The proposed approach is experimentally tested with reference to a large image set of pigmented lesions.

**Keywords:** Biomedical instrumentation; image processing; diagnosis; classification

**Citation:** Valerio De Vita, Giuseppe Di Leo, Gabriella Fabbrocini, Consolatina Liguori, Alfredo Paolillo, Paolo Sommella, Statistical techniques applied to the automatic diagnosis of dermoscopic images, Acta IMEKO, vol. 1, no. 1, article 6, July 2012, identifier: IMEKO-ACTA-01(2012)-01-06

**Editor:** Pedro Ramos, Instituto de Telecomunicações and Instituto Superior Técnico/Universidade Técnica de Lisboa, Portugal

**Received** January 10<sup>th</sup>, 2012; **In final form** July 6<sup>th</sup>, 2012; **Published** July 2012

**Copyright:** © 2012 IMEKO. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

**Funding:** Information not available

**Corresponding author:** Paolo Sommella, e-mail: psommella@unisa.it

## 1. INTRODUCTION

Malignant melanoma is nowadays one of the leading cancer in the world with a rapidly increasing incidence observed in Australia, America and Europe [1]. Fortunately the curability of skin cancer is very high, if it is treated surgically early enough.

To this aim, epiluminescence microscopy (ELM, also known as dermoscopy) has become an established noninvasive tool for improving the early detection of melanoma accuracy (10% to 30% higher sensitivity, [2] when compared with the clinical diagnosis by the naked eye).

At the same time there has been much scientific endeavor aimed at obtaining an improved and consistent differentiation between benign and malignant melanocytic skin lesions by means of digital dermoscopy analysis.

Computerized dermoscopy image analysis, in fact, adds a quantitative evaluation to the “clinical eye observation” and can be used to improve biopsy decision-making ([3]).

Therefore, different groups have been developing diagnostic systems of recorded images (slides or digital cameras). For example, in [4] an automated melanoma recognition system is proposed taking into account 21 parameters extracted from images. A computer algorithm for the diagnosis of melanocytic lesions based on the evaluation of 64 different analytical

parameters is described in [5], whereas a software module which automatically evaluates the outline of a lesion providing 50 objective parameters subdivided into three categories (geometries, texture and coloured islands) is developed in [6].

A very interesting summary of the main researches about the digital dermoscopy (in terms of recording system, calibration, image datasets, methods and approaches) is reported in [7]. As results of this survey, high accuracy may be achieved by computer aided diagnostic systems employing statistics obtained from low-levels feature and parameters. Nevertheless, it is not likely that the digital system will completely substitute the expert in dermoscopy.

According to the author’s opinion, the automated system should be in fact integrated by higher level features based on a particular diagnostic scheme in order to gain greater clinical acceptance.

More precisely, the software diagnostic system should be able to reproduce the expertise of a well-trained dermatologist and support the clinician in his/her visual inspection and diagnosis according to well-known dermoscopic methods. In detail, three different diagnostic models have become more

Table 1. Dermoscopic criteria and scores according to the 7-Point Check List method.

ELM Criterion	Definition	Score
Atypical Pigment Network	Prominent (hyper-pigmented and broad) network with irregular shape or distribution	2
Blue-whitish Veil	Irregular, confluent, gray-blue to whitish blue pigmentation not associated with red-blue lacunes or maple leaf pigmentation	2
Atypical vascular pattern	Linear, dotted or globular red structures irregularly distributed outside areas of regression	2
Irregular streaks	Radially and asymmetrically arranged linear or extensions at the edge of the lesion	1
Irregular pigmentation	Brown, gray and black areas of diffuse pigmentation with irregular shape or disruption and abrupt end	1
Irregular dots/globules	Black, brown or blue round structures irregularly distributed within the lesion	1
Regression structures	White areas, peppering, multiple blue-gray dots associated	1

widely accepted by clinicians for the interpretation of the features inspected by dermoscopy:

- i. *Pattern Analysis* [8], according to which different diagnostic classes for pigmented skin lesions (such as melanoma, Clark, Spitz/reed, blue nevi) are characterized in terms of specific global patterns and combinations of additional local features;
- ii. the *ABCD-rule* of dermatoscopy [9] which is based on a semi-quantitative assessment by the clinician of the following groups of dermoscopic criteria: asymmetry (A), border (B), color (C) and different (D) structures;
- iii. the *ELM 7-point checklist* ([10]) which is a scoring diagnosis analysis, defining only seven standard dermoscopic structures (each criterion is briefly described in Table I along with the corresponding score). The diagnostic method provides the detection of melanoma when a total score not lower than 3 is achieved summing up the partial scores.

Starting from the previous considerations, the authors have tackled the problem of defining suitable image processing algorithms for the automatic implementation of the *7-Point Check List*.

A preliminary study about the image processing techniques for the extraction of the pigmented lesion (from healthy skin) and the detection of chromatic features was reported in [11].

Further studies [12]-[13] have been led to the introduction of a software framework [14] for the automatic detection of dermoscopic criteria, where the inputs of the Computer Aided System were ELM images and the processing algorithms were derived from the clinical knowledge gained by expert dermatologists (well-trained in the 7-Point Check List application). In [15] a statistical approach is introduced for the automatic detection of a minor criterion (*Irregular dots/globules*).

Following this original trend, the present paper focuses on advanced statistical techniques adopted to improve the diagnostic capability of the automatic tool which implements the 7-Point Check-List. Basic software tasks such as feature identification and high-level classification are deeply investigated with respect to multiple dermoscopic structures. Moreover, the experimental results are extended to a large set of pigmented lesion and a comparison among the different techniques is also carried out.

Thus, the paper is organized as follows: the statistical approach is introduced in Section 2 with reference to the main stages of the image processing. The architecture of the automatic diagnostic system is presented in Section 3, where the algorithms devoted to the detection of five dermoscopic criteria are disclosed in terms of feature extraction and classification.

Experimental results are presented in Section 4, and finally some conclusions are drawn in Section 5.

## 2. METHODS

According to the scheme reported in Figure 1, the software procedure developed for the automatic analysis and diagnosis of dermoscopic images is organized into three main stages.

The *Boundary Detection* (I) allows the pigmented lesion to be extracted from the surrounding healthy skin. The *Feature Extraction* (II) aims to measure morphological and chromatic parameters (low-level features) related to the dermoscopic structures of interest. Finally, at *Feature Classification* (III) each ELM criterion (high-level feature) provided by the 7-Point Check List is detected and scored.

In the following subsections, the statistical techniques and algorithms are introduced correspondingly to the main stages of the automatic procedure.

### 2.1. Boundary detection

Boundary detection is a critical problem in ELM images because the transition between the lesion and the surrounding skin is smooth and hard to detect accurately, even for a trained dermatologist. The algorithm proposed for the skin lesion border extraction consists of three steps:

- i. color to monochrome image conversion;
- ii. image binarization using an adaptive threshold;
- iii. border identification, based on a blob-finding algorithm.

In the first step, 3 different monochrome images are obtained from the source image (RGB standard color) corresponding to the red, green and blue planes. For each component (see Figure 2a), two modes (classes) are typically evident in the pixel intensity histogram (as depicted in Figure 2b) corresponding respectively to the pigmented lesion (the image foreground) and the surrounding skin (the image background). Then, the algorithm introduced by Otsu [16] is adopted to select the optimum threshold  $S^*$  for each histogram, thus allowing the image background and foreground to be detected. In more details, the adaptive algorithm aims to minimize the intra-class variance,

$$\sigma_W^2(S) = P_0(S)\sigma_0^2(S) + P_1(S)\sigma_1^2(S), \quad (1)$$

defined as a weighted sum of variances  $\sigma_i^2$  of the two intensity classes  $C_i$  resulting from the  $S$  threshold

$$P_0(S) = \sum_{k=1}^S \frac{f_k}{N}, \quad P_1(S) = \sum_{k=S+1}^L \frac{f_k}{N}, \quad (2)$$

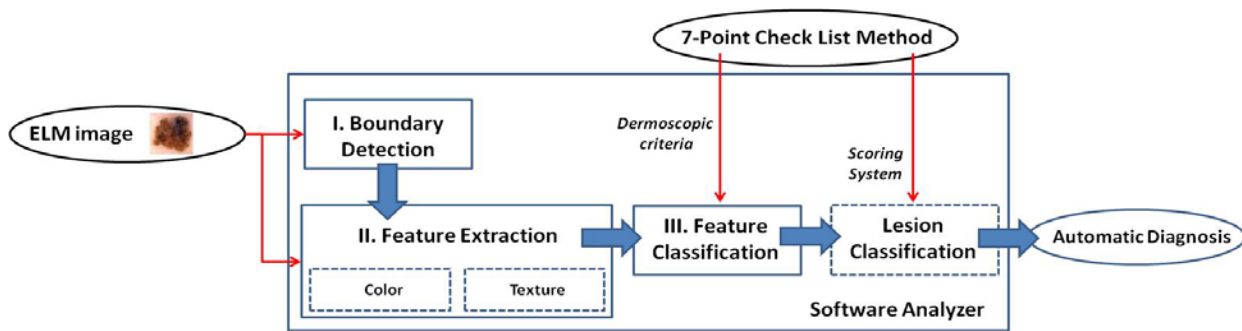


Figure 1. Scheme of the proposed procedure for the automatic diagnosis of Dermoscopic (ELM) Images.

where  $P_i$  is the probability distribution,  $N$  is the number of the image pixels,  $L$  is the number of histogram bins and  $f_k$  the number of pixels associated with  $k$  intensity value.

Otsu shows that minimizing the intra-class variance is the same as maximizing the between-class variance

$$\sigma_B^2(S) = \sigma^2 - \sigma_W^2(S) = P_0(S)P_1(S)[\mu_0(S) - \mu_1(S)]^2 \quad (3)$$

which is expressed in terms of class probabilities  $P_i$  and class means  $\mu_i$  (with  $i = 0, 1$ ).

The adoption of the Otsu's method to RGB colour image leads to three histograms and potentially different threshold values. Since the proposed approach has been experimentally revealed to be more sensitive to surrounding skin (the image background), the thresholding result corresponding to the wider skin lesion area (the image foreground) is considered as binary mask for next processing. An example of result is shown in Figure 2c.

Finally, a simple blob-finding algorithm is adopted to extract the contour of the lesion from the binary mask. According to the modified version of *Moore's Neighbour Contour Tracing* proposed in [17], the tracking algorithm collects and sorts out the contour lines (single pixel width) of the black-white image into an ordered list. In the adopted algorithm, the dependency of reaching the start pixel as stopping criterion (overhead about the redetection of each and every already traced pixel) is removed, allowing a greater efficiency to be achieved. At this point, the border is superimposed on the colour ELM image and displayed for visible inspection to the diagnostician.

## 2.2. Feature extraction

The dermoscopic criteria of interest are characterized both by chromatic and morphological structures. Thus, once the lesion is localized, several first order features are extracted and measured by means of techniques grouped into the following macro-categories:

- color segmentation
- texture analysis

**Color segmentation** – The proposed *segmentation* is based on Statistical Region Merging (SRM), a recent technique [18] belonging to the region growing and merging group. The method models segmentation as an inference problem, in which the image is treated as an observed instance  $I$  of an unknown theoretical image  $I^*$ , whose statistical (true) regions are to be disclosed. This method is typically adopted for its simplicity, computational efficiency, and excellent performance without the use of quantization or color space transformations. In more details, each pixel of the true image  $I^*$  can be modeled as

a set of  $Q$  independent random variables whereas the statistical regions represent theoretical objects sharing a common homogeneity property:

- inside any statistical region the pixels have the same expectation for each color channel (for example Red, Green and Blue);
- the expectation of adjacent regions are different for at least one color channel.

Given the homogeneity property the ideal segmentation of the observed image  $I$  relies on the frontiers between the statistical regions which are connecting pixels with differences in their color expectation.

Figure 3 depicts an example of color segmentation for the

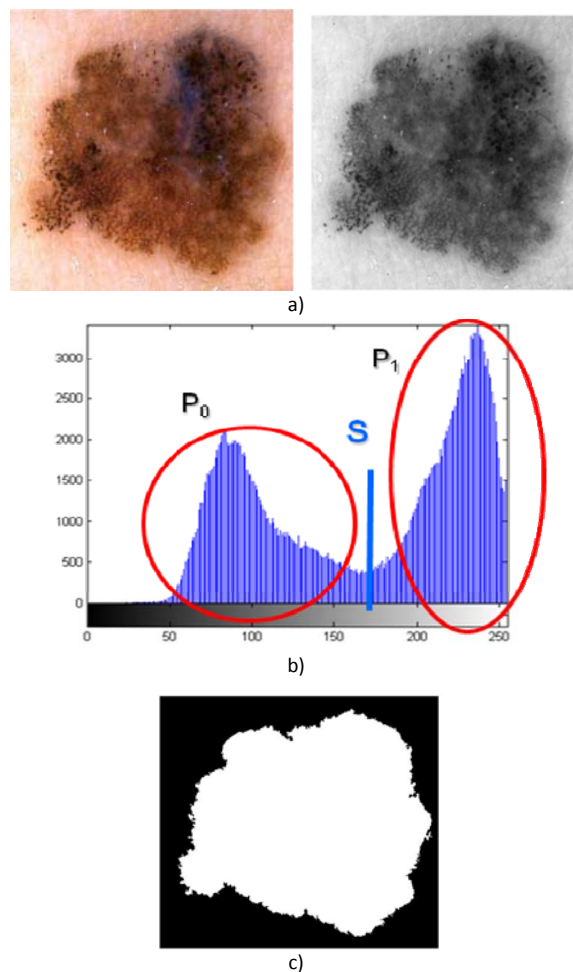


Figure 2. Example of Boundary detection: a) image conversion; b) intensity histogram; c) binary mask.

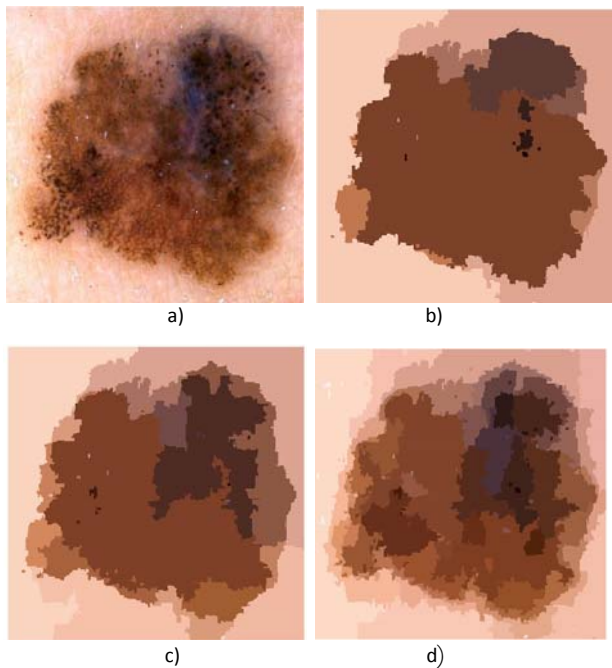


Figure 3. Segmentation using Statistical Region Merging: a) ELM Image; b) results for  $Q = 32$ ; c) results for  $Q = 64$ ; d) results for  $Q = 256$ .

ELM image performed through the SRM: each region is displayed according to its mean RGB values (averaged on pixels constituting the region). The parameter  $Q$  allows to quantify the statistical complexity of  $I^*$ , the generality of the model and finally control the coarseness of the segmentation.

Thus, the lesion map resulting from SRM segmentation can be further investigated in order to determine measurement information about the chromatic features of the local regions (such as components in the main color spaces in terms of mean value and standard deviation, relative difference among neighbors). An example is reported in Figure 4, where the first order (*low-level*) features are extracted by the lesion map resulting from a fine segmentation ( $Q = 256$ ).

**Texture extraction** – As to the search for the occurrence of texture, a combination of two different techniques (*structural* and *spectral methods* [19]) has been designed. As shown in the diagram (Figure 5), the proposed algorithm is arranged into two processing paths, which share the input 8-bit grey-level image extracted from the ELM color image at first stage:

- i. The structural technique is intended to search for primitive structures such as lines and/or points which can constitute a texture. The approach is proposed in order to

identify the main local discontinuities within the image. Thus, the monochromatic image is first compared with its version obtained by a suitable median filter and then a close-opening operation is performed to delete eventual isolated points.

- ii. The spectral technique is based on the Fourier analysis of the grey-level image. It is useful to determine the spatial period of the texture, thus allowing the identification of the only regions where a network exists. More in depth, a sequence of Fast Fourier Transform (FFT), high-pass filtering, Inverse Fast Fourier Transform (IFFT) and suitable thresholding has been adopted. As goal of this path, the local discontinuities which are not clearly associated to a network are disregarded. The result of this phase is a “regions with network” mask to be applied on the image yielded by the structural technique, in order to remove discontinuities which do not actually belong to the pigment network.

Finally the intermediate results from boundary detection (“*lesion*” mask), structural path (“*local discontinuities*” mask), and spectral path (“*regions with network*” mask) are combined according to the AND logic.

As a final result, a “network image” is achieved, where the areas constituting the pigment network are highlighted.

### 2.3. Feature classification

At this stage, each (*high-level*) dermoscopic structure provided by 7-Point Check List is automatically disclosed within the lesion through a suitable classification algorithm, which takes into account the first order (*low-level*) features previously introduced.

More in details, the classification of the chromatic features extracted from the lesion map is viewed as a problem of data mining. In this context a well-known class of solutions is represented by *Decision Tree Classifiers*, which belong to the supervised Machine Learning techniques [20].

A Decision Tree Classifier is a predictive model, trained (or induced) by adopting a suitable dataset with respect to which classification results are already available. More in depth, given a collection of objects (each one described by a set of attributes) a Decision Tree is a graph, wherein each internal node stands for an attribute, each arc toward a child node defines a property related to the parent node and finally a terminal node (or leaf) constitutes a classification result (a single value for the attribute adopted as class discriminator). The paths constituted by internal nodes with a parent-child relationship and the corresponding arcs define the rules of the

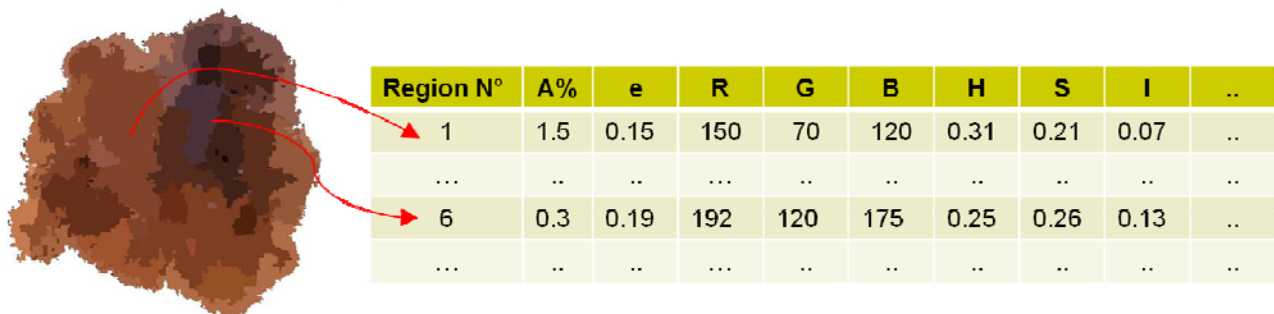


Figure 4. Example of fine color segmentation (SRM,  $Q = 256$ ) and extraction of corresponding (chromatic) feature: segment Area A% (percentage with reference to the lesion area); segment Eccentricity  $e$ , mean value of Red (R), Green (G), Blue (B), Hue (H), Saturation (S), Intensity (I) components (averaged on pixels constituting the segment).

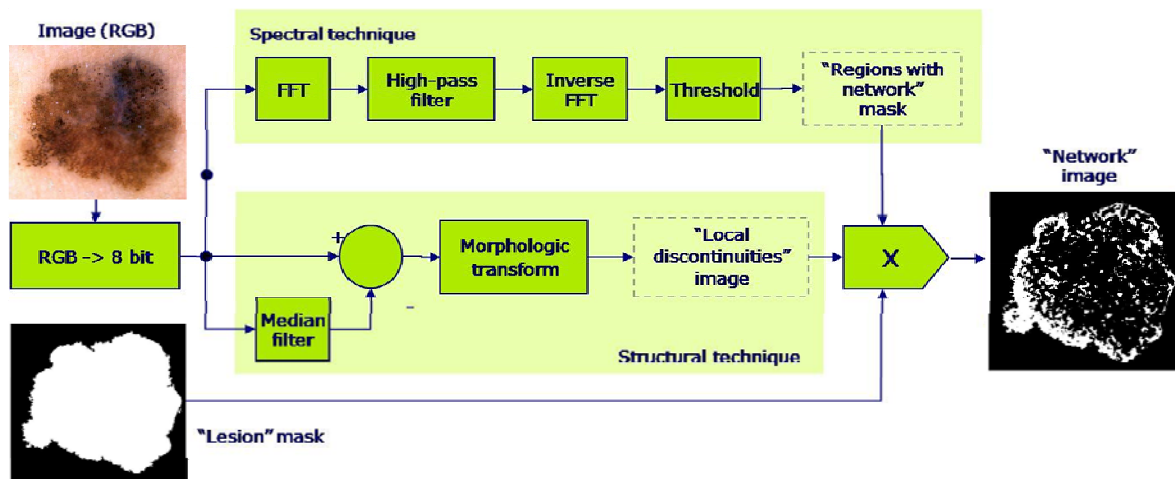


Figure 5. Proposed scheme for Texture Analysis of Pigmented Lesions.

predictive model that can be adopted for classifying new collections of objects.

The Decision Tree Technique can be generally preferred to other solutions (also including Artificial Neural Networks and Support Vector Machines) because Decision Tree Classifiers are often fast to train and apply and generate easy to understand rules. Many induction algorithms have been proposed in literature, which are different for the type (discrete and/or continuous) of attributes they can apply to and the parameter adopted as performance index for the evaluation of the goodness of induction.

Probably the *C4.5* algorithm is the most widely adopted for decision tree induction. It can be related to attributes varying into both discrete and continuous range, whereas the *information gain* (relative entropy or Kullback-Leibler divergence) is considered as leading parameter in the splitting procedure (i.e. identification of a significant attribute and its corresponding optimal value to segment the collection into suitable groups). Moreover the *C4.5* algorithm tries to prevent the *over-fitting condition* by implementing a *pruning strategy*. Given a large training set, in fact, decision tree classifiers could produce rules that perform well on the training data but do not generalize well to unseen data. In particular the *C4.5* is able to identify sub-trees that do not contribute significantly to predictive accuracy and replacing each by a leaf.

Another popular method for classification is instead *linear logistic regression*: it tries to fit a simple (linear) model to the data through a process which generally reveals quite stable, resulting in low variance but potentially high bias. The tree induction exhibits low bias but often high variance because searches a less restricted space of models, allowing it to capture nonlinear patterns in the data, but making it less stable and prone to over-fitting. Consequently a promising way to deal with classification tasks is to use a combination of a tree structure and logistic regression models resulting in a single tree according to the model proposed in [21]. Thus, the Logistic Model Tree (LMT) has been explored as suitable solution for the classification of the feature of interest (the lesion map resulting from the color segmentation).

About the classification of morphological features measured by applying the techniques previously described, a statistical approach is further proposed on the base of Test Hypothesis.

In Figure 6 an example of the feature extraction is reported. The symmetry axes (blue lines) of the lesion are computed as

the major and minor axis of the ellipse characterized by the same normalized second central moments as the region of interest.

Moreover, the main round items are highlighted (red boxes), which could correspond to texture elements of the lesion network and/or isolated dots and globules. These morphological elements can be associated to irregular dermoscopic structures within the lesion (and classified as irregular) if their spatial distribution is not uniform. In the opposite case, i.e. if the observed (spatial) round items were randomly scattered within the lesion, the number of elements in each of 4 quadrants (as resulted from the drawing of the main lesion axes) could be modeled according to the Binomial Distribution.

Therefore, a bilateral Test can be performed to estimate the casual distribution of  $N$  round objects, once the accepted risk  $\alpha$  of Type I Error is fixed. According to the proposed approach, if the paucity or plenty of objects is observed in any quadrant and/or couple of quadrants, the Null Hypothesis (i.e. the spatial symmetry of round items) is refused and the morphological structures are classified as irregular.

### 3. AUTOMATIC DETECTION

The statistical techniques and approaches previously introduced have been adopted to improve the performance of

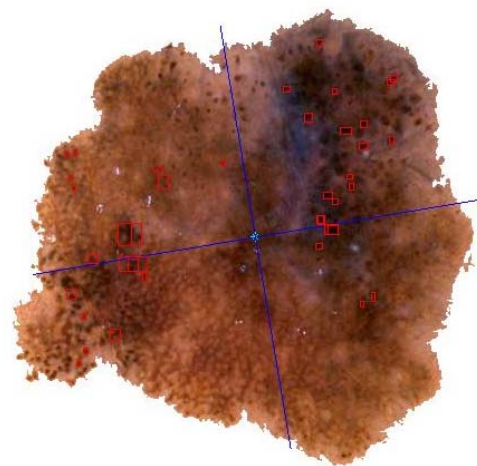


Figure 6. Result of Feature Extraction: detection of round items within the lesion area.

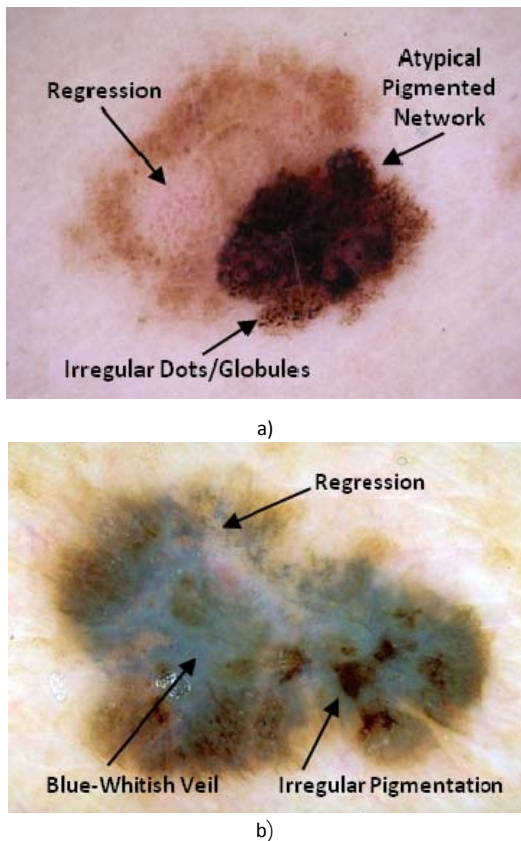


Figure 7. Detection of dermoscopic criteria according to 7-Point Check List.

the algorithms for the automatic detection of five dermoscopic criteria provided by the 7-Point Check List method.

Examples of the diagnosis (by expert dermatologist) about two pigmented lesions are shown in Figure 7, where the dermoscopic structures of interest are highlighted.

More in detail, Figure 7a shows a melanoma (total diagnostic score equal to 4) where a major criterion (*Atypical Pigmented Network*) and two minor criteria (*Regression* and *Irregular Dots/Globules*) are detected. Similarly, within the melanoma displayed in Figure 7b a major dermoscopic criterion (*Blue-whitish Veil*) and two minor structures (*Regression* and *Irregular Pigmentation*) are highlighted.

Hereinafter, two performance indexes are considered to estimate the accuracy of each classification algorithm:

- ✓ *sensitivity*, intended as the ratio of *correct detection* of the high-level structure analyzed and total number of cases where the dermoscopic criterion is present;
- ✓ *specificity*, intended as the ratio of *correct decision* about the high-level structure and total number of cases where the dermoscopic criterion of interest is absent.

The performance indexes range from 0.0 to 1.0 with the ideal classifier characterized by sensitivity and specificity both equal to the maximum value.

### 3.1. Blue-whitish Veil, Irregular Pigmentation and Regression

The approach based on the *Logistic Model Tree* is adopted for the automatic detection of the dermoscopic structures which are more closely dependent on chromatic features. The model can be suitably computed for classify the regions constituting the lesion map which results from the color segmentation. In particular for each region the components of the corresponding pixels in the *RGB*, *HSI* (Hue, Saturation and Intensity) and *CIE Luv* color spaces [19] have been considered to compute mean value and standard deviation as extracted features (vector  $\underline{x}$ ). In addition the area percentage of each region with respect to total area of the lesion is taken into account.

An example of Logistic Model Tree as obtained by training is reported in Figure 8 with reference to the detection of the *Blue-whitish Veil*. As you can see in the scheme (Figure 8a), three different logistic regression models are computed on the basis of three ranges for the Hue mean value of the region to be analyzed (which can be interpreted as corresponding to blue, red or polychromatic “path”). The regression functions  $F_i(\underline{x})$  (with  $i = 1$ , Blue-Veil region and  $i = 2$ , no Blue-Veil region) take into account the standard deviation for the Hue component, the mean and standard deviation for Saturation and Intensity components, in order to determine the probability that the color regions (displayed in Figure 8b) belong to an area characterized by the Blue Veil (the resulting detection of the criterion is depicted in Figure 8c).

Analogous LMT models are computed (with reference to the same feature vector  $\underline{x}$ ) and adopted to classify the color regions (not detected as Blue-whitish Veil) as area of either *Irregular Pigmentation* or *Regression*.

### 3.2. Irregular Dots and Globules

In order to detect the small dark areas of interest, a fine level of color segmentation is required which can be achieved by considering the Statistical Region Merging for high value of  $Q$ .

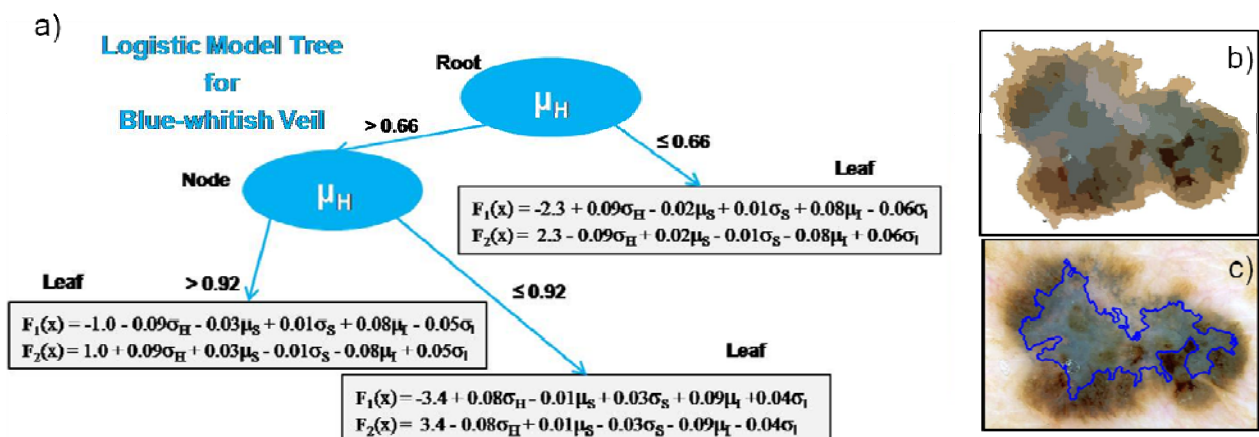


Figure 8. Detection of the *Blue-whitish Veil*: a) Logistic Model Tree; b) lesion map (feature extraction); c) regions classified as Blue-whitish Veil.

As you can easily note in the example reported in Figure 9a, the darkest segments may be deeply investigated to seek for the *structures* which represents Irregular Dots and Globules. More in details, a statistical analysis based on the histogram of the SRM image is adopted by considering and ordering the statistical regions with respect to the increasing value of Intensity value (within a suitable range for Hue component). Moreover the following morphological measurements are also considered to extract rounded items inside the lesion:

- relative dimension  $A\%$ , defined as the number of the region pixel with respect to the lesion area;
- eccentricity  $e$  of the ellipse that has the same second-moments as the region; it is computed as the ratio of the distance between the foci of the ellipse and its major axis length with value between 0 and 1 (the degenerate cases corresponding respectively to a circle and a line segment).

Once the feature extraction is completed with the detection of  $N$  round objects (see the items lightened with respect to the main symmetry axes of the lesion in Figure 9b), according to the statistical approach previously introduced the casual distribution is considered as Null Hypothesis of a bilateral Test. The following thresholds  $k_{I, \min}$ ,  $k_{I, \max}$  and  $k_2$  can be jointly adopted for estimating the irregularity of Dots and Globules so that

$$\sum_{k=0}^{k_{1, \min}} \binom{N}{k} \left(\frac{1}{4}\right)^k \left(\frac{3}{4}\right)^{N-k} + \sum_{k=k_{1, \max}}^N \binom{N}{k} \left(\frac{1}{4}\right)^k \left(\frac{3}{4}\right)^{N-k} \leq \alpha \quad (4)$$

and

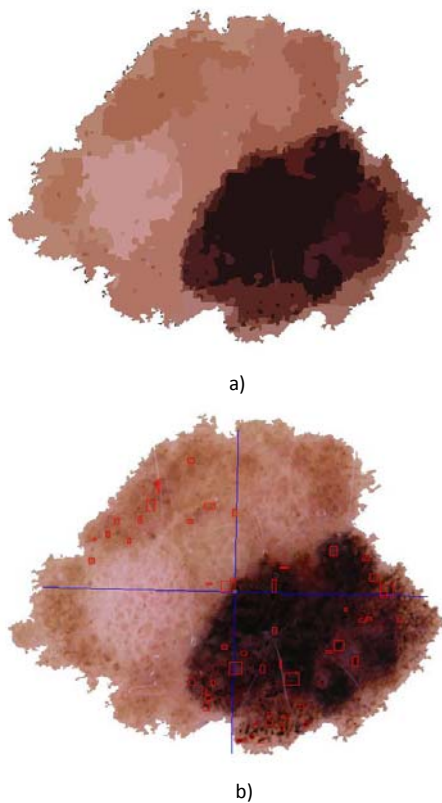


Figure 9. Detection of *Irregular Dots and Globules*: a) colour segmentation (SRM); b) feature extraction: rounded items for Hypothesis Test.

$$2 \sum_{k=0}^{k_2} \binom{N}{k} \left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{N-k} \leq \alpha, \quad (5)$$

where  $\alpha$  is the accepted risk of Type I Error.

When the round items observed in each quadrant exceed the calculated thresholds, the Null Hypothesis is refused and the corresponding Dots and Globules are classified as Irregular.

### 3.3. Atypical Pigmented Network

The approach based on the Hypothesis Test is also adopted for the classification (in terms of spatial irregularity) of the results from the feature extraction.

As introduced in the previous section, the pigmented network within the lesion of interest is detected suitably combining the texture extraction and the color segmentation.

More in details, the areas constituting the network (white objects in Figure 10a) are matched with the darkest regions of the lesion map computed through the Statistical Region Merging (see Figure 10b). Then, the  $N$  objects resulting from the coupled analysis are classified as irregular ("atypical") by performing the Hypothesis Test according to the Eq. (4)-(5). In the example reported in Figure 10c, the major criterion is detected in the darkest right-bottom area where the pigmented network is mainly distributed.

## 4. EXPERIMENTAL RESULTS

In order to develop and test the automatic procedure for the diagnosis of pigmented skin lesions, images of benign and malignant lesions were collected and stored in a *database* (200 cases were extracted from a dermoscopy atlas [22], whereas about 100 images were available from a screening activity carried out at the Dermatology Section of University "Federico II" in Naples).

For each image, the corresponding clinical and histological analysis (when available) has been considered as well as the 7-Point Check List score computed by a group of three expert physicians. About the image quality, all the pictures are 24-bit RGB color images in JPEG format with dimensions ranging from 700x447 to 2272x1520 pixels. The lesions are imaged completely with healthy skin visible at margins. Moreover, as image pre-processing for artifact removal the strategy based on mathematical morphology has been adopted as suggested in [23].

### 4.1. Boundary detection

The proposed technique based on the Adaptive Thresholding has been compared with an unsupervised approach based on the Statistical Region Merging (SRM) algorithm, which was revealed the most effective method [24] for contour detection in dermoscopy images of pigmented skin lesions. Comparison has taken into account 120 dermoscopy images (60 invasive malignant melanoma and 60 benign) randomly selected from the starting dataset.

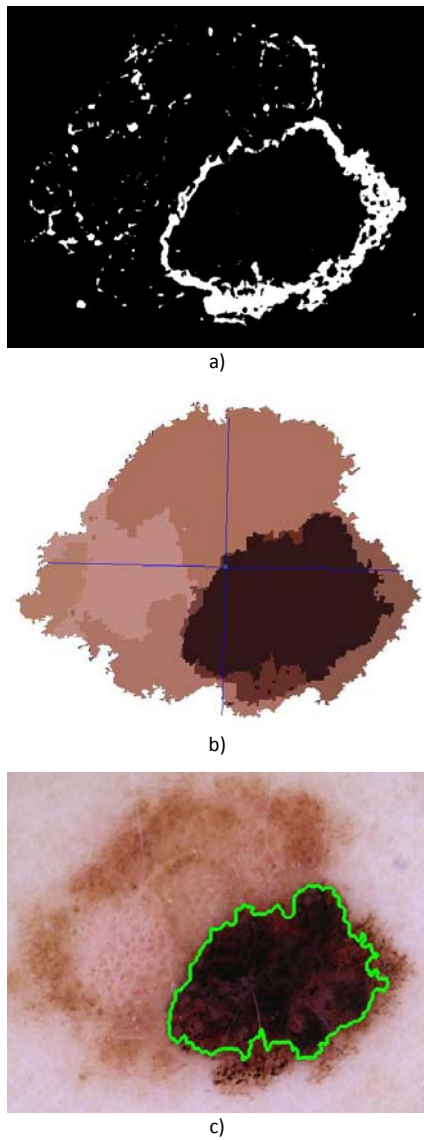


Figure 10. Detection of *Atypical Pigmented Network*: a) texture extraction; b) colour segmentation; c) classification results.

As a ground truth for the evaluation of the border detection error, a manual border was obtained by selecting a number of points on the lesion border, connecting these points by a second-order B-spline and filling the resulting closed curve. More in detail, three dermatologists were asked to select the points on the lesion border, and then the corresponding binary images were suitable combined. A majority policy is taken into account: only the image pixels resulted as inner points of the lesion by at least two dermatologist are considered as white-value pixels of the ground truth binary image (*Ref\_Binary*). Finally the tracing contour algorithm [17] is applied to determine the ground truth manual border.

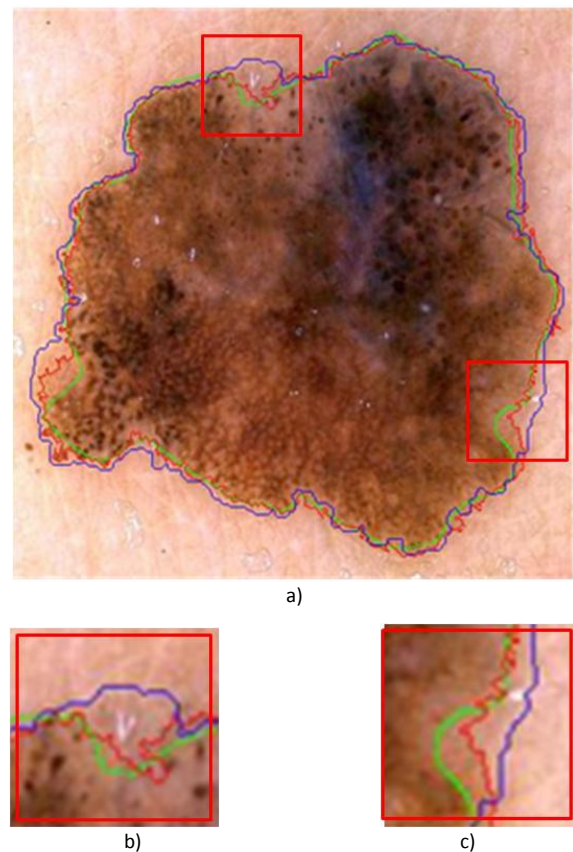


Figure 11. Comparison between automated procedures for border detection: ground truth (green line), adaptive thresholding (red line), unsupervised approach (blue line).

Using the dermatologist-determined borders, the automatic borders resulting from the Adaptive Thresholding and SRM have been compared using the metric suggested in [25]. Here, the percentage border error is given by:

$$Border\_Error = \frac{(Automatic\_Binary) XOR (Ref\_Binary)}{Area(Ref\_Binary)} 100\% \quad (6)$$

where *Automatic\_Binary* is the binary image obtained by filling the computer detected border, the exclusive-OR operation gives the pixels for which the *Automatic\_Binary* and *Ref\_Binary* disagree, and *Area(I)* denotes the number of pixels in the binary image *I*.

Table 2 shows the mean and standard deviation border error for the automated methods considered. Although the error rates increase in the melanoma group (due to the presence of higher border irregularity and color variegation in these lesions), the proposed approach has achieved the best results (lowest error values) in terms of both accuracy (mean) and consistency

Table 2. Dermoscopic criteria and scores according to the 7-Point Check List method.

Segmentation Technique	Benignant Lesions		Melanoma		All Lesions	
	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev
Adaptive Statistical Thresholding	7.4%	2.3%	10.0%	6.2%	8.7%	4.8%
Unsupervised Approach (SRM)	8.5%	3.3%	13.1%	8.7%	10.8%	6.9%



(standard deviation). An example of automatic contour extraction for a melanoma is reported in Figure 11a, where the resulting borders are compared with the manual border.

As you can see in the reported details (Figure 11b-c), the automatic border resulting from the Adaptive Thresholding is able to better match the manual border than the result from SRM. The threshold which takes into account the image as whole is, in fact, able to separate lesion and surrounding healthy skin also in critical local regions where the pixels components in RGB space are statistically close.

#### 4.2. Feature extraction and classification

As preliminary step a *Training* and *Test Set* have been suitably selected from the reference database for each dermoscopic criterion.

As guideline, the Training and Test Set have to share the same case distribution with respect to the criterion of interest. For example, 150 digital images has been adopted to develop the automatic detection of *Irregular Dots/Globules*, whereas the remaining 137 images have been adopted as Test Set to verify the software procedure.

As result, the Training and Test Set include respectively 45 and 39 skin lesions characterized by dermoscopic structures of interest.

About the color segmentation, a comparison has been carried out between the Statistical Region Merging (controlling the coarseness by varying  $Q$  from 32 to 256) and the Multi-Threshold approach (proposed in [14]) based on Principal Component Analysis [26] and 2D - histogram.

The classification results from physicians have been taken into account: 3 expert dermatologists were asked to inspect the results (lesion map) from the color segmentation in order to set the classification attribute for the regions of each image belonging to the Training Sets.

About the detection of Blue-whitish Veil, Irregular Pigmentation and Regression, multiple Logistic Model Trees (correspondingly to the different color segmentation) have been induced from the Training Set and verified (in terms of classification performance) with respect to the Test Set.

About the detection of Atypical Pigmented Network Irregular Dots/Globules, suitable thresholds have been determined from the image properties of the Training Set through ROC curves [27] for the quantities introduced in the feature extraction stage (maximum region dimension  $A\%$ , and eccentricity  $e$ , range for  $I$  component) as well as the classification (minimum number  $N_0$  of round items to perform the statistical test and the risk  $\bar{a}$ ).

As an example, the verification of the proposed approach about Irregular Dots/Globules with respect to the previously

Table 3. *Irregular Dots/Globules*: Classifier Performance (SRM,  $Q = 256$ ).

Image Set	Sensitivity	Specificity
Training Set	80%	82%
Test Set	90%	85%

introduced Test Set has resulted in 35 skin lesions correctly scored (with respect to 39 cases where the minor criterion was present).

Moreover, the classifier lead to 15 false detections (automatic score = 1). Table 3 summarizes the corresponding results (both for Training and Test Set) in terms of *sensitivity* and *specificity*. The overtraining has been avoided: similar performance of the classifier have been achieved for the two Image Sets.

Analogous results have been achieved for the detection of all dermoscopic criteria of interest.

Table 4 summarizes the performance indexes of diagnostic algorithms with reference to the images including into the corresponding Test Sets.

Goal of the comparison has been the evaluation of the color segmentation approach (between the SRM technique and Multi-threshold) that better allows to highlight the first-order features, on which the classification of the dermoscopic criteria are based. Thus, for each high-level feature, the performance corresponding to the best segmentation technique are reported in bold.

As you can easily note, the Statistical Region Merging has been revealed as the preferred solution for color segmentation.

More in details, a quite coarse segmentation ( $Q = 64$ ) is able to disclose the areas characterized by *Atypical Pigmented Network*, *Irregular Pigmentation* and *Regression*. Namely a satisfactory sensitivity (not inferior than 0.80) is achieved without downgrading the specificity (which has to be in special account for the minor criteria).

The finest segmentation ( $Q = 256$ ) has to be preferred (in terms of sensitivity) for the detection of the reduced-size objects (otherwise not revealed) which can be classified as *Irregular Dots/Globules*.

Finally, the Multi-Threshold approach is the segmentation technique able to better identify the large areas of the lesions characterized by *Blue-whitish Veil*.

An example of correct detection of the dermoscopic structures of interest is reported in Figure 12, where the experts' diagnosis about three melanomas is compared with the results of the classification algorithms.

Table 4. Automatic Diagnosis of pigmented lesions: comparison among classification results (Test Sets).

ELM Criterion	Multi-Threshold		SRM ( $Q = 64$ )		SRM ( $Q = 128$ )		SRM ( $Q = 256$ )	
	<i>Sensitivity (%)</i>	<i>Specificity (%)</i>	<i>Sensitivity (%)</i>	<i>Specificity (%)</i>	<i>Sensitivity (%)</i>	<i>Specificity (%)</i>	<i>Sensitivity (%)</i>	<i>Specificity (%)</i>
Atypical Pigment Network	80	82	<b>82</b>	<b>85</b>	85	78	87	74
Blue-whitish Veil	<b>90</b>	<b>93</b>	66	92	73	90	84	85
Irregular Pigmentation	84	77	<b>85</b>	<b>81</b>	93	66	97	63
Irregular Dots/Globules	-	-	-	-	87	88	<b>90</b>	<b>85</b>
Regression structures	72	75	<b>80</b>	<b>83</b>	83	70	85	67

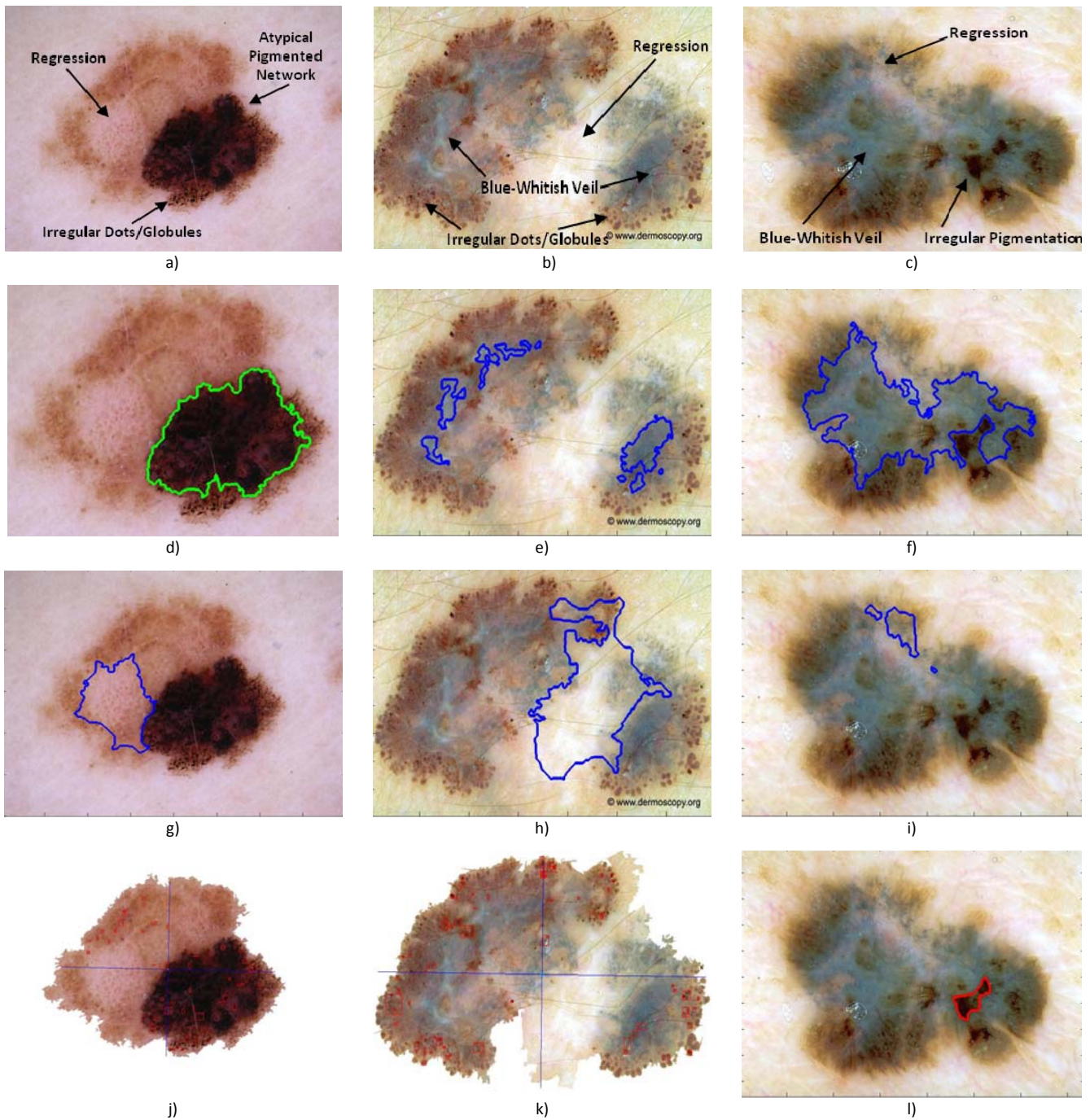


Figure 12. Analysis of Melanomas: a) - c) expert's diagnosis ; d) automatic detection of *Atypical Pigmented Network* ; e) - f) automatic detection of *Blue-whitish Veil*; g) - i) automatic detection of *Regression*; j) -k) automatic detection of *Irregular Dots/Globules*. ; l) automatic detection of *Irregular Pigmentation*.

Figure 13 reports some typical examples of incorrect and/or misleading results of the proposed approach which limit the performance indexes of the automatic procedure (sensitivity and specificity < 1.0).

More in details, sensitivity and specificity respectively for *Atypical Pigmented Network* and *Irregular Pigmentation* are the lowest values because these structures are sometimes confused by the automatic procedure (see Figure 13b-f).

## 5. CONCLUSIONS

The paper has described a statistical approach to be originally adopted for the automatic diagnosis of digital dermoscopic images according to the 7-Point Check List

method. The proposed techniques are suitable to carry out: *i)* the detection of the lesion contour; *ii)* the extraction and measurement of the main low-level features (both chromatic and morphological) within the pigmented lesion; *iii)* the classification and scoring of the high-level features (namely the automatic detection of search for *Atypical Pigmented Network*, *Blue-whitish Veil*, *Irregular Pigmentation*, *Regression* and *Irregular Dot/Globules*).

The automatic procedures have been tested with respect to a quite extensive metrological characterization (performance of each classifier estimated in terms of the sensitivity and specificity) and revealed to be satisfactory as software tool supporting the physician.

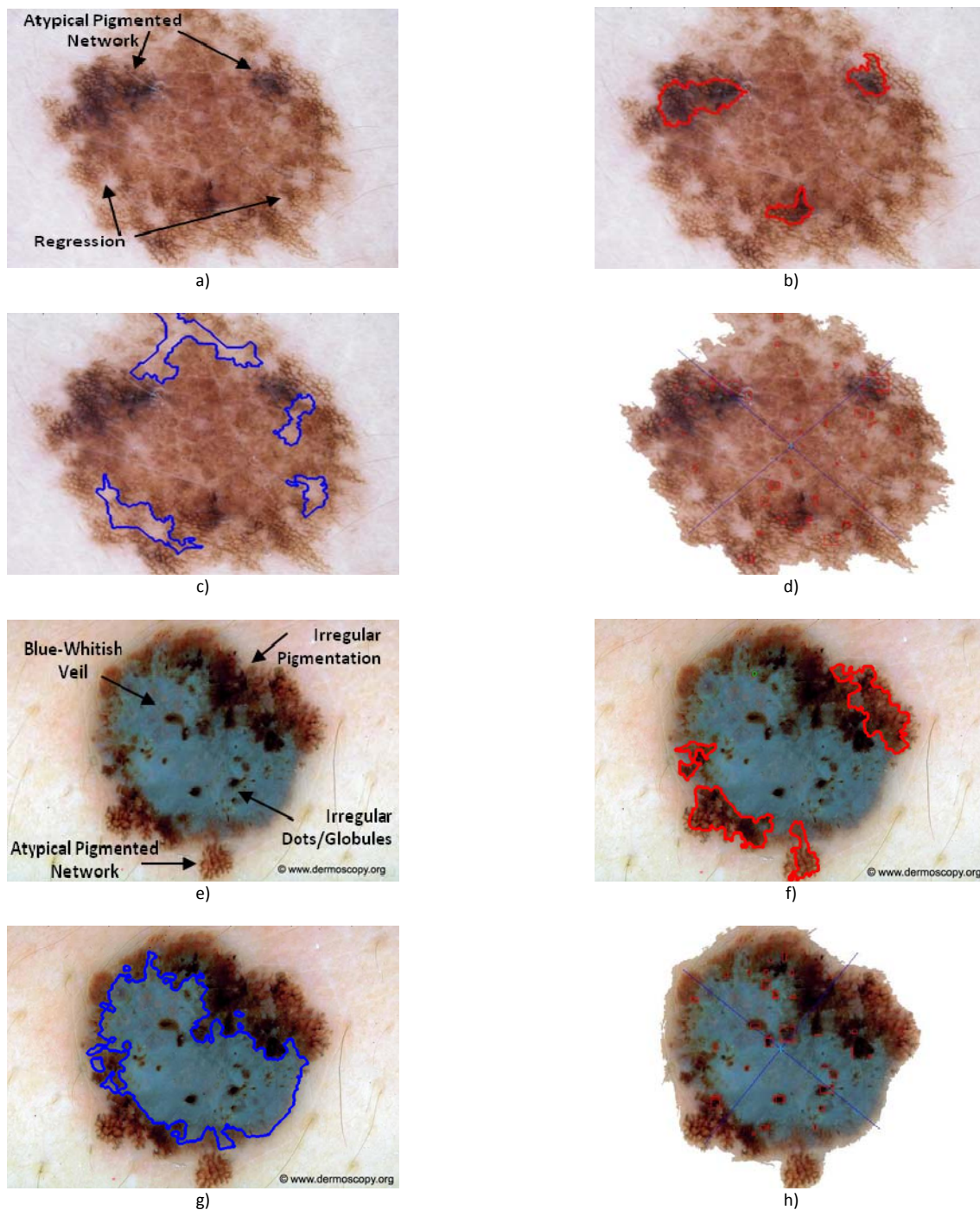


Figure 13. Analysis of Pigmented Lesions: a) - e) expert's diagnosis; b) - f) misleading detection of *Irregular Pigmentation*; c) correct detection of *Regression*; d) false detection of *Irregular Dots/Globules*; - g) correct detection of *Blue-whitish Veil*; h) correct detection of *Irregular Dots/Globules*.

Therefore, the approach will be extended and suitably integrated with other procedures (for *Atypical Vascular Pattern* and *Irregular Streaks*) in order to achieve the image-based measurement system for the diagnosis of melanoma.

Starting from the results of the present study, the research efforts will be addressed to:

- (i) deeply investigate the correlation existing among the seven dermoscopic criteria;
- (ii) compute a confidence level for each intermediate classification (for example based on a multi-resolution segmentation and a Fuzzy fusion of the corresponding

results), in order to improve sensitivity and specificity of the software system as whole;

- (iii) carry on an intensive measurement campaign and compare the diagnosis from the automatic system to the results from the interactive adoption of the software tool by two groups of physicians (respectively expert and not acquainted with the 7-Point Check List) in order to estimate the improvements in the daily clinical practice of dermatologists.

## REFERENCES

- [1] R.Marks, "Epidemiology of melanoma", *Clin. Exp. Dermatology* vol. 25, 2000, pp. 459-463.
- [2] J.Mayer, "Systematic review of the diagnostic accuracy of dermatoscopy in detecting malignant melanoma", *Med. Journal Aust* vol. 167, 1997, pp. 206-210.
- [3] M.Burroni, R.Corona, G.Dell'Eva et al., "Melanoma computer-aided diagnosis: reliability and feasibility study", *Clin. Cancer Res.* Vol. 10, 2004, pp. 1881-1886.
- [4] H.Ganster, A.Pinz et al., "Automated melanoma recognition", *IEEE Trans. Med. Imaging* vol. 20, 2001, pp. 233-239.
- [5] A.Blum, H.Luedtke et al., "Digital image analysis for diagnosis of cutaneous melanoma. Development of a highly effective computer algorithm based on analysis of 837 melanocytic lesions", *British Journal of Dermatology*, vol. 151, 2004, pp. 1029-1038.
- [6] P.Rubegni, G.Cevenini, M.Burroni et al., "Automated diagnosis of pigmented skin lesions", *International Journal of Cancer*, vol. 101, 2002, pp. 576-580.
- [7] A.Blum, I.Zalaudek, "Digital image analysis for diagnosis of skin tumors", *Semin Cutan Med Surg.*, vol. 27, 2008, pp.11-15.
- [8] H.Pehamberger, A.Steiner, K.Wolff, "In vivo epiluminescence microscopy of pigmented skin lesions: Pattern analysis of pigmented skin lesions", *J. Am. Acad. Dermatol.* vol. 17, 1987, pp. 571-583.
- [9] W.Stolz, O.Braun Falco et al., "ABCD rule of dermatoscopy: a new practical method for early recognition of malignant melanoma", *European Journal of Dermatology*, vol. 4, 1994, pp. 521-527.
- [10] G.Argenziano et al., "Epiluminescence microscopy for the diagnosis of doubtful melanocytic skin lesions: comparison of the ABCD rule of dermatoscopy and a new 7-point checklist based on pattern analysis", *Arch Dermatol.* vol. 134, 1998, pp. 1563-1570.
- [11] G.Di Leo, G.Fabbrocini et al., "ELM image processing for melanocytic skin lesion based on 7-point checklist: a preliminary discussion" *Proc. of the 13<sup>th</sup> IMEKO TC-4 Symposium*, Athens, Greece, Vol.2, 2004, pp. 474-479.
- [12] G. Betta, G. Di Leo et al., "Dermoscopic image-analysis system: estimation of atypical pigment network and atypical vascular pattern", *Proc. of Intern Work. on Medical Measurement and Applications (MeMeA)*, 2006, pp. 63-67.
- [13] G. Di Leo, C. Liguori et al., "An Improved procedure for the automatic detection of dermoscopic structures in digital ELM images of skin lesion", *IEEE Conference on Virtual Environments, Human-Computer Interfaces and Measurement Systems 2008*, 14-16 July 2008, Istanbul (Turkey), pp. 190-194.
- [14] G.Di Leo, G.Fabbrocini et al., "Automatic Diagnosis of Melanoma: a Software System based on the 7-Point Check-List", *Proceedings of the 43rd Annual Hawaii International Conference on System Sciences*, January 5-8, 2010, Computer Society Press.
- [15] V.De Vita, G.Di Leo et al., "Statistical Image Processing for the Detection of Dermoscopic Criteria", *Proc. of XVIII IMEKO TC-4 Symposium*, Sept. 27-30, 2011, Natal, Brazil.
- [16] Otsu, "A threshold selection method from gray-level histogram", *IEEE Trans. on System Man Cybernetics*, Vol. SMC-9, no. 1, 1979, pp. 62-66.
- [17] R. Pradhan, S. Kumar et al., "Contour Line Tracing Algorithm for Digital Topographic Maps", *International Journal of Image Processing (IJIP)*, vol. 4, Issue 2, 2010, pp. 156-163.
- [18] R.Nock, F.Nielsen, "Statistical Region Merging". *IEEE Trans on Pattern Analysis and Machine Intelligence*, vol. 26, no. 11, 2004, pp. 1452-1458.
- [19] R.C.Gonzalez, R.E.Woods, "Digital Image Processing Prentice Hall, New Jersey, 2002.
- [20] H.Witten, E.Frank, "Data Mining: Practical Machine Learning Tools and Techniques", Morgan Kaufmann, San Francisco, 2005.
- [21] N.Landwehr, M.Hall, E. Frank, "Logistic Model Trees", 14<sup>th</sup> Europ. Conference on Machine Learning, 2003.
- [22] G.Argenziano, H.P.Soyer, V.De Giorgi et al., *Interactive Atlas of Dermoscopy*, EDRA Medical Publishing & New Media, Milan, Italy, 2002.
- [23] P.Schmid, J.Guillod, J.P.Thiran, "Towards a computer-aided diagnosis system for pigmented skin lesions", *Computerized Medical Imaging and Graphics*, vol. 27, 2003, pp. 65-78.
- [24] M.E.Celebi et al., "Border detection in dermoscopy images using statistical region merging", *Skin Research and Technology* vol. 13, no. 4, 2008, pp. 347-353.
- [25] G.A.Hance et al., "Unsupervised color image segmentation with application to skin tumor borders", *IEEE Eng. in Medicine and Biology*, vol. 15, no. 1, 1996, pp. 104-111.
- [26] A.Levy, M.Lindenbaum, "Sequential Karhuen-Loeve basis extraction and its Application to images", *IEEE Trans. on Image Processing* vol. 9, 2000, pp. 1371-1374.
- [27] C. M. Florkowski, "Sensitivity, Specificity, Receiver-Operating Characteristic (ROC) Curves and Likelihood Ratios: Communicating the Performance of Diagnostic Tests", *Clin Biochem Rev.* 2008, pp. 83-87.