Automatic Detection of Malignant Melanoma using Macroscopic Images

Article *in* Journal of Medical Signals & Sensors · September 2014

DOI: 10.4103/2228-7477.144052

3 authors:

Maryam Ramezani
Isfahan University of Medical Sciences
1 PUBLICATION **36** CITATIONS
SEE PROFILE

Alireza Karimian
University of Isfahan
75 PUBLICATIONS **412** CITATIONS
SEE PROFILE

Payman Moallem
University of Isfahan
149 PUBLICATIONS **2,782** CITATIONS
SEE PROFILE

Original Article

www.jmss.mui.ac.ir

Automatic Detection of Malignant Melanoma using Macroscopic Images

Maryam Ramezani, Alireza Karimian, Payman Moallem[1]

Departments of Biomedical Engineering, [1]Electrical Engineering, Faculty of Engineering, University of Isfahan, Isfahan, Iran

Submission: 09-05-2014 Accepted: 17-09-2014

ABSTRACT

In order to distinguish between benign and malignant types of pigmented skin lesions, computerized procedures have been developed for images taken by different equipment that the most available one of them is conventional digital cameras. In this research, a new procedure to detect malignant melanoma from benign pigmented lesions using macroscopic images is presented. The images are taken by conventional digital cameras with spatial resolution higher than one megapixel and by considering no constraints and special conditions during imaging. In the proposed procedure, new methods to weaken the effect of nonuniform illumination, correction the effect of thick hairs and large glows on the lesion and also, a new threshold-based segmentation algorithm are presented. 187 features representing asymmetry, border irregularity, color variation, diameter and texture are extracted from the lesion area and after reducing the number of features using principal component analysis (PCA), lesions are determined as malignant or benign using support vector machine classifier. According to the dermatologist diagnosis, the proposed processing methods have the ability to detect lesions area with high accuracy. The evaluation measures of classification have indicated that 13 features extracted by PCA method leads to better results than all of the extracted features. These results are 82.2% of accuracy, 77% of sensitivity and 86.93% of specificity. The proposed method may help dermatologists to detect the malignant lesions in the primary stages due to the minimum constraints during imaging, the ease of usage by the public and nonexperts, and high accuracy in detection of the lesion type.

Key words: *Classification, malignant melanoma, melanoma diagnosis, skin lesions*

INTRODUCTION

Melanoma is a malignant pigmented skin lesion which is the most deadly type of skin cancer in the world. This cancer is the sixth most common cancer among American men and women and is the main factor of cancer death in 25-30 years old women. Also, melanoma is the most common type of cancer in 20-44 years old men in Australia and New Zealand.[1] In Iran, according to the 8877 cases of skin cancer in 1385, this type of cancer was known as the first cancer.[2] On the other hand, moles that are natural parts of the skin are benign types of pigmented skin lesions. Characteristics of both benign and malignant pigmented skin lesions are similar which makes differentiating between them a challenging problem.[3]

Dermoscope lens and conventional digital camera are the most commonly used equipments that are used to investigate characteristics of pigmented skin lesions. The usage of each one of these equipments has advantages and disadvantages. The visualization of subsurface microstructures of the epidermis and upper dermis and uniform illumination are among the benefits of dermoscopic images but, on the other hand, the dermoscopy lens is not publicly available.[4,5] While nowadays, conventional digital cameras with spatial resolution higher than one megapixel are widely used by the general public and the taken images which are called macroscopic or clinical images are nonuniformly illuminated. Figure 1 shows macroscopic and dermoscopy images of an invasive melanoma. In this figure, the differences between visible characteristics of the two mentioned types of image, is clear. So, various computer processing techniques must be used for their analysis.[3]

The ultimate goal of procedures that are developed to distinguish between benign and malignant pigmented skin lesions is simplicity of applying by nonexperts and the general public. Hence, such procedures should be developed for macroscopic images with minimum constraints of imaging conditions. Some of these conditions include the usage of special resolution camera, consideration a predetermined distance between camera and skin surface, and the usage

Address for correspondence:
Dr. Alireza Karimian, Department of Biomedical Engineering, Faculty of Engineering, University of Isfahan, Isfahan, Iran.
E-mail: karimian@eng.ui.ac.ir

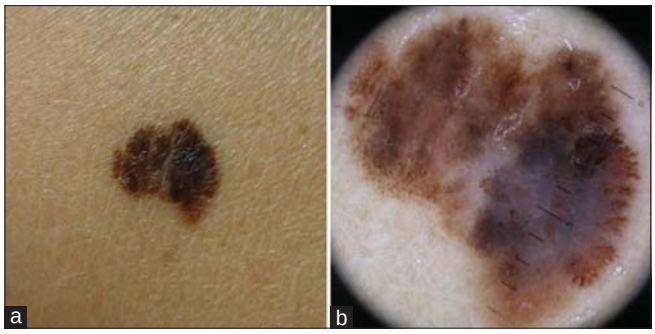

Figure 1: An invasive melanoma (a) Macroscopic image, (b) Dermoscopic image

of flash light during imaging. Up to now, studies in this area have been done assuming the above conditions and constraints are applied on the databases.[6-9]

In 2009, a study was done by Alcón *et al*. in which the algorithm to detect malignant melanoma from benign lesions by the usage of skin lesion macroscopic images is proposed. In this study, for lesion area segmentation, first the elimination of the low frequency spatial component of the image was used for background correction, and then a thresholding based method which was inspired by Otsu's algorithm, was applied to segment the lesion area. By considering ABCD criteria, 55 features were defined and extracted from the determined lesion area. Then correlation-based feature selection method and adaboost classifier were used as a feature selection step. In this algorithm, one decision support part was added which lead to the usage of the personal information including skin type, age, gender and part of the body along with the output of image classifier. Finally, 86% accuracy, 94% sensitivity and 68% specificity have been achieved.[6] In 2010, Christensen *et al*. proposed a procedure in which morphological operators were used for thick and thin hairs removal, pre and postprocessing. Otsu's thresholding algorithm was applied on blue channel of red, green and blue (RGB) color space locally to determine the lesion area, and then, 9 features describing the overall shape, border and color distribution were extracted. A prediction model was constructed based on statistical analyses of the algorithm outputs. Finally by applying an optimal threshold on output index score, 77% accuracy was achieved.[7] In 2011, the procedure is presented by Cavalcanti *et al*. in which shadow was estimated by adjusting a two degree quadric polynomial on normal skin and its effect was attenuated by removing this plane from the image. To determine the lesion area, a new three-channel image was defined, and a thresholding method inspired by Otsu's algorithm was applied on. Then by the usage of 52 extracted features, which were grouped in ABCD criteria features, and two k nearest neighborhood and decision tree classifiers in two modes, the lesion type was predicted. Finally, accuracy of 96.71%, sensitivity of 96.26% and specificity of 97.78% has been obtained.[8] In 2013, Cavalcanti *et al*. introduced 12 features based on the values of eumelanin and pheomelanin of the lesion and added them to the feature set which used in the previous study. In this way, the proposed procedure in that study resulted in 100% sensitivity, 97.78% specificity and 99.34% accuracy.[9]

The database of the mentioned studies was limited due to the conditions and constraints, which noted previously. This disadvantage prevents the proposed procedures from being appropriate to be applied on publicly available equipments that are the ultimate goal of proposing these procedures.

In this research work, a new procedure to detect malignant melanoma from benign pigmented skin lesions using macroscopic images taken by conventional digital camera with spatial resolution higher than one megapixel is presented, in which there is no constraint and special condition during imaging of the used database. In this procedure, a new method in order to weaken the effect of nonuniform illumination and also, a new threshold based algorithm in order to segment lesion area is described and applied on the database. Then, after introduction and applying new methods to correct the effect of thick hairs and large glows on the lesion, 187 features which indicate asymmetry, border irregularity, color variation, diameter and texture are extracted. The number of features is reduced using principal component analysis (PCA) algorithm and the result is used for predicting the type of lesion as benign or malignant using support vector machine (SVM) classifier.

METHODS

The proposed procedure has three stages in order to detect malignant melanoma from benign pigmented lesions. The first stage is preprocessing which includes removing effects of macroscopic images artifacts and determining lesion area with high accuracy. In the second stage, descriptor features of lesions are extracted and in the third stage which is called the classification stage, optimal features are determined and used to predict the type of lesions.

Database

The used database in this study is a set of 282 macroscopic images of pigmented skin lesions which had been collected from several online dermatology atlases such as dermnet, dermis and dermquest atlases.[10-16] This set includes RGB images of 149 benign lesions and 133 malignant which have various dimensions of 259 × 382 to 1186 × 1369 pixels. Whole area of the lesion in all of the images is visible, but lesion is not necessarily in the middle of the image and can be connected to image edges. These images are taken by conventional digital cameras with different spatial resolutions which are >1 megapixel. There was no need to adhere to a predetermined distance between the camera to skin while imaging and in some cases, flashlight is used. Thus, the used database in this study has the least restrictions and requirements for imaging.

Preprocessing

At this stage, the effects of part of artifacts in macroscopic images, including impact noise, skin lines, fine hairs, skin stains and small glows and reflections are removed by applying a median filter with mask size which is calculated using Eq. 1.[17]

$$n = \text{floor}\left(5.\sqrt{(M/768).(N/512)}\right) \quad (1)$$

In this equation, mask size n is determined for an M × N image and the floor function round down the result to the next integer.

Then, in order to weaken the effect of nonuniform illumination or shadow, image of original RGB color space is converted to hue, saturation and value (HSV) space because shadow effect in Value channel are more visible than other channels and spaces. Then, because of insensitivity of the proposed method to the location of the lesion, sampling from the normal skin is done twice as follows and as shown in Figure 2:

- Sampling from the four corners of V channel in 40 × 40 squares by taking into account the margin of 10 pixels from image edges
- Sampling from frame with a width of 40 pixels in sidelines of V channel by taking into account distance of 10 pixels from image edges.

In this way, two sets of values corresponding to the pixels of healthy skin are captured. Then 2 s and 3° polynomial functions, which are defined in equations 2 and 3, respectively, have been considered and are adapted on these two sets of samples by the least squares method.

$$Q_2(x,y) = P_1.x^2 + P_2.y^2 + P_3.xy + P_4.x + P_5.y + P_6 \quad (2)$$

$$Q_3(x,y) = P_1.x^3 + P_2.y^3 + P_3.x^2y + P_4.xy^2 + P_5.x^2 + P_6.y^2 + P_7.xy + P_8.x + P_9.y + P_{10} \quad (3)$$

In the equations 2 and 3, P_i (i = 1,…, 6 for Q_2 and i = 1,…, 10 for Q_3) determines quadric function parameter and (x, y) is image spatial location.

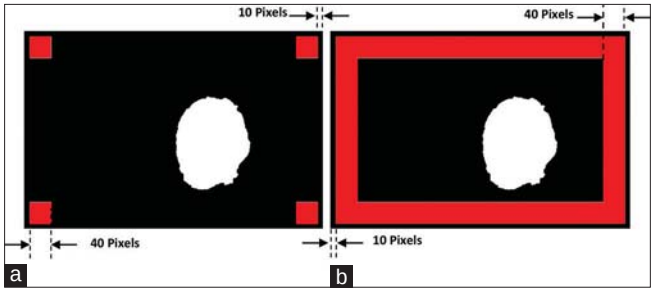

Figure 2: Sampling from the normal skin (a) From four corners, (b) From the frame

Thus, four different planes were estimated which represent four various modes of illumination distribution on the image with respect to the relative area of the lesion in image, location of lesion on body and the way of lighting while imaging. Then four *V* channels which have uniform illumination are obtained by dividing the original *V* channel on these four planes. Figure 3 shows the four estimated planes for a skin lesion image and the result of elimination of each one from the image.

If each one of these four processed V channel would be used for the following operations, images with uniform illumination are obtained that their healthy skin color is bright and different from the original image as can be seen in Figure 3. In order to retrieve true color of the skin, Eq. 4 is applied on each of processed V channels.[9]

$$V_{new}(x.y) = \frac{V_{proc}(x,y) \times \mu_{V_{orig}}}{\mu_{V_{proc}}} \quad (4)$$

In this equation, V_{proc} is the processed *V* channel, V_{orig} is the original *V* channel, μ represents mean of the respective channel and V_{new} is new *V* channel.

Then among new and original *V* channels, an image which has the least instability level, and therefore entropy, is

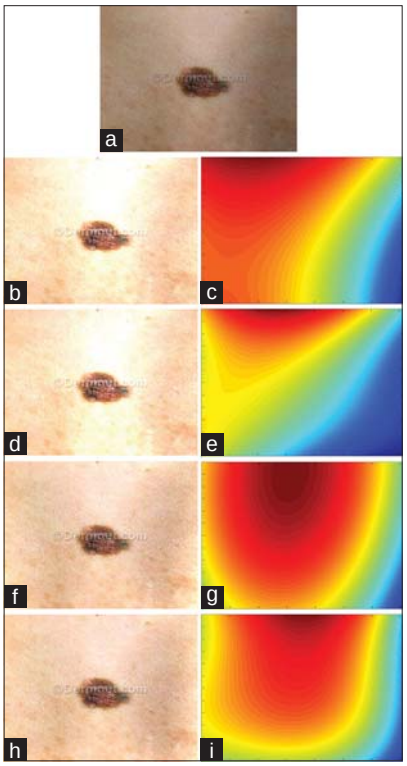

Figure 3: (a) Smoothed image of a skin lesion, (b and c) Results of adaption of two-degree polynomial function on the corners samples, (d and e) Results of adaption of three-degree polynomial function on the corners samples, (f and g) Results of adaption of two-degree polynomial function on the frame samples, (h and i) Results of adaption of three-degree polynomial function on the frame samples

chosen as the best *V* channel with uniform illumination because existence of shadow on the image leads to increased instability. This channel is replaced to the original *V* channel, and the final image is converted from HSV color space to RGB space.

Figure 4 shows the image of a skin lesion with shadow that the proposed median filter and shadow reduction method were applied on.

The second step in the preprocessing stage is segmentation of the lesion area from surrounding normal skin. For this purpose, a new, simple and accurate segmentation method, which is based on thresholding technique is introduced in which the single-channel images containing determinant factors of lesion border meaning color, illumination and texture are obtained firstly. Red channel of the RGB color space and l* channel of CIEl*u*v* color space represent color and brightness information, respectively, and the first component of principal components which are obtained from applying PCA method on the RGB image indicates texture information. The reason of using red channel as color represent or is the fact that each ethnic group has healthy skin color of reddish and skin lesions are regions of skin with altered color. The reason of using the first component is the fact that this component contains maximum changes in the image, and in skin lesion images, maximum changes as well as most of texture information occur on the lesion border.

Separation of lesion from healthy skin is more effective by using one of the three mentioned single-channel images which are determined by examining the histogram information. In general, histogram of skin lesion image has two peaks corresponding to healthy skin and lesion area which whatever they are farther and the valley between them is deeper, lesions area will be separated with higher accuracy from healthy skin. Therefore, a single-channel image is selected which distance between peaks of its smoothed histogram using local regression is maximum.

Four different thresholds are defined and calculated over the optimum single-channel image as follows:

- First threshold is calculated using Otsu thresholding algorithm ($level_o$)
- Second threshold that is the mean value of lesion and healthy skin distribution peaks of the histogram ($level_m$)
- Third threshold that is the starting point of healthy skin Gaussian distribution ($level_f$)
- Fourth threshold that is the point with the lowest height between lesion and healthy skin distribution on the histogram ($level_v$).

Then the thresholds on the image histogram which have the minimum distances to each other in terms of intensity level are selected and the largest one of them which covers results of other selected thresholds is applied on the optimal single-channel image. Since the shadow effect is corrected at first and thereafter, the threshold and borders are determined; shadow will not be mistaken by the lesion area and cannot affect on the borders determination.

Figure 5 shows a histogram of the optimum gray scale image of a skin lesion image with the four mentioned thresholds and the results of applying them and the optimal one. In the histogram of Figure 5a, the first and fourth thresholds completely matches and, therefore, are considered as the closest ones. Figure 5c shows results of using these two thresholds that indicates the lesion boundaries very accurate. As can be seen in Figure 5d-g, the boundaries of the second and third thresholds show large errors, while the selected thresholds by segmentation algorithm lead to the best results.

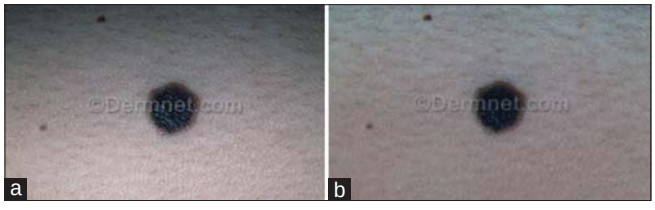

Figure 4: A skin lesion image (a) The original image, (b) The processed image after applying median filter and shadow reduction method

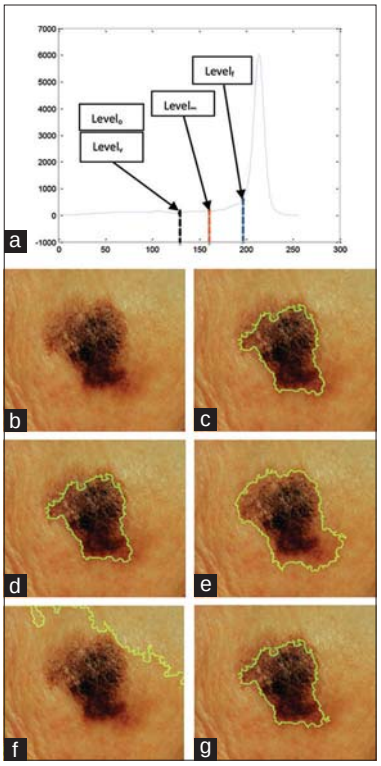

Figure 5: (a) Histogram of the optimal grayscale skin lesion image, (b) The preprocessed image of skin lesion, (c) Final result of segmentation, (d) Determined boundaries using the first threshold, (e) The second threshold, (f) The third threshold, (g) The fourth threshold

At the final step of the preprocessing stage, the effect of thick hairs and large glows and reflections on lesion is corrected to improve the results of segmentation algorithm and increase the accuracy and quality of determined boundaries. In order to correct the effect of thick hairs, bottom hat morphological transformation is applied and objects which their length to width ratio is >10 have been removed. This operation has been implemented with the assumption that hairs have long and narrow structures, while lesion has elliptical structure.

If the image has a lot of thick hairs, their remaining details on the border of lesion mask are corrected by applying morphological opening operator with circular structural element of size 3. Otherwise, this operator is not required to be applied. Following this, morphological closing operator with the same element as opening operator is applied on the whole images which removes indentations on the boundary caused by reflection of light from lines and dents of skin surface. Finally, number of pixels of each object in the image is calculated and the largest one is selected as the lesion mask.

At the end, if flash light is used, the effect of large glows and reflections on image will be corrected because intensity level of large areas on the image or its sides are increases due to intense light of flash. So the flash light effect will be checked just on the lesion area, its effect on the lesion border will be corrected and its effect on normal skin is not matter in this study. For this purpose, image is converted to cyan, magenta, yellow, and key color space and Y (yellow) channel which can show areas of glows in the best way is selected. Then elliptical-shaped area which contains entire lesion and part of surrounding healthy skin is determined as it is described in the following and the image is limited to it. For this purpose, the best-fit ellipse is defined, and length of its major and minor axes is increased to the size of largest Euclidean distance of border lesion and the defined ellipse, plus a constant value.

One of the reasons of the image limiting is that lesion area in more cases is much smaller than healthy skin and by limiting image to an ellipse, accuracy of separation glows areas on lesion will has a significant improvement. In addition, amount of processed data is also reduced which leads to an increased processing speed. Also, while defining an ellipse by increasing the lengths of ellipse axes and adding a constant value to them, it is ensured that its border is located on the healthy skin because ellipse border indicates healthy skin.

To determine glow area, k-means clustering algorithm is applied once on the limited channel Y and once more on the cluster with minimum center value, which is the output of the first run of the clustering algorithm. In each run of the algorithm, the number of clusters is selected equals to the number of smoothed histogram peaks of the input set, and five sequential iterations is performed. If the histogram has only one peak, the number of clusters is considered to be equal two. From results of second clustering, what is connected to the lesion border and is not connected to ellipse border is selected as glow mask. At the end, OR combination is applied on lesion and glow masks, and the final binary image is obtained after applying morphological closing and filling operators. Figure 6 shows the results of described sequence for determination of glow mask step by step.

Figure 7 shows the result of applying the described preprocessing step on three different skin lesions characterized by the lack and having a lot of thick hairs and large glows on lesion area.

Feature extraction

In this study similar to the traditional process of visual inspection, after determining lesion area, a set of main features is extracted from the area and are combined in order to distinguish between benign and malignant skin lesions. ABCD criteria are selected among methods used to diagnose melanoma because of measurability of its features using the information contained in macroscopic images. These criteria have four features of asymmetry, border irregularity, color variation and diameter.[18] In addition to these features, texture feature plays a decisive role in distinguishing between benign and malignant lesions. Finally, five groups of descriptors which are defined to measure these features are extracted and combined.

Features of asymmetry group are based on characteristics descripting center of gravity and inertia moments of lesion, which each one tries to measure the lesion asymmetry in the best way. This group includes 32 features such as orientation angle, asymmetry indices,[19-23] mean squared error of nonoverlapping area with the respect to major axes, eccentricity,[6,24] equivalent diameter,[6,8,25] circularity indices,[6,8] excircle and circumcenter index, sphericity index,[6] four features based on areas of both sides of the

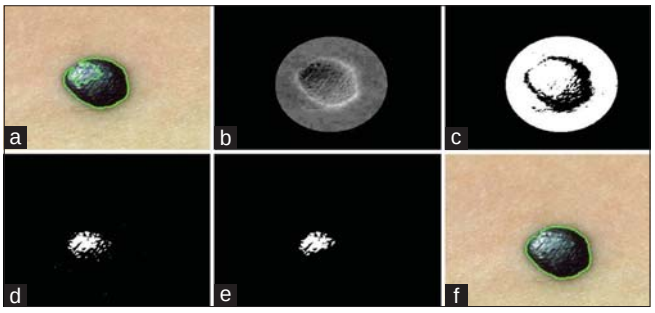

Figure 6: (a) A skin lesion with large glow on lesion area and the determined borders by lesion mask, (b) The limited Y channel, (c) The cluster with minimum center value (result of first run of clustering algorithm), (d) Result of second run of clustering algorithm, (e) Glow mask, (f) The determined borders by the final binary image

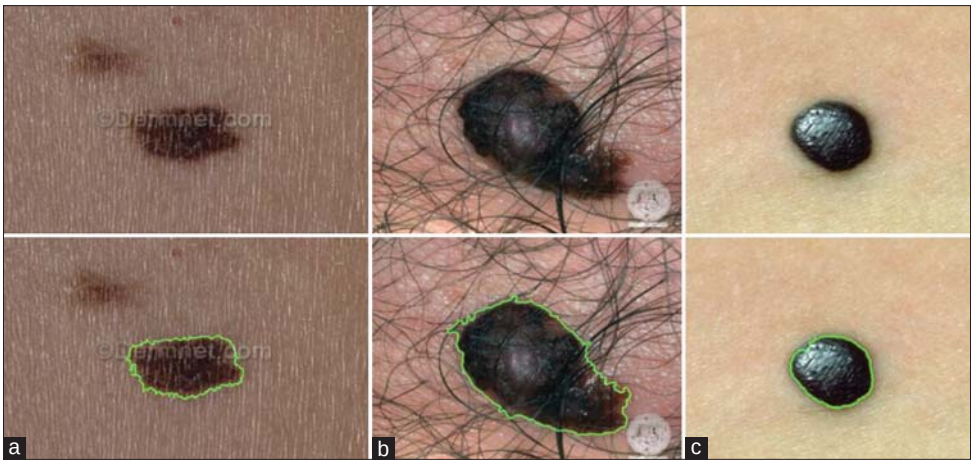

Figure 7: The original (top) and preprocessed (bottom) image of skin lesion (a) Without hairs and large glows, (b) With a lot of thick hairs, and (c) With large glows on lesion area

major axes,[8] normalized contour moments,[24] dimensionless moments,[26] extent index, elongation index[8] and area of bounding box. Border irregularity group is consisted of 34 features, which can be categorized in sets based on area and perimeter of the lesion, irregularity index, best-fit ellipse, convex hull, gradient and fractal geometry, and are area, two perimeters, four features based on radius,[25] border irregularity indices,[27,28] ccompactness index,[22] heywood circularity index, mean curvature,[7] best-fit ellipse indices,[7,29] bulkiness index,[30] bending energy, area and perimeter of convex hull, cconvexity index,[24] solidity index,[8] indentation and protrusion index, fractal dimensions[20,30] and ten border resolution.[6,8,20] Some of the features have different descriptions which are all considered.

Color variation group comprises 72 features of RGB and non-RGB color spaces components and gray scale image. Most descriptors of this group are statistical and are extracted from lesion mask which doesn't contain glows areas. Moreover, pixels with values <70% of the maximum of each channel are removed in calculations related to a healthy skin to ensure that there is no effect of hairs. Among the statistical characteristics of RGB color space components and gray scale image can be noted to the minimum, maximum, range of values, mean, standard deviation, coefficient of variation and variance and skewness, normalized standard deviation, ratio of mean values of RGB components, six basic colors counters,[6] relative chromaticity.[22] The statistical characteristics of non-RGB color space components are mean and standard deviation. These spaces include CIElch,[31] CIEl*a*b*,[32] HSI and spherical colour space which is defined by three Br, angle-α and angle-β components.[26]

Lesion diameter features contain 7 features of best-fit ellipse diameter, major diameter and the maximum distance between two nonadjacent points on the lesion border. Lesion texture features are extracted from gray level co-occurrence matrixes. These features include mean and range of values of 21 descriptors which are calculated for each of the four co-occurrence matrix for four different orientations of 0°, 45°, 90° and 135° and overall, describe 42 features for lesion texture. Among the co-occurrence matrixes descriptors can be noted to the auto-correlation, contrast, correlation, cluster prominence, cluster shade, dissimilarity, energy, entropy, homogeneity, maximum probability,[33] variance, sum average, sum entropy, sum variance, difference variance, difference entropy, information measures of correlation,[34] inverse difference, inverse difference normalized and inverse difference moment normalized.[35]

Typically, values of extracted descriptors are at different ranges which if set in a certain range leads to significantly improved classification performance. For this reason, values of descriptors are normalized using z-score conversion and Eq. 5.

$$Z_{i,j} = \frac{\left(f_{i,j} - \mu_j\right) / \left(3\sigma_j\right) + 1}{2} \quad (5)$$

In the above equation, $f_{i,j}$ is the amount of j-th descriptor of i-th image and μ_j and σ_j are mean and standard deviation of j-th descriptor, respectively. This conversion ensures that 99% of $Z_{i,j}$ values are in the range of zero and one. What is out of this range is rounded to zero or one.[36]

Classification

After the feature extraction stage, a set of high-dimensional data is obtained which high number of them is effective on the accuracy and required time for accurate classification. Moreover, extraction of this number of features will lead to a high computational cost in terms of time and storage. At this stage, the number of features in order to achieve better performance will be reduced.

For this purpose, the PCA, which is an unsupervised linear feature extraction method, is used. In the

PCA, due to the different units in the feature set, the correlation matrix is used instead of the covariance matrix. After implementing this process and calculating eigenvalues and variances, a set of principal components is obtained which are arranged in order to their ability in distinguishing between benign and malignant lesions. In order to determine the number of features which leads to the best classification results, the k number of sorted features is determined, and their efficiency is examined during classification. Finally, the features with maximum efficiency are selected.[24,37,38]

Classification is the last step in the computerized analysis of pigmented skin lesions images in which lesion is predicted as benign or malignant. According to the previous studies, SVM have performed well in the field of skin lesions classification. In addition, this algorithm has various parameters that different data models can be separated by changing them. Therefore, SVM with radial basis function kernel is used as the classifier in this study.

Radial basis function kernel has two parameters of C and γ which their optimal values are determined by a grid-search on two sequences of $C = \{2^{-5}, 2^{-4},\ldots, 2^{9}, 2^{10}\}$ and $\gamma = \{2^{-8}, 2^{-7},\ldots, 2^{3}, 2^{4}\}$. During the grid search procedure, ten-fold stratified cross-validation is performed to evaluate how well a particular combination of parameters is. After the grid-search, the target database is divided into two training and test sets. 70% of the database is used for training and the remaining 30% formed the test set. In both sets, the ratio of two benign and malignant classes is the same. Then SVM classifier with optimal parameters is trained and then tested on these two sets. In order to estimate the classification error, this procedure is performed 100 times and each time by changing the members of training and test sets, and the mean and standard deviation of the following evaluation criterion are calculated:

- Sensitivity: Percentage of patients who have been diagnosed correctly as patients.

$$\text{Sensitivity} = TP/(TP + FN) \quad (6)$$

Where *TP* and *FN* represent the number of patients who have been diagnosed correctly as patient and incorrectly as healthy, respectively.[39]

- Specificity: Percentage of healthy people who are correctly diagnosed as healthy.

$$\text{Specificity} = TN/(TN + FP) \quad (7)$$

Where *TN* and *FP* represent the number of healthy people who have been diagnosed correctly as healthy and incorrectly as patient, respectively.[39]

- Accuracy: Percentage of patient and healthy individuals who have been diagnosed correctly.[39]

$$Accuracy = (TP + TN)/(TP + FP + TN + FN) \quad (8)$$

- The area under the curve (AUC): Receiver operating characteristic curve is the trace of *TP* rate versus *FP* rate and AUC provides the expected performance of prediction as a numerical value.[40]

The described classification procedure is implemented on the database with all the extracted features and also, on the k number of features obtained by the reducing method. Then the optimal feature set which has the lowest number of members and also, the highest AUC value, is selected.

RESULTS

The results of the proposed methods in preprocessing stage have been examined by dermatologist. According to the medical doctor diagnosis, these methods detect the boundaries of lesions with high accuracy and determine the lesion area with accuracy of extent of 100% for the used database in this study.

In the classification stage, initial experiment was conducted on the database with all the extracted features. The optimal parameters which are found by the grid search and 10-fold cross-validation for this input set, have values of $(C^*, \gamma^*) = (32, 0.0078)$. Applying 100 times of SVM classifier with these optimal parameters on the training and test sets of 197 and 85 members respectively, leads to 81.13% ± 3.25% accuracy, 75.66% ± 6.87% sensitivity, 86.14% ± 5.27% specificity and 0.87 ± 0.03 area under the characteristic curve.

Then, with the goal of reducing the computation time and increasing the efficiency of the classifier, the classification procedure runs on the k features obtained by dimension reduction method. Due to the complexity of the problem in this study, it does not seem that small number of features has ability to make distinctions between classes very well. On the other hand, the large number of features may result in poor performance of the classifier. With these assumptions, k ranges between 5 and 60.

Figure 8 shows mean values of the AUC of 100 times classification with the optimal parameters versus the subset size of the principal components which are obtained by PCA. In this figure, it is observed that highest value of the AUC corresponds to the subset of principle components with size of 13, which has the mean value of 0.881. The mean and standard deviation of accuracy, sensitivity and specificity for this subset are 82.2% ± 3.57%, 77.02% ± 5.97% and 86.93% ± 5.46%, respectively which are obtained for optimal parameters $(C^*, \gamma^*) = (256, 0.0078)$. In Figure 9 which shows the mean values of accuracy, sensitivity and specificity versus the size of principal components subsets, the mentioned values can be observed.

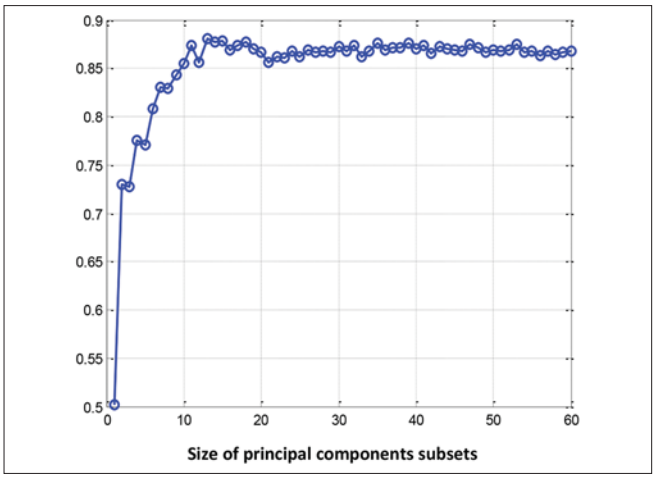

Figure 8: The mean values of area under the curve versus the size of principal components subset

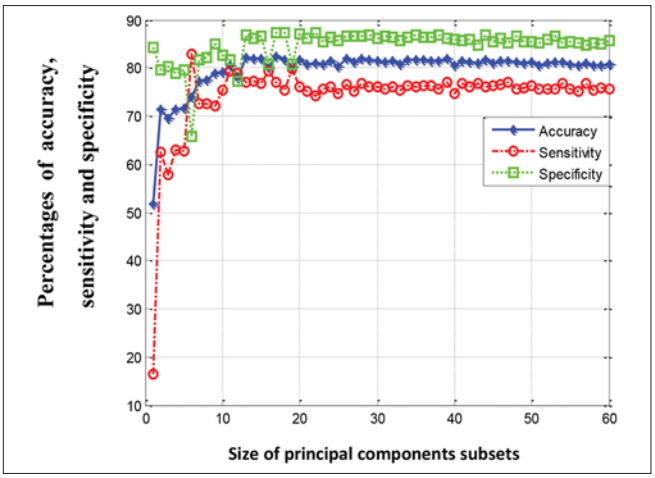

Figure 9: The mean values of accuracy, sensitivity and specificity versus the size of principal components subset

Table 1: The classification results for all features and the optimal features determined by feature selection method

Feature selection algorithm	Number of features	AUC	Accuracy*	Sensitivity*	Specificity*
-	187	0.87	81.13±3.25	75.66±6.87	86.14±5.27
PCA	13	0.881	82.2±3.57	77.02±5.97	86.93±5.46

*The results are listed as mean±SD. SD – Standard deviation; AUC – Area under the curve; PCA – Principal component analysis

Table 1 shows results of classification for the optimal number of features selected by the feature selection method and also, for all the extracted features. By comparing the values listed in this table, it is observed that values of evaluation criteria are increased using 13 principal components which are extracted by PCA method than 187 original features which are extracted from lesion area. As a result, much smaller number of features can lead to better classification results.

DISCUSSION

This paper presents a new procedure for classifying pigmented skin lesions as benign or malignant using macroscopic images, which are taken by conventional digital cameras with spatial resolution higher than one megapixel. While imaging the used database, any constraints and specific conditions are avoided that is an important difference between this study and the previous ones in this area and makes the proposed procedure appropriate for implementation by public and nonspecialists.

In this study, new methods to enhance the quality of processing and analysis of macroscopic images of skin lesions have been proposed; including new method which weakened effect of nonuniform illumination on the image in the best way, a new thresholding based algorithm which by review the existing information on the image histogram exploits extractable information of the image and a new method which corrects effect of thick hairs and large glows on the lesion that appear while imaging using flash light and greatly increases accuracy of the boundaries set by the segmentation algorithm. In this study, 187 features representing asymmetry, border irregularity, color variation, diameter and texture which are the maximum number of extractable features from the lesion are extracted and by using the PCA algorithm, 13 optimal features are selected. Finally, SVM classifier predicts lesion types with accuracy of 82.2%, sensitivity of 77% and specificity of 86.93%. Because of dissimilarity between the used databases in this study and the other ones, the achieved results cannot be compared. However the accuracy improved significantly against the 64% accuracy of naked eye specialist, which is a worthy conclusion. According to the dermatologist report, the proposed method in this research due to its sensitivity, accuracy and specificity may help dermatologists in detection the malignant melanoma in more priority stages which may help their treatment more effectively. Moreover, if there was access to the personality and imaging information such as tumor site, patient's skin and eye color, the distance between camera and skin and the lesion diameter which is a limitation for this procedure, the achieved results could be improved significantly.

REFERENCES

1. Kutlubay Z, Engin B, Serdaroğlu S, Tüzün Y. Current management of malignant melanoma: State of the art. In: Vereecken DP, editor. Highlights in Skin Cancer. Istanbul: InTech.; 2013. p. 67-125.
2. Azly M, Ramezani R, Dounlu M, Sedighi Z, Dabiry A, Partoie A. National Report of Cancer Registration 1385. Tehran: Ministry of Health and Medical Education; 1386. p. 621.
3. Korotkov K, Garcia R. Computerized analysis of pigmented skin lesions: A review. Artif Intell Med 2012;56:69-90.
4. Rigel DS, Russak J, Friedman R. The evolution of melanoma diagnosis: 25 years beyond the ABCDs. CA Cancer J Clin 2010;60:301-16.
5. Wurm EM, Soyer HP. Scanning for melanoma. Aust Prescr 2010;33:150-5.
6. Alcón JF, Ciuhu C, Kate W, Heinrich A, Uzunbajakava N, Krekels G, *et al*.

Automatic imaging system with decision support for inspection of pigmented skin lesions and melanoma diagnosis. IEEE J Sel Top Signal Process 2009;3:14-25.
7. Christensen JH, Soerensen MB, Linghui Z, Chen S, Jensen MO. Pre-diagnostic digital imaging prediction model to discriminate between malignant melanoma and benign pigmented skin lesion. Skin Res Technol 2010;16:98-108.
8. Cavalcanti PG, Scharcanski J. Automated prescreening of pigmented skin lesions using standard cameras. Comput Med Imaging Graph 2011;35:481-91.
9. Cavalcanti PG, Scharcanski J, Baranoski GV. A two-stage approach for discriminating melanocytic skin lesions using standard cameras. Expert Syst Appl 2013;40:4054-64.
10. Cohen BA, Cohen MR, Janjua SA, Hosler G, Waseem M. Dermatlas Beta. Available from: http://www.dermatlas.org/. [Last cited on 2014 Feb 02].
11. Skin Disease Atlas. Dermnet: Portsmouth, NH; 1998. Available from: http://www.dermnet.com/. [Last cited on 2014 Feb 02].
12. National Cancer Institute. NCI Visuals Online. Available from: https://www.visualsonline.cancer.gov. [Last cited on 2014 Feb 02].
13. Oakley A. DermNet NZ: The Dermatology Resource. Available from: http://www.dermnetnz.org/. [Last cited on 2014 Feb 02].
14. Hexsel DM, Eastern J, Feldman SR, Spencer JM, Miyachi Y, Aractingi S, Pandya AG. Personalised learning and teaching resources for dermatologists today and tomorrow. Available from: http://www.dermquest.com. [Last cited on 2014 Feb 02].
15. Silva SF, Calheiros DB. Dermatology Atlas. Available from: http://www.atlasdermatologico.com.br/. [Last cited on 2014 Feb 02].
16. DOIA Team. Dermatology Information System. Available from: http://www.dermis.net. [Last cited on 2014 Feb 02].
17. Celebi ME, Kingravi HA, Iyatomi H, Aslandogan YA, Stoecker WV, Moss RH, *et al.* Border detection in dermoscopy images using statistical region merging. Skin Res Technol 2008;14:347-53.
18. Friedman RJ, Rigel DS, Kopf AW. Early detection of malignant melanoma: The role of physician examination and self-examination of the skin. CA Cancer J Clin 1985;35:130-51.
19. Cheung K. Image Processing for Skin Cancer Detection: Malignant Melanoma Recognition [Master's Thesis]. Canada: National Library of Canada; 1997. p. 109.
20. Zagrouba E, Barhoumi W. A prelimary approach for the automated recognition of malignant melanoma. Image Anal Stereology 2004;23:121-35.
21. She Z, Liu Y, Damatoa A. Combination of features from skin pattern and ABCD analysis for lesion classification. Skin Res Technol 2007;13:25-33.
22. Parolin A, Herzer E, Jung CR. Semi-automated diagnosis of melanoma through the analysis of dermatological images. In: The 23rd SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI); 2010. Gramado: IEEE; 2010. p. 71-8.
23. Stoecker WV, Li WW, Moss RH. Automatic detection of asymmetry in skin tumors. Comput Med Imaging Graph 1992;16:191-7.
24. Celebi ME, Aslandogan YA. Content-based image retrieval incorporating models of human perception. In: Proceedings of International Conference on Information Technology: Coding and Computing (ITCC 2004); 2004 April, 5-7: IEEE; 2004. p. 241-45.
25. Manousaki AG, Manios AG, Tsompanaki EI, Panayiotides JG, Tsiftsis DD, Kostaki AK, *et al.* A simple digital image processing system to aid in melanoma diagnosis in an everyday melanocytic skin lesion unit: A preliminary report. Int J Dermatol 2006;45:402-10.
26. Kusumoputro B, Ariyanto A. Neural network diagnosis of malignant skin cancers using principal component analysis as a preprocessor. In: The 1998 IEEE International Joint Conference on Neural Networks Proceedings. IEEE World Congress on Computational Intelligence; 1998 May, 4-8. Anchorage, AK: IEEE; 1998. p. 310-15.
27. Kjoelen A, Thompson MJ, Umbaugh SE, Moss RH, Stoecker WV. Performance of AI methods in detecting melanoma. IEEE Eng Med Biol Mag 1995;14:411-6.
28. Maglogiannis IG, Zafiropoulos EP. Characterization of digital medical images utilizing support vector machines. BMC Med Inform Decis Mak 2004;4:4.
29. Chang Y, Stanley RJ, Moss RH, Van Stoecker W. A systematic heuristic approach for feature selection for melanoma discrimination using clinical images. Skin Res Technol 2005;11:165-78.
30. Claridge E, Hall PN, Keefe M, Allen JP. Shape analysis for classification of malignant melanoma. J Biomed Eng 1992;14:229-34.
31. X-Rite, Inc. A Guide to Understanding Color Communication. USA: X-Rite; 2007. p. 24.
32. Plataniotis KN, Venetsanopoulos AN. Color Image Processing and Applications. Color Spaces. Ch. 1. Berlin Heidelberg: Springer; 2000. p. 1-49.
33. Soh LK, Tsatsoulis C. Texture analysis of SAR sea ice imagery using gray level co-occurrence matrices. IEEE Trans Geosci Remote Sens 1999;37:780-95.
34. Haralick RM, Shanmugam K, Dinstein IH. Textural features for image classification. IEEE Trans Syst Man Cybern 1973;SMC-3:610-21.
35. Clausi DA. An analysis of co-occurrence texture statistics as a function of grey level quantization. Can J Remote Sens 2002;28:45-62.
36. Aksoy S, Haralick RM. Probabilistic vs. geometric similarity measures for image retrieval. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2000 Jun, 13-15. Hilton Head Island, SC: IEEE; 2000. p. 357-62.
37. Han J, Kamber M. Data Mining: Concepts and Techniques. 2nd ed. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.; 2006. p. 743.
38. Murtagh F, Heck A. Multivariate Data Analysis with Fortran, C and Java Code. Northern Ireland: Queen's University Belfast, Astronomical Observatory Strasbourg; 2000. p. 272.
39. Zhu W, Zeng N, Wang N. Sensitivity, specificity, accuracy, associated confidence interval and ROC analysis with practical SAS® implementations. In: Proceedings of NESUG: Health Care and Life Sciences; 2010 Nov, 14-17. Baltimore, Maryland; 2010. p. 9.
40. Huang J, Ling CX. Using AUC and accuracy in evaluating learning algorithms. IEEE Trans Knowl Data Eng 2005;17:299-310.

How to cite this article: ???

Source of Support: Nil, **Conflict of Interest:** None declared

BIOGRAPHIES

Maryam Ramezani received B.Sc. degree in electronics engineering from Isfahan University of Technology, Isfahan, Iran, in 2011, and she received M.Sc. degree in biomedical engineering from University of Isfahan, Isfahan, Iran, in 2014.

E-mail: m.ramezani85@yahoo.com

Alireza Karimian received his B.Sc. degree in electronics engineering from Ferdowsi University, Mashhad, Iran. He also received his M.Sc. and Ph.D. degrees in nuclear engineering in the field of medicine from Amirkabir University of Technology, Tehran, Iran. Furthermore He has passed successfully a one year fellowship research in the field of medical physics in the La Sapienza University, Rome, Italy under grant of ICTP. In 2006, He joined the Department of Biomedical Engineering at University of Isfahan, Iran, as an Assistant Professor, and was the lecturer of some courses such as medical imaging systems, Simulation and its application in medicine, radiation shielding, Dosimetry and radiation detection, Biophysics and medical physics. Since February of 2013, He was successful to be as Associate Professor in University of Isfahan. He has published more than 130 research papers in peer-reviewed journals and conferences. His research interests are: Imaging systems - Image processing – Dosimetry - Radiotherapy- Monte Carlo simulation and its applications in medicine.

E-mail: Karimian@eng.ui.ac.ir

Payman Moallem received B.Sc. and M.Sc. degrees in electronics engineering from Isfahan University of Technology, Isfahan, Iran, in 1992, and Amirkabir University of Technology, Tehran, Iran, in 1996, respectively. He also received a PhD degree in electrical engineering from Amirkabir University of Technology. In 2003, he joined the Department of Electrical Engineering at University of Isfahan, Iran, as an assistant professor, and was promoted clude image processing, machine vision, neural networks, pattern recognition, intelligent systems, and real-time signal and video processing. Since 2006, he has been a member of the editorial boards of Majlesi Journal of Electrical Engineering and Majlesi Journal of Multimedia Processing. He has published more than 235 papers in peer-reviewed journals and conferences.

E-mail: P_moallem@eng.ui.ac.ir