

Automatic Diagnosis of Melanoma Based on the 7-Point Checklist

Gabriella Fabbrocini, Valerio De Vita, Sara Cacciapuoti,
Giuseppe Di Leo, Consolatina Liguori, Alfredo Paolillo,
Antonio Pietrosanto and Paolo Sommella

Abstract 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 detecting the dermoscopic structures provided by the 7-Point Checklist. Advanced techniques are introduced at different stages of the image processing pipeline, including the border detection, the extraction of low-level features and scoring of high order features.

Keywords Melanoma · Pigmented lesions · Dermoscopy · Image analysis · Semi-automatic diagnosis

Introduction

Malignant melanoma is currently one of the leading cancers among many fair skinned populations around the world. Change of recreational behavior together with the increase in ultraviolet radiation have caused dramatic increase in the number of melanomas diagnosed [1–3].

Currently, the possibility to increase the accuracy of diagnosis of melanoma is one of the most important tools to reduce the mortality rate of this tumor. The aim

G. Fabbrocini · V. De Vita · S. Cacciapuoti · G. Di Leo
Dermatology Research, University Federico II of Naples, Via S. Pansini 5, 80131 Naples, Italy

G. Di Leo · C. Liguori(✉) · A. Paolillo · A. Pietrosanto · P. Sommella(✉)
Department of Industrial Engineering (DIIn), University of Salerno, Via Giovanni Paolo II 132,
84084 Fisciano, Sa, Italy
e-mail: tliguori@unisa.it

P. Somella
e-mail: psommella@unisa.it

of prevention campaigns is to increase public awareness of early warning signs. The possibility of new diagnostic methodologies and algorithms can contribute to perform an earlier diagnosis and to reduce the metastatic risk. Investigations have shown that the curability rate of thin melanomas (<1 mm) is between 91.8 and 98.1 % [4]. Epiluminescence microscopy (ELM) is an *in vivo*, non invasive technique that has disclosed a new dimension of the clinical morphologic features of pigmented skin lesions, using various incident light magnification systems with an oil immersion technique [2, 3]. Results of previous studies demonstrated that ELM improves accuracy in diagnosing pigmented skin lesions from 10 to 27 % when compared to clinical diagnosis by the naked eye [5].

Three diagnostic models with similar reliability have become more widely accepted by clinicians: 1. pattern analysis, which is based on the “expert” qualitative assessment of numerous individual ELM criteria; 2. the ABCD-rule of dermoscopy which is based on a semi-quantitative analysis of the following criteria: asymmetry (A), border (B), color (C) and different dermoscopic (D) structures; 3. the ELM 7-Point Checklist scoring diagnosis analysis, proposed by Argenziano et al. defining seven standard ELM criteria: Atypical pigment network, Blue-whitish veil, Atypical vascular pattern, Irregular streaks, Irregular pigmentation, Irregular dots/globules, Regression structures. The ELM 7-Point Checklist provides a simplification of standard pattern analysis and, if compared to ABCD, allows less experienced observers to achieve higher diagnostic accuracy values [6].

The 7-Point Checklist

The 7-point checklist is a diagnostic method that requires the identification of only 7 dermoscopic criteria to help clinicians to use dermoscopy. This simplified algorithm has been shown to be reproducible with non-expert dermatologists, who were able to classify a high percentage of melanomas [6].

In the original paper on the 7-point checklist, dermoscopic images of melanocytic skin lesions were studied to evaluate the incidence of 7 standard criteria. These features were selected for their frequent association with melanoma.

The 7 standard criteria are briefly defined in Table 1 along with the corresponding histological correlates and the scoring system (the 3 major criteria have a score of 2 points and the 4 minor criteria a score of 1 point). The differences between melanomas and nevi were evaluated by a univariate statistical test and the significant variables were used for stepwise logistic regression analysis to determine their diagnostic weights in the diagnosis of melanoma, as expressed by odds ratios. Using the odds ratios calculated with multivariate analysis, a score of 2 was given to the 3 criteria with odds ratios >5, termed “major” criteria, and a score of 1 to the 4 criteria with odds ratios <5, termed “minor” criteria. The total score for the lesion is obtained by simple addition of the individual scores for each detected criterion. In order to diagnose a melanoma, the identification of at least 2 melanoma-specific dermoscopic criteria is required, i.e. a minimum total score of 3 (1 major plus 1 minor or 3 minor

Table 1 Dermoscopic criteria and scores according to the 7 point checklist 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

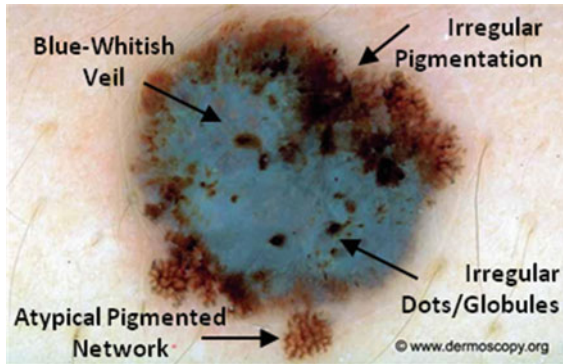


Fig. 1 Application of the 7-point checklist to the digital ELM image of a pigmented skin lesion: blue-whitish veil (score 2) + atypical pigment network (score 2) + irregular pigmentation (score 1) + irregular dots/globules (score 1) = total score 6. In order to diagnose a melanoma, the identification of at least 2 melanoma-specific dermoscopic criteria is required: in fact, a minimum total score of 3 is required (1 major plus 1 minor or 3 minor criteria). Histology confirmed the diagnosis of melanoma. Source of this figure: www.dermoscopy.org

criteria). Figure 1 shows a diagnosis example where different structures are present and scores corresponding to melanomas are computed.

Like most other diagnostic algorithms, the 7-point checklist was developed and validated retrospectively on a set of 342 dermoscopic images of histologically proven

melanocytic skin lesions (the sensitivity and specificity of melanoma detection were 95 and 75 %, respectively). Argenziano et al. showed that the ELM 7-point checklist, in the hands of experienced observers, gave the greatest sensitivity value (95 %), especially in the subgroup of early cutaneous melanoma. Compared with overall ELM diagnosis, the specificity was lower (75 vs. 90 %) because of the tendency to overclassify melanocytic nevi (especially the atypical types) as melanomas with the scoring diagnostic systems. A decrease in specificity may result in some increase in biopsy examinations of benign lesions, but the increase in sensitivity would decrease the chances of missing melanomas. The authors designed a model that requires the identification of only 7 standard ELM criteria, thus enabling even the less experienced clinician to use the method. In fact, this simplified scored pattern analysis was shown to be reproducible not only with a test set performed by experts but also by less experienced dermatologists, who were able to classify a high percentage of melanomas (85–93 %). The lower specificity values (45–48 %) obtained by the less experienced observers could be explained by the fact that most of the non-melanomas used to determine specificity were clinically atypical (leading to the decision to perform a biopsy); thus, they require more experience to perform correct assessments. However, use of the model would have avoided the excision of almost half of those lesions. For a cutaneous melanoma to be diagnosed, identification of at least 1 major and 1 minor ELM criterion (or 3 minor criteria) was required. This confirmed the previously reported rule that a single criterion usually does not suffice to make a diagnosis [6].

More recently, Haenssle et al. have assessed the sensitivity, specificity, and diagnostic accuracy of the 7-Point Checklist in the setting of a prospective long-term study. They have screened patients at increased melanoma risk at regular intervals by naked-eye examination, the dermoscopic 7-point checklist, and digital dermoscopy follow-up (10-year study interval). They have detected 127 melanomas including 50 melanomas in situ. The mean Breslow thickness of invasive melanomas has been 0.57 mm. A total of 79 melanomas were detected using the 7-point checklist melanoma threshold of 3 or more points (62 % sensitivity, compared with 78–95 % in retrospective settings). In all, 48 melanomas have scored fewer than 3 points and have been excised because of complementary information (e.g., lesional history, dynamic changes detected by digital dermoscopy). The specificity of the 7-point checklist has been 97 % (compared with 65–87 % in retrospective settings). Regression patterns, atypical vascular patterns, and radial streaming have been associated with the highest relative risk for melanoma (odds ratio 3.26, 95 % confidence interval 2.05–5.16; odds ratio 3.04, 95 % confidence interval 1.70–5.46; odds ratio 2.91, 95 % confidence interval 1.64–5.15; $P < 0.0003$, respectively). Melanomas thicker than 0.5 mm have exhibited significantly more regression patterns and atypical vascular patterns ($P < 0.02$). The malignant versus benign ratio for all excised lesions has been 1:8.6 (127 melanomas, 1092 non-melanomas). As a consequence, the 7-point checklist has appeared less sensitive but highly specific in this prospective clinical setting. Complementary information has clearly increased sensitivity. Thus, the authors have suggested that regression patterns or radial streaming in nevi of patients

at high risk should raise a higher melanoma suspicion than might be concluded from retrospective studies [7].

Further studies have proposed the use of simplified diagnostic algorithms, such as 3-point. Firstly Soyer et al. have evaluated the diagnostic performance of non-experts using a new 3-point checklist based on a simplified dermoscopic pattern analysis. Clinical and dermoscopic images of 231 clinically equivocal and histopathologically proven pigmented skin lesions were examined by 6 non-experts and 1 expert in dermoscopy. For each lesion the non-experts assessed 3 dermoscopic criteria (asymmetry, atypical network and blue-whitish veil) constituting the 3-point method. In addition, all examiners made an overall diagnosis by using standard pattern analysis of dermoscopy. Asymmetry, atypical network and blue-white structures were shown to be reproducible dermoscopic criteria, with a kappa value ranging from 0.52 to 0.55. When making the overall diagnosis, the expert had 89.6% sensitivity for malignant lesions (tested on 68 melanomas and 9 pigmented basal cell carcinomas), compared to 69.7% sensitivity achieved by the non-experts. Remarkably, the sensitivity of the non-experts using the 3-point checklist reached 96.3%. The specificity of the expert using overall diagnosis was 94.2% compared to 82.8 and 32.8% achieved by the non-experts using overall diagnosis and 3-point checklist, respectively. These data showed that the 3-point checklist can be considered a valid and reproducible dermoscopic algorithm with high sensitivity for the diagnosis of melanoma in the hands of non-experts. Thus, the authors suggested that the 3-point checklist might be applied as a screening procedure for the early detection of melanoma [8]. Afterwards, Zalaudek et al. reevaluated these preliminary results in a large number of observers independently from their expertise in dermoscopy. The three-point checklist showed good interobserver reproducibility (kappa value: 0.53). Sensitivity for skin cancer (melanoma and basal cell carcinoma) was 91.0% and this value remained basically uninfluenced by the observers' professional profile. These results confirmed that the three-point checklist was a feasible, simple, accurate and reproducible skin cancer screening tool [9].

In 2010, Gerely et al. compared the sensitivity, specificity, and diagnostic accuracy of the seven-point and three-point checklist methods in the diagnosis of clinically atypical pigmented skin lesions and melanoma. The sensitivity, specificity, and positive and negative predictive values of the seven-point checklist method were 87.50, 16.17, 51.22, and 57.14%, respectively. The sensitivity, specificity, and positive and negative predictive values of the three-point checklist method were 89.58, 31.25, 56.58, and 75%, respectively. Thus, this study highlighted that the three-point checklist was observed to be a superior screening test. The seven-point checklist provided a more detailed analysis, especially for thin melanomas. In comparison with the seven-point method, the three-point method may be useful for less experienced observers when they need to obtain greater diagnostic accuracy [10].

However, it has been demonstrated that the 7-point checklist and the other above-cited diagnostic algorithms have actually a lower diagnostic accuracy if they are performed by inexperienced dermatologists [11]. Dermoscopy requires high training to optimize diagnosis of pigmented skin lesions. Indeed, Binder et al. demonstrated

that ELM pattern analysis increases the quality of diagnostic performance of ELM experts but decreases the performance of clinicians not specially trained in ELM [12].

To avoid these problems, to enhance the reproducibility of clinical diagnosis and to help clinicians with poor dermoscopic experience, computer-assisted analysis of dermoscopic images has been investigated. Computer-aided diagnosis can help clinicians in the diagnosis of skin lesions. The aim of these systems is to increase the specificity and the sensitivity in melanoma recognition and reduce unnecessary biopsies. Most of these automated systems are based on the afore-mentioned melanoma diagnosis methods. In general, image processing techniques are used to locate the lesions, extract image parameters describing the dermatological features of the lesions, and, based on these parameters, perform the diagnosis. Their potential benefits are very promising, but there are considerable difficulties involved in their development and their use in clinical practice. It has been widely stated that their accuracy can achieve the same range as dermoscopic diagnosis performed by experts or even that they can obtain better accuracy [13]. Computer-aided diagnosis, based on mathematical analysis of pigmented skin lesions, can be a tool to transform a qualitative evaluation into a quantitative one and to increase sensitivity of dermatologists with low dermoscopic experience. Suggest clinicians step by step the dermoscopic criteria arising from observation can mean to implement the dermoscopy use without reduce its diagnostic accuracy.

Automatic Diagnostic Systems

As previously introduced, there has been much research 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 [14].

Therefore, different groups have been developing diagnostic systems of recorded images (slides or digital cameras) techniques to assist clinicians in differentiating early melanoma from benign skin lesions [15].

For example, in [16] an automated melanoma recognition system is proposed taking into account 21 parameters extracted from images.

Schmid [17] proposed a color based segmentation scheme without extracting features, whereas a new procedure based on the Catmull-Rom spline method and the computation of the gray-level gradient of points extracted by interpolation of normal direction on spline points was employed in [18].

A computer algorithm for the diagnosis of melanocytic lesions based on the evaluation of 64 different analytical parameters is described in [19], 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 [14].

Very interesting summaries of the main researches about the digital dermoscopy (in terms of acquisition, calibration, image datasets and processing methods) are reported in [20, 21]. As a results of these surveys, high accuracy may be achieved by computer aided diagnostic systems employing statistics obtained from low-level features and parameters. Nevertheless, it is not likely that the digital system will completely substitute the expert in dermoscopy.

The Proposed Framework

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

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 Checklist.

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 [22].

Further studies [23, 24] have led to the introduction of a software framework [25] for the automatic detection of dermoscopic criteria. Following the example of the Computer Aided System architecture proposed for digital ELM images by Schmid [26] and the methodological approach to the classification suggested in [27], the software framework includes all the processing algorithms derived from the clinical knowledge gained by expert dermatologists (well-trained in the 7-Point Checklist application).

In [28] a statistical approach is introduced for the automatic detection of a minor criterion (Irregular dots/globules).

The present chapter reviews the main image processing techniques adopted to provide and 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. Finally, the experimental results are extended to a large set of pigmented lesion and a comparison among the different techniques is also carried out.

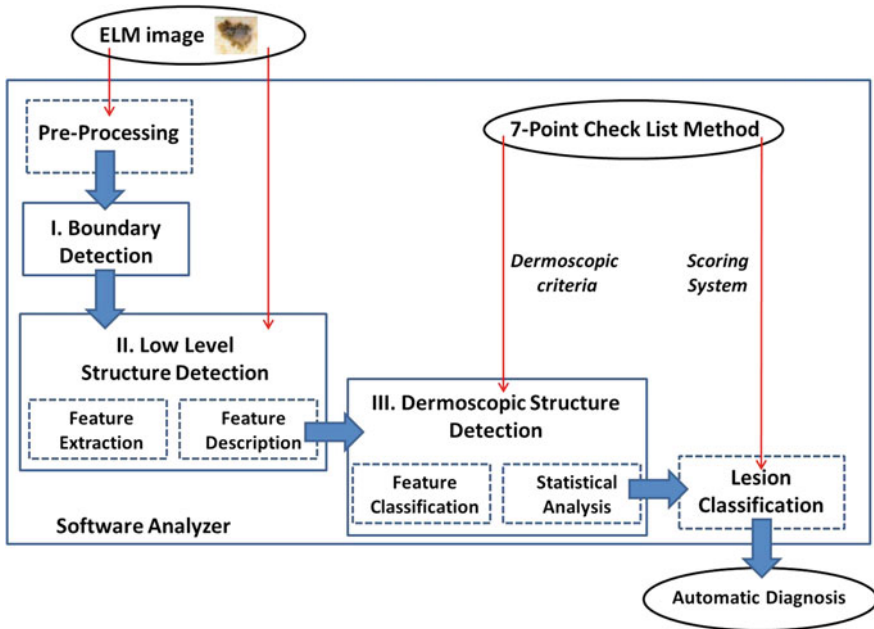


Fig. 2 Scheme of the software framework for the automatic diagnosis of ELM images

Methods

According to the scheme reported in Fig. 2, the software procedure developed for the automatic analysis and diagnosis of dermoscopic images is organized into three main detection stages.

After a preliminary processing stage designed to remove hair and/or artifacts, the *Boundary Detection* (I) allows the pigmented lesion to be extracted from the surrounding healthy skin. Then, the *Low-Level Structure Detection* (II) aims to identify and measure the main morphological and chromatic *features* throughout the lesion. Finally, at *Dermoscopic Structure Detection* (III), the feature classification and analysis are performed in order to detect each ELM criterion (high-level structure) provided by the 7-Point Checklist.

In the following subsections, for each stage of the automated procedure the remarkable literature is reviewed as well as the authors' novel approach is detailed in terms of advanced statistical techniques.

Pre-processing

Lesion segmentation in the presence of hair is usually doomed to failure. Because shaving the lesion area before the acquisition often interferes with the clinical practice, a computer-aided system for the analysis of dermoscopic images should always include an automated hair removal algorithm.

A well known hair removal algorithm was proposed in [29]: it identifies the image segments that approximate the structure of the hair, and then the regions that contain these segments are interpolated using the information of the surrounding pixels.

A similar approach is proposed by Schmid in [30], concerned with uniform color spaces, such as $L^*u^*v^*$ (the main advantage is that color difference can be measured and used for comparisons between pixels or distance measures in the spectral domain). According to the latter approach the morphological closing operator with a spherical structuring element is applied to the luminance component L^* , then the threshold operation is carried on the difference with the original image. More sophisticated techniques [31, 32] have also been introduced using image inpainting. Nevertheless, similar results are achievable. Thus, the algorithm disclosed in [30] has been preferred as preliminary stage in the proposed framework.

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. Consequently, different approaches [33] have been developed for automatic detection of lesion border in both clinical and dermoscopy images.

Many studies have introduced techniques based on color clustering [17, 34–36]. For example, in [17] the first two principal components of the CIE $L^*u^*v^*$ color space are adopted to determine a 2D histogram. Then, initial cluster centers are calculated from the peaks using a perceptron classifier and, finally, the lesion image is segmented using a modified version of the fuzzy c-means (FCM) clustering algorithm. Other color clustering algorithms provide for median cut, k-means, FCM and mean-shift [37]. Further approaches investigating on digital lesion images include snakes methods based on gradient vector flow [38, 39], improved region-based active contour algorithms [40], morphological flooding [30] and optimized JESG segmentation [41]. Finally, the Histogram Thresholding represents a widely-adopted strategy, upon which the latest investigations have been focused by introducing color channel optimization, hybrid (i.e. combined global and local) thresholding [42], and/or fusion within Markov Random Field framework [43].

A very interesting comparison of the main proposed approaches is reported in [44], where a new algorithm based on Statistical Region Merging (SRM) is also

introduced. As results of this survey, two approached emerged as the most effective methods: SRM and DTEA (Dermatologist-like Tumor Extraction Algorithm, [45]).

The Statistical Region Merging is a recent technique [46] 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 determined. This method is typically adopted for its simplicity, computational efficiency, and excellent performance without the use of quantization or color space transformations. Specifically, 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 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 detect the inner regions constituting the pigmented lesion to be contoured. According to the method suggested in [44], the background skin color is estimated as mean R , G and B colors of the pixels belonging to four patches (20×20 sized) from the corners of the image. Post-processing provides for the deletion of light-colored and bounding regions (including the regions whose mean color has an Euclidean distance less than 60 to the background skin color, the regions that touch the image frame and those with rectangular borders). The initial border detection result is obtained by removing the isolated regions and then merging the remaining regions. Finally a morphological dilation with a circular structuring element is applied to obtain the automatic border.

The DTEA algorithm is based on thresholding followed by iterative region growing. Following the same approach, the authors suggested in [25] a novel lesion border detection. The developed algorithm, referred to as Adaptive Thresholding, 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 Fig. 4a), two modes (classes) are typically evident in the pixel intensity histogram (as depicted in Fig. 4b) corresponding respectively to the pigmented

lesion (the image foreground) and the surrounding skin (the image background). Then, the algorithm introduced by Otsu [47] is adopted to select the optimum threshold S^* for each histogram, thus allowing the image background and foreground to be detected. The adaptive algorithm aims to minimize the intra-class variance σ_W :

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

defined as a weighted sum of variances σ_i of the two intensity classes C_i resulting from the S threshold:

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

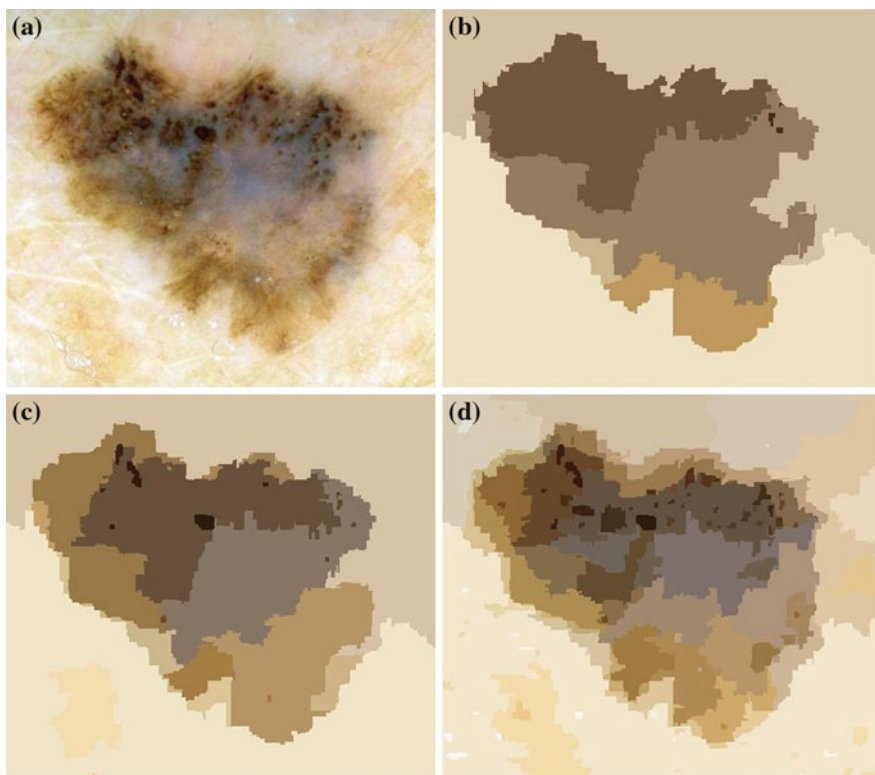


Fig. 3 Segmentation using statistical region merging: **a** ELM image; **b** results for $Q = 32$; **c** results for $Q = 64$; **d** results for $Q = 256$

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 σ_B :

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

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

The adoption of the Otsu's method to RGB color image leads to three histograms and potentially different thresholds values. Since the proposed approach has been

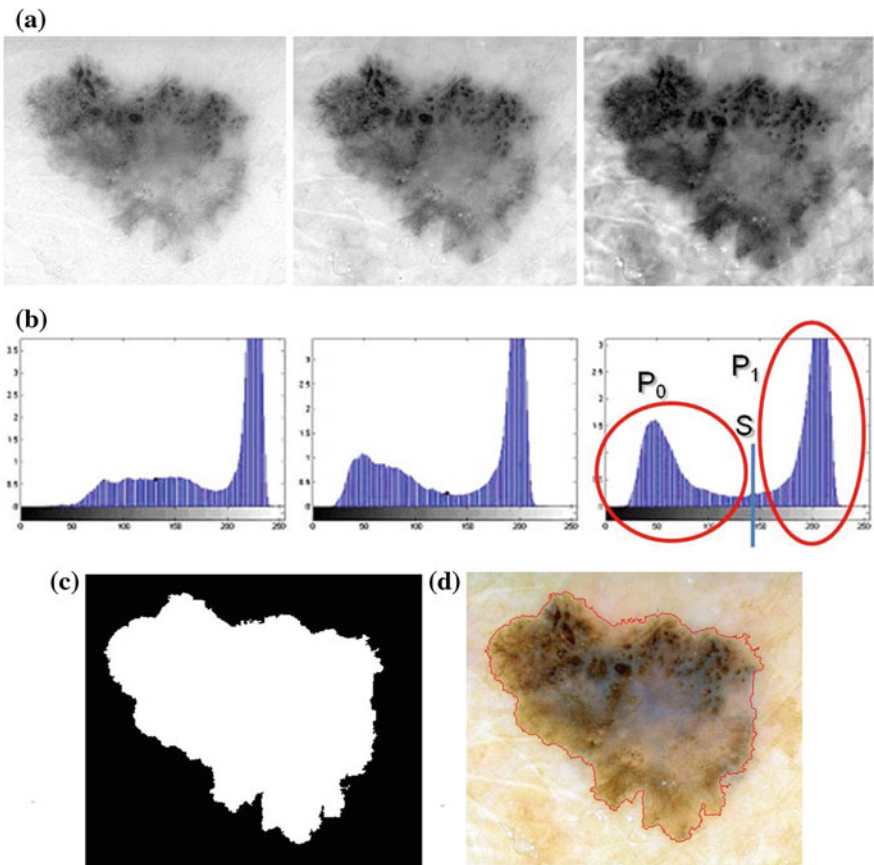


Fig. 4 Example of boundary detection: **a** image conversion (*Red, Green, Blue planes*); **b** intensity histogram; **c** binary mask; **d** lesion contour

experimentally revealed to be more sensitive to surrounding skin (the image background), the largest binary mask (the image foreground) is considered for next processing. An example of result is shown in Fig. 4c.

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 Neighbor Contour Tracing* proposed in [48], the tracking algorithm collects and sorts the contour lines (single pixel width) of the binary mask into an ordered list (the adopted algorithm also reveals to be computationally efficient by deleting the stopping criterion concerned with the start pixel). At this point, the border is superimposed on the color ELM image and displayed for visual inspection to the diagnostician (Fig. 4d).

Low Level Structure Detection

The dermoscopic criteria as defined by 7-Point Checklist Method are characterized both by chromatic and morphological low level structures (features). Thus, once the lesion is localized, *feature extraction* is performed by adopting suitable statistical techniques, which may be grouped into the following macro-categories:

- color segmentation
- texture analysis

Color Segmentation

Starting from the source image and the binary mask, the color segmentation stage is carried out with the aim of splitting the internal area into multiple chromatically homogenous regions (the lesion map).

To this aim, the SRM algorithm previously introduced may be adopted, by regulating the coarseness through a suitable choice of the Q parameter.

An alternative approach, proposed in [49] and investigated by the authors for dermoscopic images in [25], is represented by the Multi-Thresholding of the color image. In particular the following steps are proposed: (i) Principal Component Analysis (PCA); (ii) 2D histogram construction; (iii) peaks picking algorithm; (iv) histogram partitioning; (v) lesion partitioning.

(i) The Principal Component Analysis (also known as the discrete Karhunen-Loeve Transform or Hotelling Transform [50]) is a technique for reducing the dataset dimensionality while retaining those characteristics that contribute most to dataset variance. As for the application of the Principal Component Analysis to the ELM image, the RGB components of the pixels corresponding to lesion area (selected through the binary mask obtained as final result of image segmentation) constitute the starting dataset (belonging to a state space with dimension $N = 3$). A new 3D representation of the lesion pixels can be obtained from the Hotelling Transform equation. An example is reported in Fig. 5a–c: the decreasing variability in each

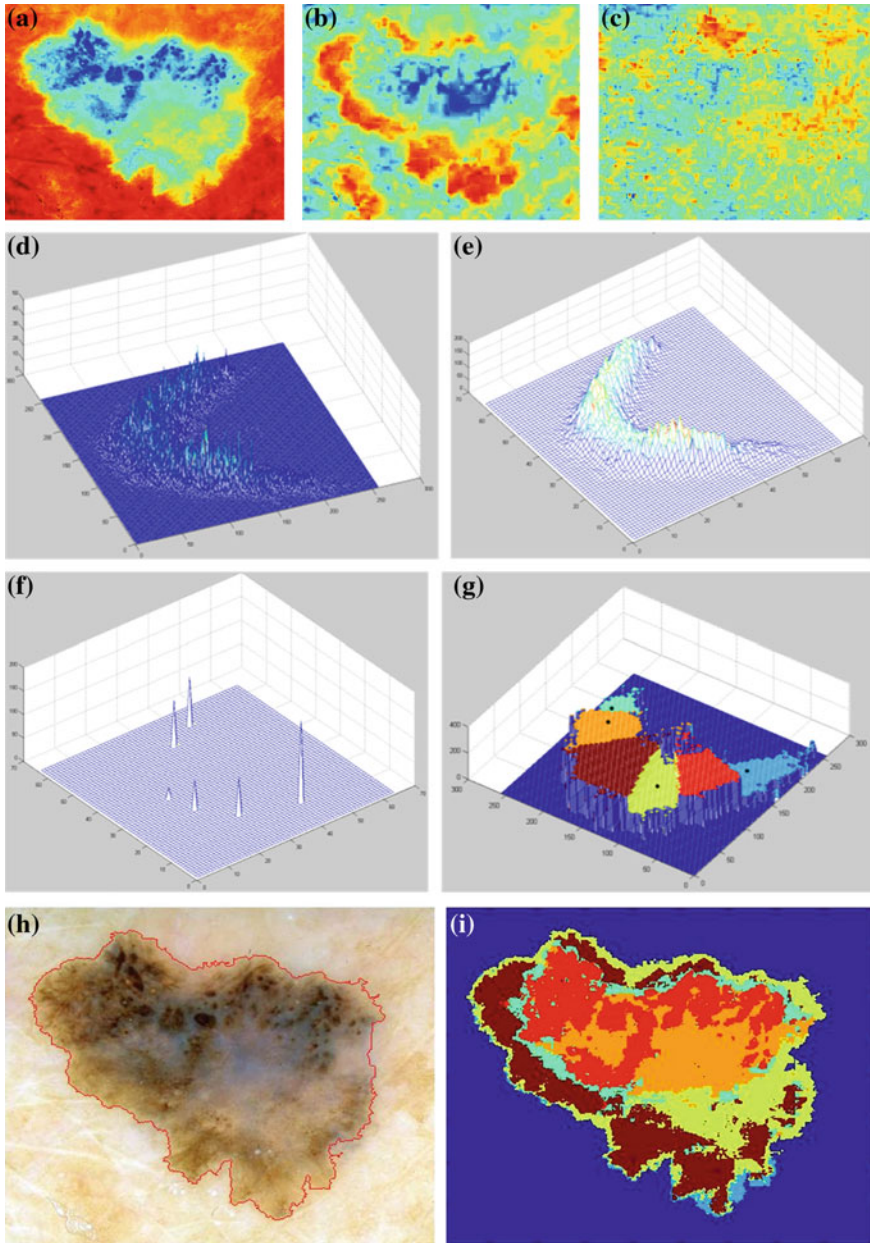


Fig. 5 Color segmentation based on Multi-thresholding: **a** 1st principal component; **b** 2nd principal component; **c** 3rd principal component; **d** joint histogram of the first 2 principal components; **e** down-sampled 2-D histogram; **f** result of peak-picking method; **g** partitioned 2-D histogram; **h** ELM image and lesion contour; **i** lesion map

individual band as long as the order of the principal component increases can be noted easily (for more evidence, PCA is related to all image pixels taking into account also the surrounding skin).

(ii) Since the low order components preserve sufficient information in order to obtain reliable information (Fig. 5a, b) whereas the third component contains most of the image noise (Fig. 5c), a joint histogram is created from the first 2 principle components (referring to which the multithresholding has to be carried out). An example of 2-D histogram is depicted in Fig. 5d. Because the estimated histograms are, in general, noisy due to the scarcity of data, it is advantageous to smooth and down-sample the histograms to eliminate noise effects. In particular the original histogram is reduced from size 256×256 to size 64×64 (see Fig. 5e).

(iii) The multithresholding is carried out by finding peaks in the 2-D histogram with significant mass around them. It is expected that these peaks will correspond to the cluster centroids in 2-D space and consequently will be well-representative of corresponding color regions (or segments) in the starting image. The knowledge of the number of segments is implicit in the peak search, and so is the maximum number K of peaks which have to be determined in the 2-D histogram. In our application the algorithm of Koonty [51] has been considered as peaks-picking method. As an example, in Fig. 5f the result of the peaks-picking algorithm is depicted with reference to the 2-D histogram shown in Fig. 5e when K equal to 10 is selected as maximum number of different color regions.

(iv) Once the peaks are identified, each corresponding hopefully to a segment, the other (non-peaks) histogram bins are attributed to the nearest dominant peak, constituting effectively their domains. Thus, a 2-D histogram is partitioned using its peak bins and an assignment rule (gravity force) which takes into account the strength (height) of the peak and the distance from the pick to the histogram bin under consideration. Figure 5g shows a partitioned 2-D histogram (after the partitioned 64×64 2-D histogram has been sampled back to its original size by simple replication of the bin labels by 4×4 fold): each color represents a histogram region.

Once the partitioned 2-D is computed, each pixel in the starting image (see Fig. 5h) can be directly labeled by taking into account the corresponding values for the two principal components. In particular assigning to each histogram region (or segment) an arbitrary intensity value, a gray-level image (or alternatively a false-color image) can be obtained where different regions are easily identified (see Fig. 5i).

Texture Extraction

As to the search for morphological (low-level) structures within the lesion, several approaches have been proposed in literature, including both structural and spectral methods [52]).

The structural techniques, which are intended to search for primitive structures such as points, lines and circles, have been extensively adopted for automatically detecting texture and/or local networks in dermoscopic images. For example, one of the most recent studies about the pigment network [53] introduces a feature extraction

based on the Laplacian of Gaussian (LOG) filtering. More in detail, the result of the edge detection step is a binary image which is subsequently converted into a graph to find the lesions meshes. Similarly, the detection of cyclic structures representing the pigment network is performed in [54] on the basis of the matching filtering principle and the adoption of suitable directional filters (namely 2-D Gabor filters).

The spectral technique is based on the Fourier analysis of the grey-level image. About computerized dermoscopic analysis, the approach is useful to determine the spatial period of the texture, thus allowing the identification of the regions where typically a network exists.

Thus, in order to disclose the pigment network (the dermoscopic criterion mainly correlated to low level morphological structures), a feature extraction combining structural and spectral methods has been introduced by the authors in [23]. With reference to the diagram in Fig. 6, 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 proposed in order to identify the main local discontinuities within the image: the monochromatic image is first compared with its version obtained by a suitable median filter, then a close-opening operation is performed, which deletes eventual isolated points.

(ii) A sequence of Fast Fourier Transform (FFT), high-pass filtering, Inverse Fast Fourier Transform (IFFT) and suitable thresholding has been adopted. As goal of the spectral 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.

Feature Description

Following the automatic extraction of low level structures, the feature analysis is proposed in order to determine measurement information in terms of both chromatic and morphological descriptors.

Thus, the lesion map resulting from segmentation (performed through the SRM technique or the Multi-Thresholding algorithm) is investigated in order to identify the most significant descriptors of the local regions (*features*), such as components in the main color spaces in terms of mean value and standard deviation as well as relative difference among neighbors. An example is reported in Fig. 7, where the corresponding feature extraction resulted from a color fine segmentation ($Q = 256$).

More in detail, for each local region the pixel components in *RGB*, *HSI* and *Luv* color space are considered and the corresponding mean value and standard deviation

are computed. A percentage (30%) dilation is also considered in order to compute further chromatic descriptors as the relative difference (mean value) with respect to neighbor regions. Moreover, the following morphological descriptors are computed for further analysis:

- relative dimension $A\%$, defined as the number of the region pixels 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).

About the feature extraction results concerning with the pigment network, some feature descriptors are computed from the statistical distribution of observed intensities in the *network image* at specified positions relative to each other. According to the number of intensity points (pixels) in each combination, statistics may be classified into first-order, second-order and higher-order statistics.

The Gray Level Co-occurrence Matrix (GLCM) method [55] is an extensively adopted way of extracting second order statistical texture features.

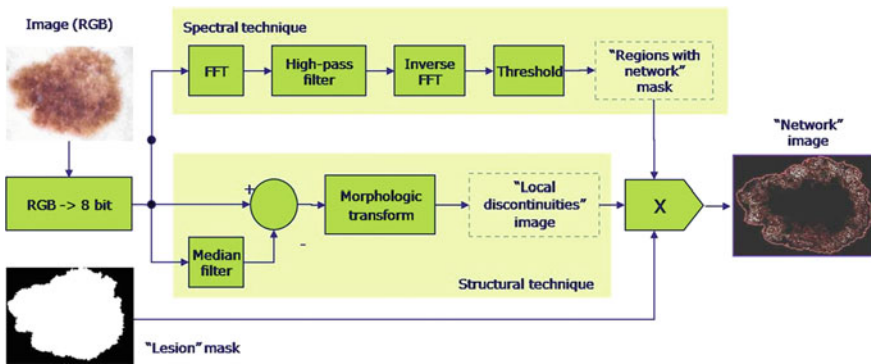


Fig. 6 Proposed scheme for texture analysis of pigmented lesions

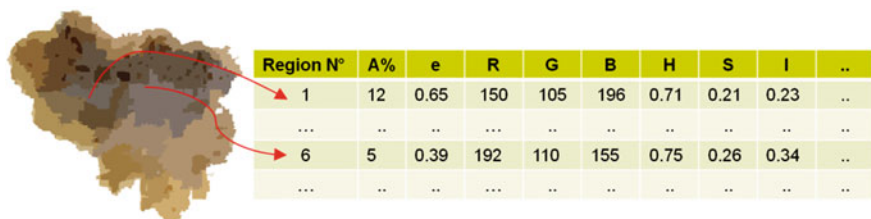


Fig. 7 Example of fine color segmentation (SRM, $Q = 256$) feature description: segment area $A\%$ (percentage with reference to the lesion area); segment Eccentricity e , mean value of Red (R), Green (G), Blue (B), Hue (H), Saturation (S), Intensity (I) components (averaged on pixels constituting the segment)

Generally speaking, a GLCM is a matrix where the number of rows and columns is equal to the number of gray levels, G , in the image; each matrix element $P(i, j|d, \theta)$ contains the second order statistical probability values for changes between gray levels i and j at a particular displacement distance d and at a particular angle θ . A very interesting example of GLCM application to the computerized analysis of digital dermoscopic images is reported in [56], where 176 texture descriptors (on a total of 428 objective descriptors including color and asymmetry properties) were derived from 11 different-sized co-occurrence matrices with distance value d ranging from $1/2$ to $1/64$ of the length L of the major axis of the lesion. Texture descriptors mainly contributed to PCA-based classifiers able to effectively discriminate between melanomas and nevi as well as ridges and furrows.

Following this example, in order to avoid dependency of direction, one may calculate an average (isotropic) matrix out of four matrices ($\theta = 0^\circ, 45^\circ, 90^\circ, 135^\circ$), whereas the parameter d is suitable chosen according to the image resolution (as $d = L/32$). Finally, from the isotropic GLCM, a set of texture descriptors is computed, which includes entropy, inverse difference moment and correlation.

Dermoscopic Structure Detection

At this stage, each (high-level) dermoscopic structure provided by 7-Point Checklist is automatically disclosed within the lesion through suitable classification algorithm and/or statistical analysis, which take into account the features descriptors previously introduced.

Feature Classification

Most literature concerning with computerized dermoscopy has been focused on supervised learning as typical approach to classify discriminative features inspired from both ABCD rule and identification of specific patterns within the lesion. Supervised learning is, in fact, a general technique of estimating model parameters given a set of training examples. Thus, dermoscopic features are fed into a classifier and supervised learning is typically used to diagnose unseen images.

Although a general model using supervised learning and Maximum A Posteriori Probability (MAP) estimation has been recently proposed in [57] to perform common tasks in automated skin lesion diagnosis (also including border detection, artifact detection) with interesting and promising results, it is usually the case that a supervised learning is only performed in the final stage of feature classification.

Following this trend, the classification of the chromatic features (the lesion map) is straightforwardly viewed as a problem of data mining from feature descriptors. In this context a well-known class of solutions is represented by Decision Tree Classifiers, which belong to the Machine Learning techniques [58]. This type of classifiers has been firstly introduced in the computer-assisted analysis of ELM images by

Debeir et al. in [59], where Decision Tree were suitably learned and adopted both for skin-lesion segmentation and pigmented lesion classification (between five lesion patterns). Moreover, decision trees have been successfully adopted as pixel classification technique in [60] in order to automatically detect the blue-white veil areas in dermoscopic images.

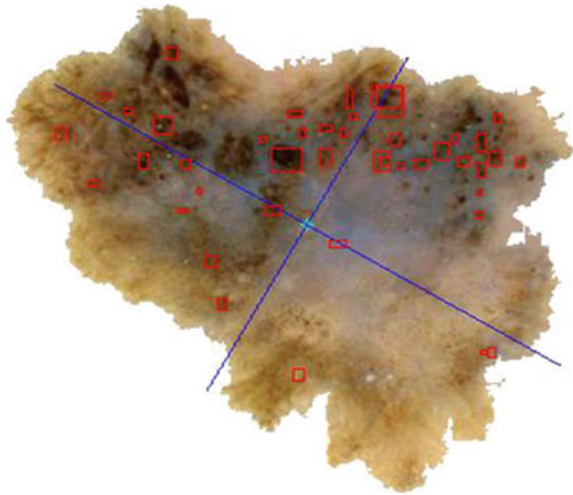
A Decision Tree Classifier is a predictive model, trained (or induced) by adopting a suitable dataset with respect to which classification results are already available. More in depth, given a collection of objects (each one described by a set of attributes) a Decision Tree is a graph, wherein each internal 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 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 [61] 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. For example, in [54, 62], the SimpleLogistic classifier is proposed to perform the automatic detection of pigment network and irregular streaks respectively. Generally speaking, logistic regression tries to fit a simple (linear) model to the data through a process which typically 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 explored by the authors for performing the classification tasks is a combination of a tree structure and logistic regression models resulting in a single tree according to the model proposed in [63].

Thus, the Logistic Model Tree (LMT) has been proposed for classifying the chromatic features (the lesion map resulting from the color segmentation) on the basis of the corresponding descriptors, as detailed further in the text.

Fig. 8 Result of feature extraction: detection of round items within the lesion area



Statistical Analysis

About the classification of morphological features, a statistical approach based on the Test Hypothesis is proposed in order to verify the irregular distribution of the dermoscopic structures of interest. For better explaining the underlying idea, the method description refers to the example of feature extraction reported in Fig. 8. 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 highlighted as red boxes are the chromatic and/or morphological features resulting from color segmentation and texture extraction (more details are given further in the text). They could correspond to texture elements of the lesion network and/or isolated dots and globules. Thus, the candidate features could 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 Binomial Test can be performed to estimate the casual distribution of N round objects, once the accepted risk α 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 structured are classified as irregular.

The approaches previously introduced have been adopted to perform the automatic detection of dermoscopic high-level structures (criteria) provided by the 7-Point Checklist method.

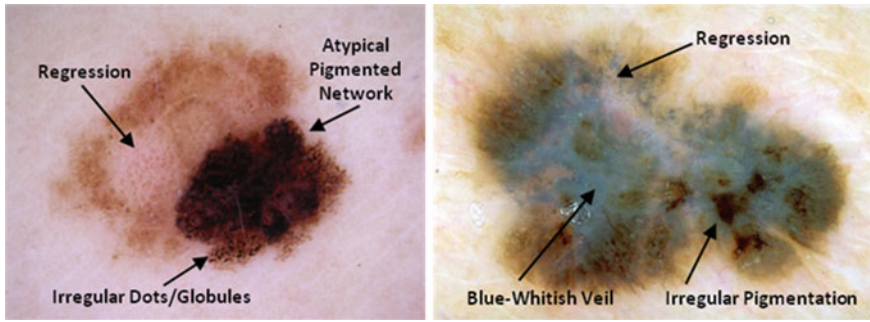


Fig. 9 Detection of dermoscopic criteria according to 7-point checklist

Examples of the diagnosis (by expert dermatologist) about two pigmented lesions are shown in Fig. 9, where the dermoscopic structures are highlighted.

More in detail, Fig. 9a shows a melanoma (total diagnostic score equal to 4) where a major criterion (Atypical Pigment Network) and two minor criteria (Regression and Irregular Dots/Globules) are detected. Similarly, within the melanoma displayed in Fig. 9b 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, defined as the ratio of correct detection of the high-level structure analyzed and total number of cases where the dermoscopic criterion is present;
- specificity, defined 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.

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.

As already mentioned, for each region the components of the corresponding pixels in the *RGB*, *HSI* (Hue, Saturation and Intensity) and *CIE Luv* color spaces have been considered to compute mean value and standard deviation as feature descriptors (vector 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 Fig. 8 with reference to the detection of the Blue-whitish Veil.

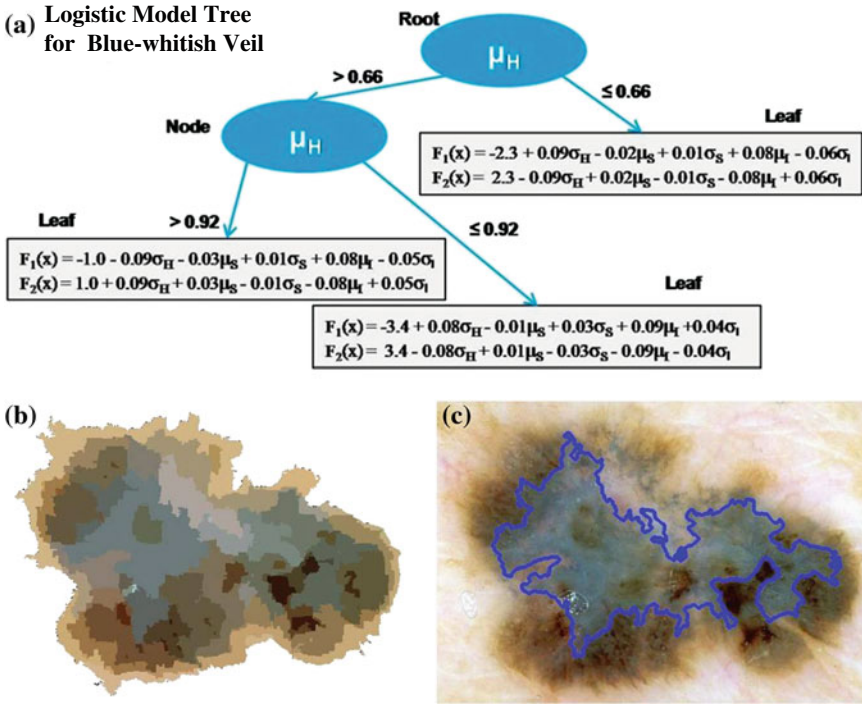


Fig. 10 Detection of the blue-whitish veil: **a** logistic model tree; **b** lesion map (feature extraction); **c** regions classified as blue-whitish veil

As you can see in the scheme (Fig. 10a), 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 $F_i(x)$ (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 chromatic features (color regions displayed in Fig. 10b) belong to an area characterized by the Blue Veil (the resulting detection of the criterion is depicted in Fig. 10c).

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

About the automatic detection of regression structures, two different logistic models (see Fig. 11a) have been computed according to the range wherein the mean value for the Saturation component of the region segment falls. Just five feature descriptors are truly significant to determine the probability that the color region belongs to an area characterized by Regression (an example of the resulting detection is also reported in Fig. 11b–d).

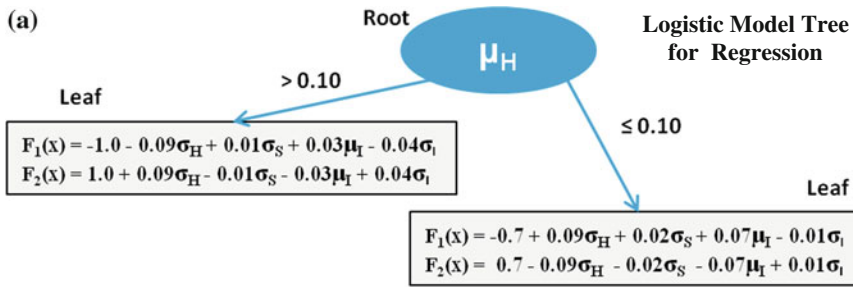


Fig. 11 a LMT for classification with respect to regression; b ELM image and contour; c lesion map; d detection results

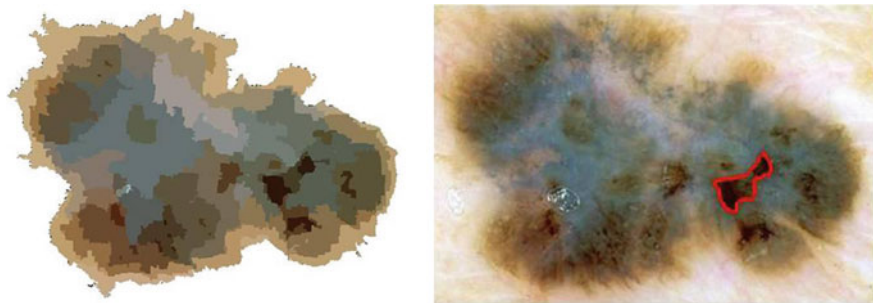


Fig. 12 LMT for automatic detection of irregular pigmentation

About the Irregular Pigmentation, a very simple LMT has been obtained: it computes the class probabilities taking into account the Intensity and L components (mean and standard deviation), and the area percentage measured for each color segment of the image (see Fig. 12).

$$F_1(x) = -0.5 - 0.09\mu_L + 0.07S\% + 0.04\mu_I - 0.02\sigma_I \quad F_2(x) = -F_1(x)$$

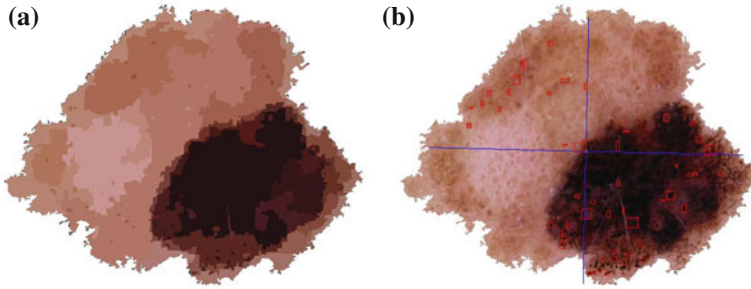


Fig. 13 Detection of irregular dots and globules: **a** color segmentation (SRM); **b** feature extraction: rounded items for hypothesis test

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 . As you can easily note in the example reported in Fig. 13a, the darkest segments may be deeply investigated to seek for the structures which represents Irregular Dots and Globules.

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 morphological feature descriptors previously introduced (percentage area $A\%$ and eccentricity e) are also compared with corresponding thresholds derived by experimental testing and tuned in order to extract rounded items inside the lesion.

Once the feature identification and analysis is completed with the detection of N round objects (see the items lightened with respect to the main symmetry axes of the lesion in Fig. 13b), according to the statistical approach previously introduced, the casual distribution is considered as Null Hypothesis of a Binomial Test. The following thresholds $k_{1,min}$, $k_{1,MAX}$ and k_2 can be jointly adopted for estimating the irregularity of Dots and Globules:

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

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

where α 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.

Atypical Pigment 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 pigment network within the lesion of interest is detected suitably combining the texture extraction and the color segmentation. Specifically, the areas constituting the network (white objects in Fig. 14a) are matched with the darkest regions of the lesion map computed through the Statistical Region Merging (see Fig. 14b). Then, the N objects resulting from the coupled analy-

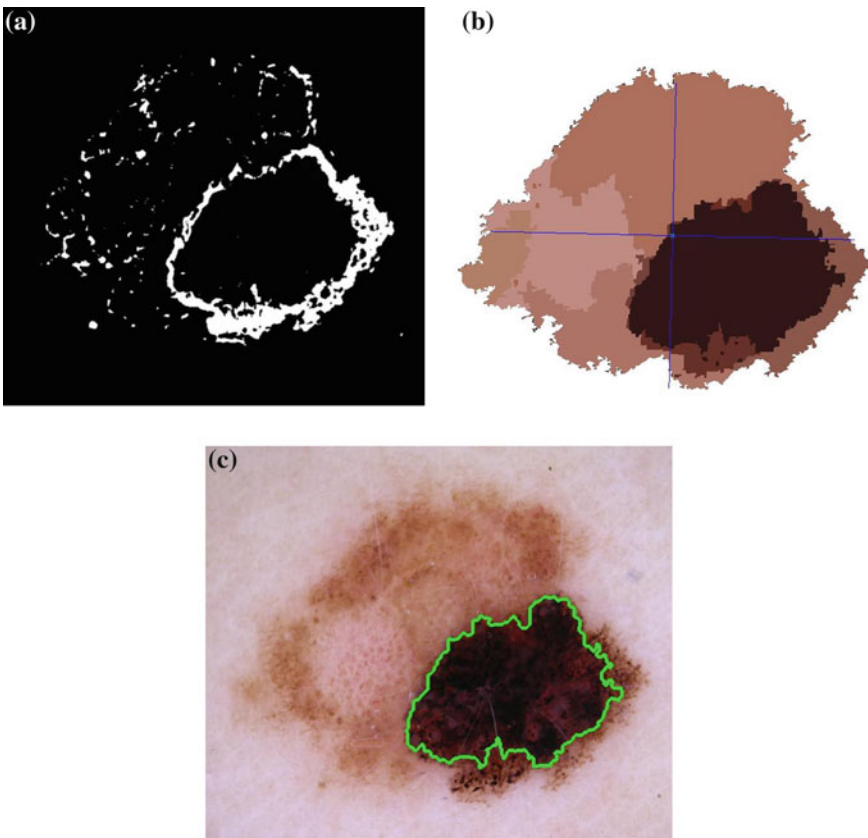


Fig. 14 Detection of atypical pigment network: **a** texture extraction; **b** color segmentation; **c** classification results

sis are classified as irregularly distributed (“atypical”) by performing the Hypothesis Test according to the Eqs. (4) and (5). In the example reported in Fig. 14c, the major criterion is detected in the darkest right-bottom area where the pigment network is mainly distributed.

Irregular Streaks

Although the presence of Irregular streaks is highly suggestive for malignancy of a lesion, the modeling, detection and analysis of streak lines and starburst pattern have rarely been used. A summary of the previous studies is reported in [62], where an original graph-based approach and very interesting feature set are also proposed. Nevertheless, the main hypothesis (the lesion modeling as an ellipse) leads to reduced detection accuracy when the algorithm is applied to generalized image set. A shape independent approach is proposed by the authors combining structural technique and color segmentation. The presence of asymmetrically arranged (linear or bulbous) extensions at the edge of the lesion can be detected by searching for the simultaneous occurrence of two different structures:

- (i) brown pigmentation localized in the same restricted region, and
- (ii) finger-like track of the contour of the lesion.

Both structures are detected by mean a local analysis of the lesion contour, this last is split into 10 equally length segments. A color segmentation of the region of interest is performed through the Statistical Region Merging in order to seek for the black/brown dermoscopic structures. Then a morphological irregularity index is computed and compared with a suitable threshold.

The irregularity index is defined as ratio of number of pixels constituting the lesion contour and the shortest path”, where the “shortest path” pixels are the points belonging to the line that connects the farther contour points in the region (see Fig. 15).

Atypical Vascular Pattern

About the automatic detection of Atypical Vascular Pattern, an approach combining color segmentation and structural analysis is proposed similarly to the methodology concerned with the Irregular Dots/Globules. Again the inner area of the lesion (see Fig. 16a) is considered and segmented through the Statistical Region Merging at fine level ($Q = 256$).

The resulting SRM image (see Fig. 16b) is firstly matched with the texture descriptors (entropy, inverse difference moment and correlation) based on the gray level co-occurrence matrix in order to exclude texture areas.

Then, a statistical analysis of the candidate SRM segments is performed by comparing the corresponding Hue range and eccentricity with suitable thresholds (experimentally tuned through a supervised learning approach) in order to detect linear or globular red structures irregularly distributed within the lesion. An example of linear vascular pattern is reported in Fig. 16c.

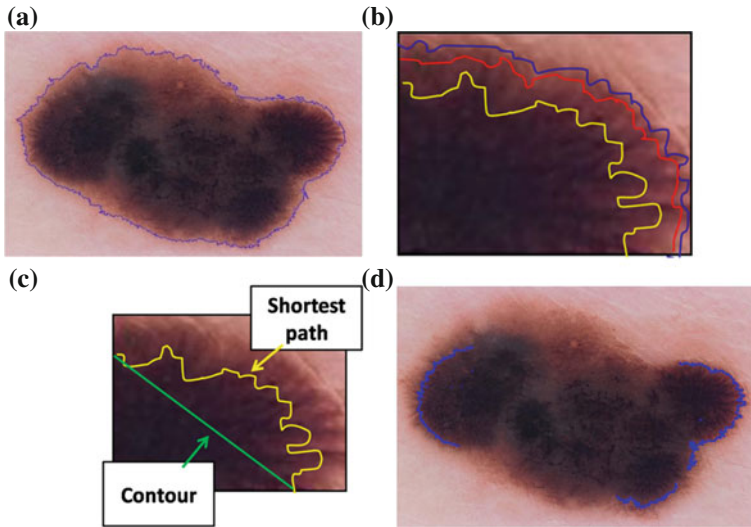


Fig. 15 Detection of irregular streaks: **a** ELM image and lesion contour; **b** the lesion area next to the border segment (*blue line*) is investigated (colour segmentation based on SRM) searching for darkest regions; **c** finger-like structures are detected through a quantitative comparison between the edge (*yellow line*) of the brown pigmentation and the corresponding straight line (*green line*); **d** results of automatic detection

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 [64] and observed by epiluminescence microscopy by two different dermatologists (M.S., G.F.) to evaluate the grade of accuracy in the management of 7 Point Checklist algorithm. Moreover, three dermatologists specifically trained in dermoscopy were asked to assess digital images of 100 melanocytic skin lesions selected among a digital collection of lesions screened between 2010 and 2012 at the *Department of Dermatology of the University of Naples Federico II*. In this department, the imaging is performed by a digital camera (Canon Power-Shot G9 with Heine Dermaphot Optics) that is combined with an epiluminescence microscope in order to produce digitized ELM images of skin lesions. The observers were first asked to use pattern analysis to score each lesion as naevus, melanoma or lesion to be excised, using the individual criteria listed in the seven-point checklist. For each image, the corresponding clinical and/or histological analyses were available. The images were extracted in order to obtain a quite homogeneous diagnosis distribution of the cases with respect to the criteria of interest.

As consequence, the overall database refers both to cutaneous melanomas and melanocytic nevi (also including Clark, Spitz, Reed nevi). About the image qual-

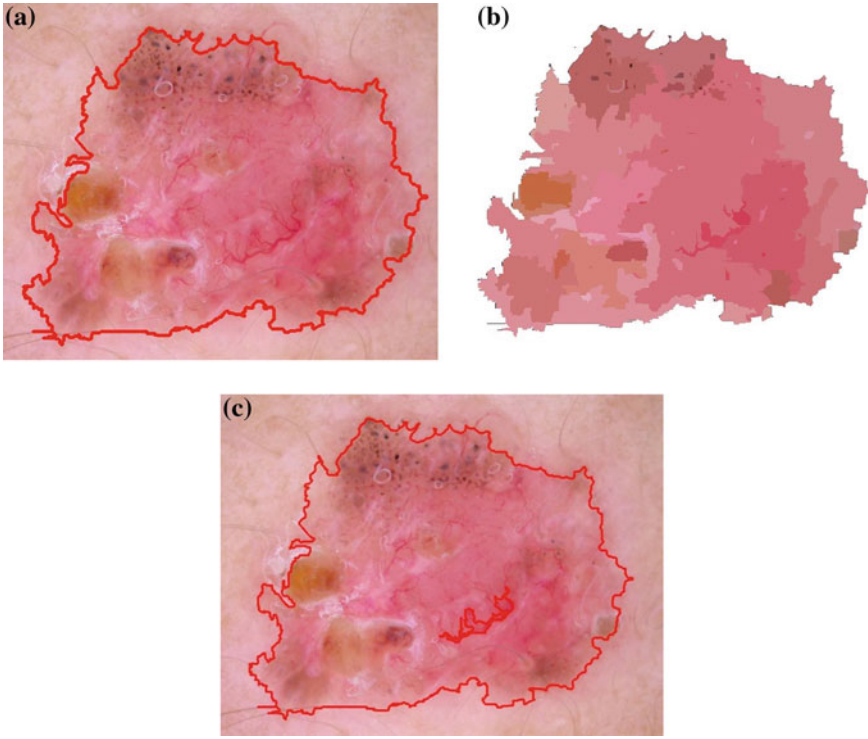


Fig. 16 **a** ELM image and lesion contour; **b** SRM segmentation of the lesion area; **c** automatic detection of atypical vascular pattern

ity, all the pictures are 24-bit RGB color images in JPEG format with dimensions ranging from 700×447 to 2272×1520 pixels. The lesions are imaged completely with healthy skin visible at margins. As previously mentioned, the image pre-processing strategy based on mathematical morphology [30] has been adopted for artifact removal.

Border Detection

The proposed technique based on the Adaptive Thresholding has been compared with the unsupervised approach based on the SRM algorithm, which was revealed the most effective method [44] 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.

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, 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 [48] is applied to determine the ground truth manual border.

Using the dermatologist-determined borders, the automatic borders resulting from the Adaptive Thresholding and SRM have been compared using the metric suggested in [65]. Here, the percentage border error is given by:

$$\text{Border Error} = \frac{(\text{Automatic_Binary})\text{XOR}(\text{Ref_Binary})}{\text{Area}(\text{Ref_Binary})} 100 \% \quad (6)$$

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 (standard deviation).

An example of automatic contour extraction for a melanoma is reported in Fig. 17a, where the resulting borders are compared with the manual border.

As you can see in the reported details (Fig. 17b, 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

Table 2 Dermoscopic criteria and scores according to the 7-Point Checklist method

Segmentation technique	Benignant lesions		Melanoma		All lesions	
	Mean (%)	Standard deviation (%)	Mean (%)	Standard deviation (%)	Mean (%)	Standard deviation (%)
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

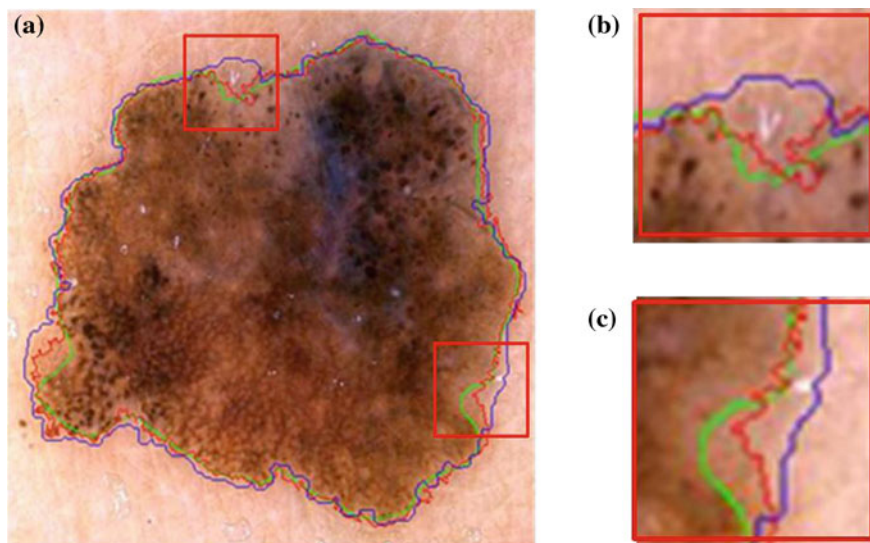


Fig. 17 Comparison between automated procedures for border detection: ground truth (*green line*), adaptive thresholding (*red line*), unsupervised approach (*blue line*)

fact, able to separate lesion and surrounding healthy skin also in critical local regions where the pixels components in RGB space are statistically close.

Automatic Detection of Dermoscopic Structures

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 have 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 based on Principal Component Analysis 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 features (local 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. Moreover, on the basis of experts' observations concerned with the Training Set, suitable thresholds (about the minimum detection area) have been derived to aggregate the *per-feature labeling* into *per-image classification accuracies*.

About the detection of Atypical Pigment Network Irregular Dots/Globules, suitable thresholds have been determined from the image properties of the Training Set through ROC curves [66] 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 α).

As an example, the verification of the proposed approach about Irregular Dots/Globules with respect to the previously 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 per-image 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 chromatic and morphological features, on which the classification of the dermoscopic criteria are based. Thus, the per-image 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.

A quite coarse segmentation ($Q = 64$) is able to disclose the areas characterized by Atypical Pigment 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).

Table 3 Irregular dots/globules: classifier performance (SRM, $Q = 256$)

Image set	Sensitivity (%)	Specificity (%)
Training set	80	82
Test set	90	85

Table 4 Automatic diagnosis of pigmented lesions: comparison among per-image classification results (test sets)

ELM criterion	Multi-threshold		SRM ($Q = 64$)		SRM ($Q = 128$)		SRM ($Q = 256$)	
	Sensitivity (%)	Specificity (%)	Sensitivity (%)	Specificity (%)	Sensitivity (%)	Specificity (%)	Sensitivity (%)	Specificity (%)
Atypical pigment network	80	82	82	85	85	78	87	74
Blue-whitish veil	90	93	66	92	73	90	84	85
Irregular pigmentation	84	77	85	81	93	66	97	63
Irregular dots/globules	-	-	-	-	87	88	90	85
Regression structures	72	75	80	83	83	70	85	67

Table 5 Classifier performance for atypical vascular pattern and irregular streaks

ELM criterion	Sensitivity (%)	Specificity (%)
Atypical vascular pattern	80	78
Irregular streaks	86	88

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.

Finally, Table 5 summarizes the classification performance in corresponding Test Sets of the proposed approach for the automatic detection of Atypical Vascular Pattern and Irregular Streaks.

Discussion and Future Work

On the basis of advanced techniques of image processing a Computer-Aided System has been achieved for the analysis of digital dermoscopic images according to the 7-Point Checklist method. The software routines are suitable to carry out: (i) the detection of the lesion contour; (ii) the extraction and measurement of the main chromatic and morphological features within the pigmented lesion; (iii) the classification and scoring of the dermoscopic structures.

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 a very promising software tool supporting the physician. Using pooled data obtained from expert observation and from computer detection, the diagnostic outcomes of pattern analysis were compared. The *sensitivity* of the system was calculated as the percentage of dermoscopic images scored as melanomas from the computer system and diagnosed as melanoma from expert dermatologists and confirmed by histology (the gold standard): it was resulted 97%. The *specificity* was calculated as the percentage of dermoscopic images scored both by the observers and by computer analysis as benignant melanocytic and naevi: it was resulted 87%.

Starting from the present framework, further research efforts will be firstly addressed to compare and integrate the very promising approaches and corresponding feature descriptors reported in the most recent literature [53, 54, 57, 62], in order to improve the classification accuracy of the dermoscopic structures.

Then, the correlations existing among the seven dermoscopic criteria will be deeply investigate and a confidence level will be computed for each intermediate classification (for example on the basis of multi-resolution segmentation, Fuzzy

fusion and/or Markov Random Field approach [43]). The corresponding information could be effectively adopted at Lesion Diagnosis stage for improving sensitivity and specificity of the software system as whole.

Finally, an intensive measurement campaign will be carried out aiming to a double goal. The control of the image acquisition and the availability of a large image database will allow to deeply investigate the influence of the color calibration on the proposed processing algorithms. Moreover, the diagnosis from the automatic system will be compared with 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.

This system will help dermatologists to deliver a fast and non-invasive diagnosis. Thus, it will help prevent skin cancer and treat it in its early stages, because patients will be more comfortable when tracking their lesions. However, because such instrumentation will never achieve 100 % diagnostic accuracy, and because the gold standard of histopathologic diagnosis suffers from significant interobserver disagreement, the diagnosis can't be only performed by computer, but a semiautomatic computer diagnosis can help the clinicians to achieve the best diagnostic accuracy. These technologies can be used in accordance with the patient history and clinical examination to enhance the ability to diagnose melanoma while avoiding unnecessary biopsies.

References

1. De Vries, W., Coebergh, J.W.: Cutaneous melanoma in Europe. *Eur. J. Cancer* **40**, 2355–2366 (2004)
2. Parkin, D.M., Whelan, S.L. et al.: *Cancer Incidence in Five Continents*, vol. I–VIII, no. 7. IARC Cancer Base, Lyon (2007)
3. Curado, M.P., Edwards, B., et al.: *Cancer Incidence in Five Continents IX*, no. 160. IARC Scientific Publications (IARC), Lyon (2007)
4. Leiter, U., Buttner, P.G., Eigentler, T.K., Garbe, C.: Prognostic factors of thin cutaneous melanoma: an analysis of the central malignant melanoma registry of the German Dermatological Society. *J. Clin. Oncol.* **22**(18), 3660–3667 (2004)
5. Mayer, J.: Systematic review of the diagnostic accuracy of dermatoscopy in detecting malignant melanoma. *Med. J. Aust.* **167**, 206–210 (1997)
6. Argenziano, G., Fabbrocini, G., Carli, P. et al.: Epiluminescence microscopy for the diagnosis of doubtful melanocytic skin lesions. Comparison of the ABCD rule of dermatoscopy and a new 7-point checklist based on pattern analysis. *Arch. Dermatol.* **134**(12), 1563–1570 (1998)
7. Haenssle, B., Korpas, Hansen-Hagg, C. et al.: Seven-point checklist for dermatoscopy: performance during 10 years of prospective surveillance of patients at increased melanoma risk. *J. Am. Acad. Dermatol.* **62**(5), 785–793 (2010)
8. Soyer, H.P., Argenziano, G., Zalaudek, I. et al.: Three-point checklist of dermoscopy. A new screening method for early detection of melanoma. *Dermatology* **208**(1), 27–31 (2004)
9. Zalaudek, I., Argenziano, G., Soyer, H.P.G. et al.: Three-point checklist of dermoscopy: an open internet study. *Br. J. Dermatol.* **154**(3), 431–437 (2006)

10. Gereli, M.C., Onsun, N., Atilganoglu, U. et al.: Comparison of two dermoscopic techniques in the diagnosis of clinically atypical pigmented skin lesions and melanoma: seven-point and three-point checklists. *Int. J. Dermatol.* **49**(1), 33–38 (2010)
11. Tromme, I., Sacré, L., Hammouch, F. et al.: Availability of digital dermoscopy in daily practice dramatically reduces the number of excised melanocytic lesions: results from an observational study. *Br. J. Dermatol.* **167**(4), 778–786 (2012)
12. Binder, M., Schwarz, M., Winkler, A., et al.: Epiluminescence microscopy a useful tool for the diagnosis of pigmented skin lesions for formally trained dermatologists. *Arch. Dermatol.* **131**, 286–291 (1995)
13. Blum, A., Luedtke, A.H., Ellwanger, U., Schwabe, R., Rassner, G., Garbe, C.: Digital image analysis for diagnosis of cutaneous melanoma. Development of a highly effective computer algorithm based on analysis of 837 melanocytic lesions. *Br. J. Dermatol.* **151**(5), 1029–1038 (2004)
14. Burrioni, M., Corona, R., Dell'Eva, G., et al.: Melanoma computer-aided diagnosis: reliability and feasibility study. *Clin. Cancer Res.* **10**, 1881–1886 (2004)
15. Rubegni, P., Cevenini, G., Burrioni, M., et al.: Automated diagnosis of pigmented skin lesions. *Int. J. Cancer* **101**, 576–580 (2002)
16. Ganster, H., Pinz, A., et al.: Automated melanoma recognition. *IEEE Trans. Med. Imag.* **20**, 233–239 (2001)
17. Schmid, P.: Segmentation of digitized dermoscopic images by two-dimensional color clustering. *IEEE Trans. Med. Imag.* **18**, 164–171 (1999)
18. Grana, C., Pellacani, G., Cucchiara, R., Seidenari, S.: A new algorithm for border description of polarized light surface microscopic images of pigmented skin lesions. *IEEE Trans. Med. Imag.* **22**, 959–964 (2003)
19. Blum, A., Luedtke, H., et al.: Digital image analysis for diagnosis of cutaneous melanoma. Development of a highly effective computer algorithm based on analysis of 837 melanocytic lesions. *Br. J. Dermatol.* **151**, 1029–1038 (2004)
20. Korotkov, K., Garcia, R.: Computerized analysis of pigmented skin lesions: a review. *Artif. Intell. Med.* **56**(2), 69–90 (2012)
21. Celebi, M. E., Stoecker, W.V., Moss, R. H.: Advances in skin cancer image analysis. *Comput. Med. Imaging Graph.* **35**(2), 83–84 (2011)
22. Di Leo, G., Fabbrocini, G. et al.: ELM image processing for melanocytic skin lesion based on 7-point checklist: a preliminary discussion. Proceedings of the 13th IMEKO TC-4 Symposium, vol. 2, pp. 474–479. Athens, Greece (2004)
23. Betta, G., Di Leo, G. et al.: Dermoscopic image-analysis system: estimation of atypical pigment network and atypical vascular pattern. Proceedings of Intern Work on Medical Measurement and Applications (MeMeA), pp. 63–67 (2006)
24. Di Leo, G., Liguori, C. 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*, pp. 190–194, Istanbul, Turkey, 14–16 July 2008
25. Di Leo, G., Fabbrocini, G. 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, Computer Society Press, 5–8 Jan 2010
26. Schmid, P., Guillodb, J.: Towards a computer-aided diagnosis system for pigmented skin lesions. *Comput. Med. Imag. Graphics* **27**, 65–78 (2003)
27. Celebi, M.E., Kingravi, H.A. et al.: A methodological approach to the classification of dermoscopy images. *Comput. Med. Imaging Graph.* **31**(6), 362–373 (2007)
28. De Vita, V., Di Leo, G. et al.: Statistical image processing for the detection of dermoscopic criteria. Proceedings of XVIII IMEKO TC-4 Symposium, Natal, Brazil, 27–30 Sept 2011
29. Lee, T., Ng, V., Gallagher, R. et al.: DullRazor: a software approach to hair removal from images. *Comput. Biol. Med.* **27**(6), 533–543 (1997)
30. Schmid, P.: Lesion detection in dermatoscopic images using anisotropic diffusion and morphological flooding. *IEEE*, pp. 449–453 (1999)

31. Abbas, Q., Garcia, I.F., Celebi, M.E., Ahmad, W.: A feature-preserving hair removal algorithm for dermoscopy images. *Skin Res. Technol.* **19**(1), e27–e36 (2013)
32. Abbas, Q., Celebi, M.E., Garcia, I.F.: Hair removal methods: a comparative study for dermoscopy images. *Biomed. Signal Process. Control* **6**(4), 395–404 (2011)
33. Celebi, M.E., Iyatomi, H., Schaefer, G., Stoecker, W.V.: Lesion border detection in dermoscopy images. *Comput. Med. Imag. Graph.* **33**(2), 148–153 (2009)
34. Donadey, T., Serruys, C. et al.: Boundary detection of black skin tumors using an adaptive radial-based approach. *SPIE Med. Imag.* **3379**, 810–816 (2000)
35. Cucchiara, R., Grana, C. et al.: Exploiting color and topological features for region segmentation with recursive fuzzy C-means. *Mach. Graphics Vis.* **11**, 169–182 (2002)
36. Melli, R., Grana, C. et al.: Comparison of color clustering algorithms for segmentation of dermatological images. *SPIE Med. Imag.* **6144**, 351–359 (2006)
37. Zhou, H., Schaefer, G., Sadka, A., Celebi, M.E.: Anisotropic mean shift based fuzzy C-means segmentation of dermoscopy images. *IEEE J. Sel. Top. Signal Process.* **3**(1), 26–34 (2009)
38. Erkol, B., Moss, R.H. et al.: Automatic lesion boundary detection in dermoscopy images using gradient vector flow snakes. *Skin Res. Technol.* **11**, 17–26 (2005)
39. Zhou, H., Schaefer, G., Celebi, M.E., Lin, F., Liu, T.: Gradient vector flow with mean shift for skin lesion segmentation. *Comput. Med. Imag. Graph.* **35**(2), 121–127 (2011)
40. Abbas, Q., Celebi, M.E., Garcia, I.F.: A novel perceptually-oriented approach for skin tumor segmentation. *Int. J. Innovative Comput. Inf. Control* **8**(3), 1837–1848 (2012)
41. Celebi, M.E., Aslandogan, Y.A. et al.: Unsupervised border detection in dermoscopy images. *Skin Res. Technol.* **13**(4), 454–462 (2007)
42. Garnavi, R., Aldeen, M., Celebi, M.E., Varigos, G., Finch, S.: Border detection in dermoscopy images using hybrid thresholding on optimized color channels. *Comput. Med. Imag. Graphics* **35**(2), 105–115 (2011)
43. Celebi, M.E., Wen, Q., Hwang, S., Iyatomi, H., Schaefer, G.: Lesion border detection in dermoscopy images using ensembles of thresholding methods. *Skin Res. Technol.* **19**(1), e252–e258 (2013)
44. Celebi, M.E., Kingravi, H.A. et al.: Border detection in dermoscopy images using statistical region merging. *Skin Res. Technol.* **13**(4), 347–353 (2008)
45. Iyatomi, H., Oka, H., Celebi, M.E., Hashimoto, M., Hagiwara, M., Tanaka, M., Ogawa, K.: An improved internet-based melanoma screening system with dermatologist-like tumor area extraction algorithm. *Comput. Med. Imag. Graphics* **32**(7), 566–579 (2008)
46. Nock, R., Nielsen, F.: Statistical region merging. *IEEE Trans. Pattern Anal. Mach. Intell.* **26**(11), 1452–1458 (2004)
47. Otsu, N.: A threshold selection method from gray-level histogram. *IEEE Trans. Syst. Man Cybern.* **SMC-9**(1), 62–66 (1979)
48. Pradhan, R., Kumar, S., et al.: Contour line tracing algorithm for digital topographic maps. *Int. J. Image Process.* **4**(2), 156–163 (2010)
49. Kurugollu, F., Sankur, B., Harmanci, A.E.: Color image segmentation using histogram multi-thresholding and fusion. *Imag. Vis. Comput.* **19**, 915–928 (2001)
50. Levy, A., Lindenbaum, M.: Sequential Karhuen-Loeve basis extraction and its application to images. *IEEE Trans. Image Process.* **9**, 1371–1374 (2000)
51. Koonty, W., Narendra, P.M., Fukunya, F.: A graph theoretic approach to non-parametric cluster analysis. *IEEE Trans. Comput.* **25**, 936–944 (1976)
52. Gonzalez, R.C., Woods, R.E.: *Digital Image Processing*. Prentice Hall, New Jersey (2002)
53. Sadeghi, M., Razmara, M., Wighton, P., Lee, T.K., Atkins, M.S.: A novel method for detection of pigment network in dermoscopic images using graphs. *Comput. Med. Imag. Graphics* **35**(2), 137–143 (2011)
54. Barata, C., Marques, J.S., Rozeira, J.: A system for the detection of pigment network in dermoscopy images using directional filters. *IEEE Trans. Biomed. Eng.* **59**(10), 2744–2754 (2012)
55. Haralick, R.M., Shanmugam, K., Dinstein, I.: Textural features for image classification. *IEEE Trans. Syst. Man Cybern.* **SMC-3**(6), 610–621 (1973)

56. Iyatomi, H., Oka, H., Celebi, M.E., Ogawa, K., Argenziano, G., Soyer, H., Koga, H., Saida, T., Ohara, K., Tanaka, M.: Computer-based classification of dermoscopy images of melanocytic lesions on acral volar skin. *J. Invest. Dermatol.* **128**(8), 2049–2054 (2008)
57. Wighton, P., Lee, T.K., Lui, H., McLean, D.I., Atkins, M.S.: Generalizing common tasks in automated skin lesion diagnosis. *IEEE Trans. Inf. Technol. Biomed.* **15**(4), 622–629 (2011)
58. Witten, H., Frank, E.: *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann, San Francisco (2005)
59. Debeir, O., Decaestecker, C.: Computer-assisted analysis of epiluminescence microscopy images of pigmented skin lesions. *Cytometry* **37**, 255–266 (1999)
60. Celebi, M.E., Iyatomi, H. et al.: Automatic detection of blue-white veil and related structures in dermoscopy images. *Comput. Med. Imag. Graphics* **32**(8), 670–677 (2008)
61. Quinlan, J.R.: *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers, San Mateo, CA (1993)
62. Sadeghi, M., Lee, T.K., McLean, D.I., Lui, H., Atkins, M.S.: Detection and analysis of irregular streaks in dermoscopic images of skin lesions. *IEEE Trans. Med. Imag.* **32**(5), 849–861 (2013). doi:[10.1109/TMI.2013.2239307](https://doi.org/10.1109/TMI.2013.2239307)
63. Landwehr, N., Hall, M., Frank, E.: Logistic model trees. 14th European Conference on, Machine Learning (2003)
64. Argenziano, G., Soyer, H.P., De Giorgi, V., et al.: *Interactive Atlas of Dermoscopy*. EDRA Medical Publishing & New Media, Milan, Italy (2002)
65. Hance, G.A. et al.: Unsupervised color image segmentation with application to skin tumor borders. *IEEE Engineering in Medicine and Biology Magazine* **15**(1), 104–111 (1996)
66. Florkowski, Christopher, M.: Sensitivity, specificity, receiver-operating characteristic (ROC) curves and likelihood ratios: communicating the performance of diagnostic tests. *Clin. Biochem. Rev.* **29**(Supplement (i)), S83–S87 (2008)