

RESEARCH

Open Access



Melanoma recognition using extended set of descriptors and classifiers

Michał Kruk^{1*}, Bartosz Świdorski¹, Stanisław Osowski^{2,3}, Jarosław Kurek¹, Monika Słowińska⁴ and Irena Walecka⁴

Abstract

The paper presents a novel method of melanoma recognition on the basis of dermoscopic images. We use color images of skin lesions, advanced image processing, and different classifiers to distinguish melanoma from the other non-melanoma lesions. Different families of descriptors are used for generation of the image diagnostic features for final pattern recognition. To increase the efficiency of the system, we apply different selection procedures to find the best set of features and different solutions of classifier. The numerical results concerning the accuracy of the proposed recognition system have confirmed good accuracy of the proposed method and high sensitivity in melanoma recognition.

Keywords: Melanoma recognition, Diagnostic features, SVM, Random forest

1 Introduction

Melanoma is a potentially life-threatening neoplasm. It is manifested by the growing, unusual-looking skin lesion, of the odd-shaped, uneven, or uncertain borders and multiple colors in advanced cases. Thin melanomas a few millimeters in diameter can mimic benign nevi and cannot be detected by the “naked eye” examination. The only possibility to diagnose them is using the dermoscopy as a tool. Early recognition and surgical excision can be curative for the patient.

However, the number of yearly deaths from melanoma continues to increase, and the overall melanoma mortality rate is one of the few cancer mortality rates not on the decline [1–3]. These realities combined with increasing evidence of the lack of efficacy of the clinically assessed ABCDE criteria (“A” for “asymmetry,” “B” for “border irregularity,” “C” for “color variation,” “D” for “diameter,” and “E” for “evolving lesions”) have necessitated ongoing efforts to enhance the earlier clinical detection of melanoma [3, 4].

Most approaches to melanoma diagnosis have included emphases on recognition of changing lesions, recognition of outlier (“ugly duckling”) lesions, and specific melanoma features, with the most utilized criteria being the ABCDE descriptors. Some recently published

strategies have rejected the diameter criterion as well as abandoned all or portions of the ABCDE mnemonics [3–5]. The additional problems with application of the ABCDE descriptors appear for extensive lesions, for which the borders of lesions are outside the dermoscopic image or there is a smooth transition between the lesions and the healthy skin. Therefore, the development of other diagnostic features well characterizing the skin lesions is needed.

There is a growing interest in developing automatic systems which support the dermatologists in early recognition of melanoma [6]. Such systems include composition of few main steps: (1) image segmentation, (2) feature extraction and selection, and (3) lesion classification. Recently, many approaches to these topics have been proposed. The paper [7] proposed new mathematical descriptors for the border of pigmented skin lesion images, like lesion slope and lesion slope regularity. The other works proposed different approaches to melanoma segmentation and characterization. They include color clustering [6], wavelet analysis [5], Markov tree features [8], use of color texture [9], application of global and dynamic thresholding [10], GVF snakes [10], etc.

Different classifier solutions applying the selected descriptors have been proposed. They include clustering approach, linear discriminant analysis, neural networks, fuzzy and neuro-fuzzy systems, support vector machines (SVM), K-nearest neighbors (KNN), naïve Bayes, random

* Correspondence: michal_kruk@sggw.pl

¹Warsaw University of Life Sciences, Warsaw, Poland

Full list of author information is available at the end of the article

forest, etc. Ganster et al. [6] has achieved a sensitivity of 87 % and a specificity of 92 % for a large data set with more than 5300 dermoscopy images. Recent results show a sensitivity of 83.06 % and specificity of 90.05 % of cascade classifiers in tenfold cross-validation mode for recognition of melanoma in clinical images [11]. The research of melanoma reported in [5] show the accuracy of 91.26 % and area under the curve (AUC) value of 0.937 on the set of 289 dermoscopic images (114 malignant, 175 benign) partitioned into train, validation, and test image sets. The paper [8] has declared the accuracy of SVM classifier varying from 40 to 75 %, depending on the test performed and features used. The paper [12] has reported the sensitivity of 93 % and specificity of 92 %. The work [13] has reported the accuracy rate changing from 69.9 up to 93.7 % depending on the combination of training and testing results. The paper [14] has reported the accuracy of 85 % for recognition of non-invasive melanomas based on the ABCD rule and pattern-recognition image-processing algorithms. The other paper [15] has compared the application of neural and neuro-fuzzy networks to skin cancer recognition, reporting the accuracy rate of 90.67 % for neural and 91.26 % for neuro-fuzzy networks. The recent paper [16] has investigated two systems (global and local) of detection of melanoma and declared the sensitivity of 96 % and specificity of 80 %.

In this paper, we propose the application of different types of image descriptors to characterize the dermoscopy image of the skin lesions. Textural features are based on the standard Haralick descriptors [17], while statistical features apply the segmentation-based fractal texture analysis (SFTA) [18, 19], Kolmogorov–Smirnov distance [20, 21], percolation descriptors [22], and maximum subregion descriptors [20]. The selection procedure based on application of Fisher discrimination measure, feature correlation, and fast correlation-based filter is used to choose the features, which is able to recognize melanoma with the best accuracy. The selected descriptors are used as the input attributes to the system of classifiers, responsible for the final recognition of melanoma.

2 Materials

Input data were taken from the Department of Soft Tissue/Bone Sarcoma and Melanoma, Warsaw Memorial Cancer Center and Institute of Oncology, as dermoscopic images. We obtained 92 RGB images of non-melanoma and 84 images of melanoma (176 cases in total). They were acquired by dermatologists during clinical exams using a dermatoscope of the magnification of $\times 20$. The data base was medically assessed by the expert dermatologists on the basis of the ABCDE dermoscopic criteria and exact pathomorphological inspection, including medical segmentation of the lesion and clinical and histological

diagnosis. The detailed contents of the database based on their assessment are presented in Table 1. The registered images of the lesions were of different sizes extending from 465×599 to 1077×1899 pixels. They were stored in JPEG format.

The representative examples of both class images, which participated in our experiments, are shown in Fig. 1. It is easy to observe that the standard ABCD criteria cannot be applied directly to images of melanoma because the borders of the lesions have not been registered or it is impossible to segment the nevus region from the skin background with a satisfactory precision.

The additional experiments have been performed using the data set PH2 available in the Internet [23]. The dermoscopic images forming the basis were obtained by the Dermatology Service of Hospital Pedro Hispano (Matosinhos, Portugal) using the Tuebinger Mole Analyzer system at the magnification of $\times 20$. They are 8-bit RGB color images of a resolution of 768×560 pixels. The database contains 200 dermoscopic images of melanocytic lesions, including 80 common nevi, 80 atypical nevi, and 40 melanomas. The PH2 database includes medical annotation of all the images based on medical segmentation of the lesion, clinical and histological diagnosis, and the assessment of several dermoscopic criteria (colors, pigment network, dots/globules, streaks, regression areas, blue-whitish veil). The assessment of each parameter was performed by an expert dermatologist [23].

3 Methods

We propose the computerized system implementing few stages of processing to recognizing melanoma on the basis of the color images of the skin lesions. The first stage is an acquisition of the original RGB image containing the lesions area using dermoscopy. The next step is image filtering, aimed at minimizing the influence of the noise, like thin hair or small air bubbles. The following step is generation of the numerical descriptors (diagnostic features) of the image, which represent the potential input attributes to the classifier system. These features undergo the assessment of their class discrimination ability in the feature selection process. The selected features are treated as the input attributes to the classification

Table 1 The database of skin lesions used in experiments

Type of lesions	Number of samples	
Melanoma	Lentigo maligna melanoma	69
	Nodular melanoma	17
Non-melanoma	Seborrheic keratosis	4
	Angioma	1
	Pigmented nevus	29
	Atypical nevus	59

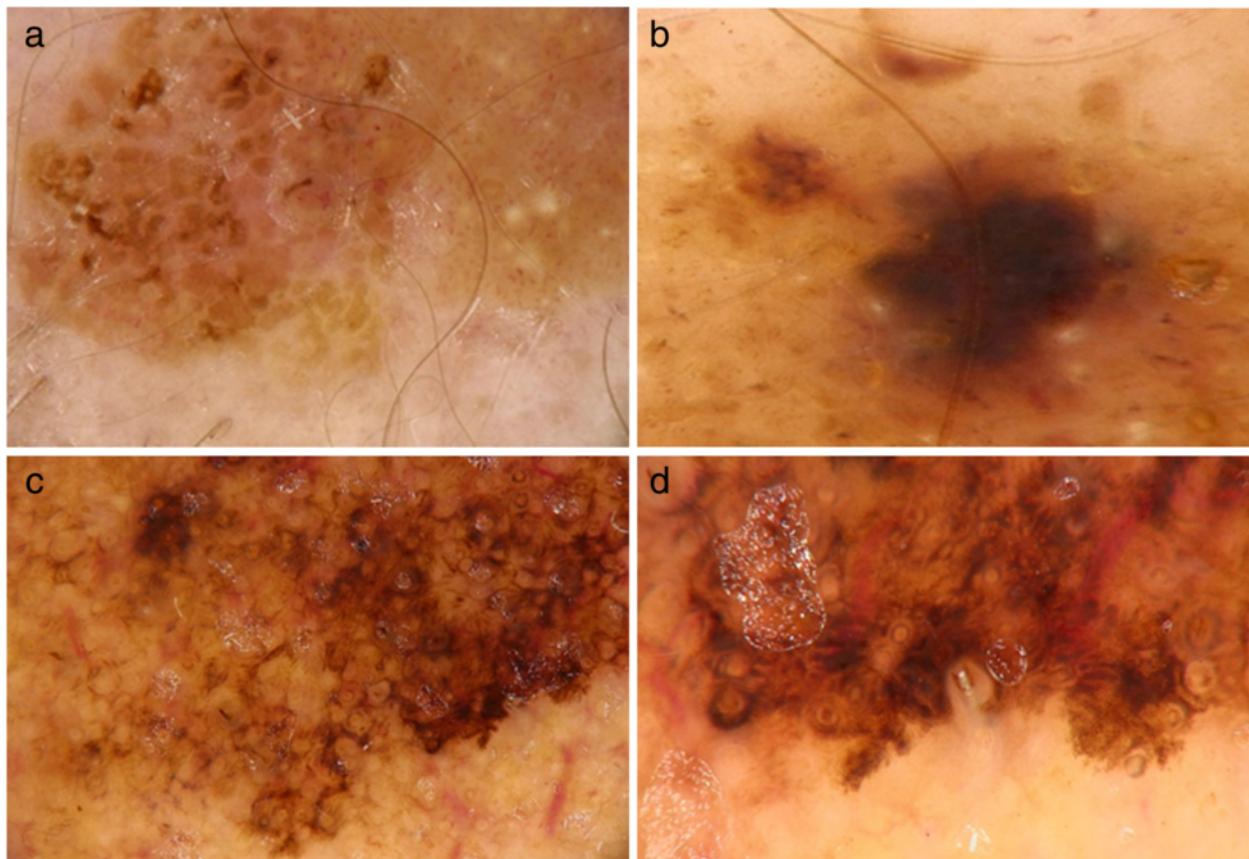


Fig. 1 The examples of input data. The left-side image (a, c) presents the non-melanoma cases and right-side image (b, d) the melanoma cases

system, responsible for the final recognition of the non-melanoma (first class) and melanoma cases (second class). The general scheme of the proposed system is presented in Fig. 2.

3.1 Image filtering

Image filtering is aimed at removing small structures and artifacts from skin image to reduce future over-segmentation in further processing steps of the image. The artifacts are treated as an impulse noise and are removed by applying median filtering.

In the presence of thick hair, the median filtering may be not sufficient. Therefore, we have applied an additional

procedure based on the improved DullRazor technique [24]. It identifies the hair areas and replaces the hair pixels by nearby non-hair pixels. The exemplary result of initial filtering of the skin lesions is presented in Fig. 3.

3.2 Generation of diagnostic features

To create the effective classification system, we have to generate the appropriate set of diagnostic features, which will form the input signals to the classifier. Good features should allow distinguishing different classes with the highest precision. It means that they should assume similar values for the images belonging to the same class and different values for the representatives

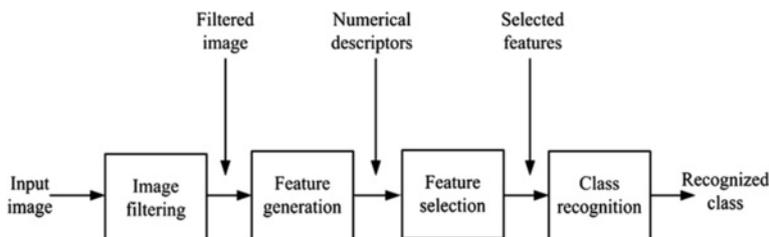


Fig. 2 The proposed system of melanoma recognition

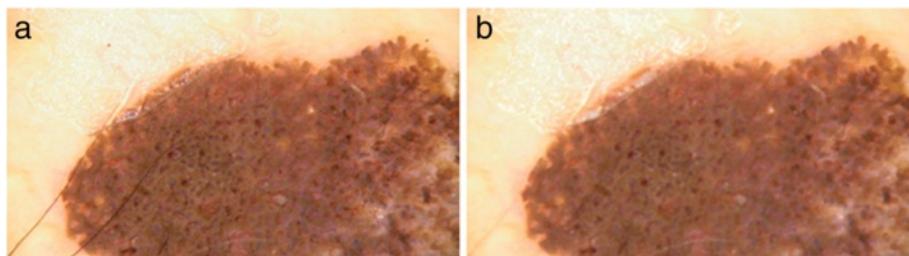


Fig. 3 The illustration of filtering the skin lesion image. **a** Original image. **b** Result of filtering

of the opposite class. In the proposed solution, we will exploit the statistical and textural descriptions of the image. They are divided into few subgroups: the numerical descriptors based on the Kolmogorov–Smirnov (KS) statistical distance [20, 21], maximum subregion principle, percolation theory [22], classical Haralick descriptors [17], and descriptors based on fractal texture analysis [18].

3.2.1 Kolmogorov–Smirnov descriptors

Kolmogorov–Smirnov descriptors reflect the changing distribution of intensity of pixels placed in the rings of the increasing geometrical distances from the central point [20, 21]. The exemplary division of the images into coaxial rings is illustrated in Fig. 4. It represents four concentric rings of equal number (56) of pixels in each ring (equal number of pixels causes more stable distribution of KS statistics).

The central point is traveling along pixels distributed uniformly every ten pixels in the image. The results of statistical analyses of these coaxial rings will be combined together by concatenating the pixel intensities corresponding to the rings placed in equal distances from each other and at different positions of the central pixel. Then, the cumulative KS distance [21] between the intensity of pixels x_i and x_j belonging to two different rings using the KS test is estimated. The KS statistics

determines if the samples of both rings are drawn from the same continuous population characterized by the cumulative distributions $F(x_i)$ and $F(x_j)$. The distance between these two populations is defined in the KS test as

$$d_{KS} = \max |F(x_i) - F(x_j)| \tag{1}$$

over all x . This distance is treated as the measure of difference between the distributions of both populations.

The KS distance for all combinations of two rings is calculated. As a result, we get a set of KS statistics corresponding to different levels of such differences. Level 1 corresponds to KS differences of the succeeding rings, i.e., rings 1 and 2, 2 and 3, 3 and 4, etc. Level 2 corresponds to the KS differences of rings distant by 2, for example 1 and 3, 2 and 4, etc. As a result, we collect the KS distances corresponding to the same differences of rings for each level.

Figure 5 presents the mean values of d_{KS} for different levels l and its linear regression estimated for the image of Fig. 1b. The measured (known) values of d_{KS} are given by three square points and a solid line while its linear approximation by the dashed line. The horizontal axis

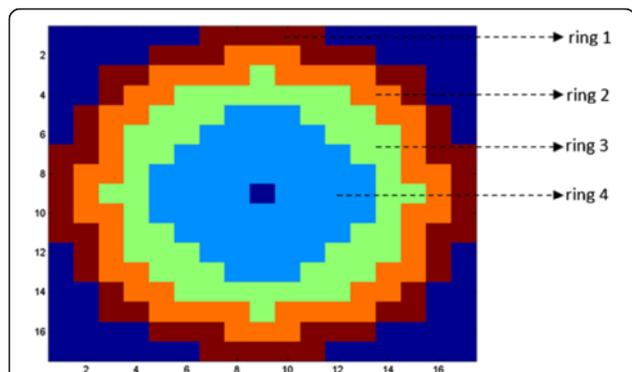


Fig. 4 The illustration of four coaxial rings around the central pixel

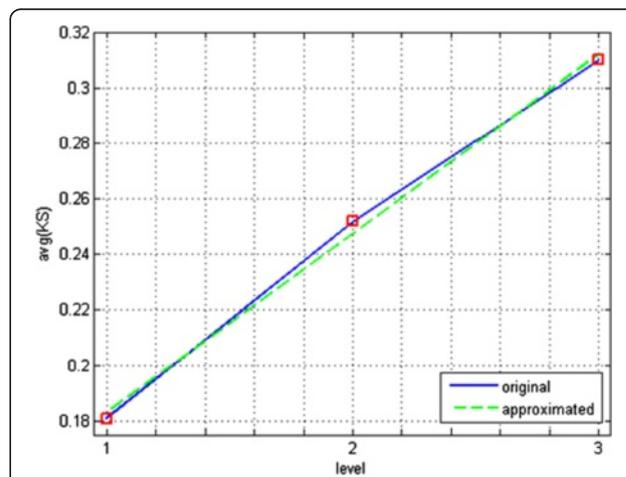


Fig. 5 The linear fit of the relationship of KS distance versus the levels of differences between the rings

represents the succeeding levels l and the vertical one the average KS distance.

As the features, we have assumed:

- a) d_{KS12} (the mean of KS statistics between ring no. 1 and ring no. 2)
- b) d_{KS13} (the mean of KS statistics between ring no. 1 and ring no. 3)
- c) d_{KS14} (the mean of KS statistics between ring no. 1 and ring no. 4)
- d) The ratio d_{KS13}/d_{KS12}
- e) The ratio d_{KS14}/d_{KS12}
- f) The coefficient α_0 of the approximation line $d_{KS} = \alpha_0 + \alpha_1 l + \varepsilon$
- g) The slope coefficient α_1 of the approximation line $d_{KS} = \alpha_0 + \alpha_1 l + \varepsilon$

In this way, the total population of KS features is equal to seven.

3.2.2 Maximum subregion descriptors

This set of descriptors is generated by applying thresholding and splitting the grayscale image into smaller consistent subgroups. We search for such level of the threshold value which creates the largest number of the compact subgroups of pixels of the intensity either lower or higher than the threshold value. For computational reason, this searching will be conducted using percentiles of the pixel's intensity. Let us assume that the threshold value th_a corresponds to the p th percentile for the grayscale image of the intensity changing from 0 to 255. We scale this threshold value to the range of percentiles from 1 to 99 (to avoid the effect of saturation of black and white colors). For each image, we determine the intensity levels f_1 corresponding to the first percentile and the intensity f_{99} associated with the 99th percentile. Then, the normalized threshold nth_a is recalculated according to the formula

$$nth_a = (th_a - f_1) \frac{255}{f_{99} - f_1} \tag{2}$$

The value of percentile p th and the corresponding normalized threshold nth_a represent the features. The additional descriptor is the area (in pixels) of the largest subgroup in the image after thresholding. Because of two types of thresholding procedures (the intensity of compact subgroups of pixels higher or lower than the threshold value), the number of these features is duplicated (six features in total).

3.2.3 Percolation descriptors

These descriptors are based on percolation theory [22] and are defined on the grayscale image. The main idea

of the method is to segment the image into subimages and then “set fire” inside of the square region of nine pixels (central pixels and its eight neighbors). Assuming that fire can be spread simultaneously on all subimages created in the thresholding process, we observe the duration of the longest fire, measured by the number of iterations. The threshold value is changed successively in the grayscale intensity range [0 255] according to the decile steps from $q = 1$ up to $q = 9$. The more jagged image the longer fire duration. For each threshold, we note the duration of fire. As a feature, we define the weighted average indicator q_w of quantiles

$$q_w = \frac{\sum_{i=1}^9 q_i d_i}{\sum_{i=1}^9 d_i} \tag{3}$$

where q_i is a quantile changing from 0.1 to 0.9 with step equal to 0.1 and d_i is the duration of fire at the threshold value corresponding to the i th decile. The segmentation of the image may be continued on the pixels of the intensity higher or lower than the threshold value. In this way, we can define two features q_w corresponding to these two percolation processes.

Table 2 presents the results of the percolation process performed on the image of Fig. 6. The values q_i and d_i represent the deciles and duration of fire, respectively, at the threshold value corresponding to the i th decile, calculated for the regions corresponding to the pixel intensity higher than the threshold value.

In this case, the descriptor q_w takes the value 0.469.



Fig. 6 The exemplary image subject to percolation process

Table 2 The duration of “fire” of the image of Fig. 6 as a function of the threshold values measured in quantiles q_i changing from 0.1 to 0.9

q_i	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
d_i	31	40	31	30	40	24	29	27	22

3.2.4 Haralick descriptors

Haralick (GLCM) descriptors of the texture are based on the co-occurrence matrix concept [17] and focus on the relationships among the intensity levels of the neighboring pixels in the image. In this application, we have limited our considerations to the statistics concerning the local contrast of the image, correlation of pixel pairs, energy representing the occurrence of repeated pairs within an image, and homogeneity coefficient characterizing the distribution of the elements of a GLCM with its diagonal. They have been generated separately for three RGB channels of the color image. Up to 48 features have been defined in this way.

3.2.5 Descriptors based on fractal texture analysis

The next set of image descriptors uses the segmentation-based fractal texture analysis (SFTA) [18]. The input grayscale image is decomposed into a set of binary images by selecting pairs of lower and upper threshold values (multi-level Otsu algorithm) for each region until the desired number n_t of thresholds is obtained. In this way, the number of the resulting binary images is equal to $2n_t$.

The SFTA feature vector is constructed by representing the resulting binary images through their size, mean gray level, and characterization of boundary fractal dimension through the box-counting method. In this application, we have used six binary images, each described by three abovementioned measures. In this way, 18 SFTA features have been defined.

The total size of image descriptors taken into account in the next steps of processing is equal to 81. The first seven represent the KS features, six represent the maximum subregion descriptors, two represent the percolation descriptors, the next 18 (from 16 to 34) are based on the SFTA approach, and the last 48 are the Haralick features.

3.3 Feature selection

A central problem in constructing the efficient classification system is identifying a representative set of features from which to construct a classification model for a particular task. Good feature should take similar values for the representatives of the same class and differ significantly for different classes. Thus, the main problem in the classification process is to find out the features of the highest importance for the problem solution. Elimination

of the features of the weakest class discrimination ability (treated as the noise) leads to smaller dimension of the feature space and improvement of the generalization ability of the classifier in the testing mode for the data not taking part in learning.

In our numerical experiments, we have implemented and compared three methods of feature selection: Fisher discriminant measure (FD) [25], correlation feature selection (CFS) [26], and fast correlation-based filter (FCBF) [27].

In Fisher criterion, the importance of the feature f is measured on the basis of the so-called discrimination coefficient $S_{ab}(f)$. For two classes A and B, it is defined as follows

$$S_{AB}(f) = \frac{|c_A(f) - c_B(f)|}{\sigma_A(f) + \sigma_B(f)} \tag{4}$$

The parameters c_A and c_B are the mean values of the feature f in classes A and B, respectively. The variables σ_A and σ_B represent the standard deviations determined for both classes. The larger the value of $S_{AB}(f)$ the better is the separation ability of the feature f for these two classes. Figure 7 presents the actual values of the Fisher measure of all features generated for the set of images representing the investigated 176 lesions of the skin.

According to the Fisher method, the best discrimination ability possesses the maximum subregion descriptors and KS features. The poorest performance is associated with the Haralick features. By trying different values of threshold levels and checking the maximum classification accuracy on the learning set, we have found the optimum cutoff value of Fisher measure equal to 0.18 (the horizontal continuous line in the figure). It resulted in 21 important features, which are then treated as the elements of the input vector to the classification system. The representative features from all families have been chosen in the selected set: 6—KS, 7—maximum subregions, 1—percolation, 5—SFTA, and 2—Haralick.

The second approach to the feature selection is based on correlation and called correlation feature selection (CFS) [26]. We assume that a good feature set is one that contains features highly correlated with the predicted class, and not correlated with each other. Knowing the correlation between each potential feature of the process and the class, and also the inter-correlation between each pair of components, the correlation measure between a composite test consisting of the summed features and the class is estimated as [26]

$$R_{cf} = \frac{NR_{ci}}{\sqrt{N + N(N-1)R_{ii}}} \tag{5}$$

where R_{cf} is the estimated total correlation measure between the summed features and class c , N is the number

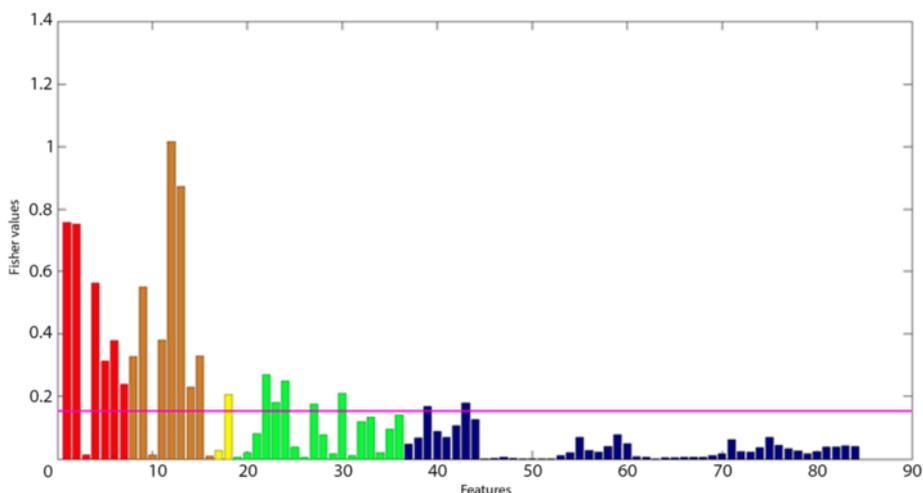


Fig. 7 The values of Fisher discrimination measure of the features. The features are arranged in the following way: the first 7, the KS features; the next 9, the maximum subregions descriptors; the next 2, the percolation descriptors; the next 18, the SFTA features; and the last 48, the Haralick features

of components, \bar{R}_{ci} is the average of Pearson’s correlations between the set of features and the class, and \bar{R}_{ii} is the average inter-correlation between features [26]. This equation is used as a heuristic measure of the “merit” of feature subsets in classification tasks. The set resulting in the highest value of this merit measure is treated as an optimal one.

Application of the CFS method has resulted in the set of 15 features, covering the members of all feature families. The following features have been selected: 1, 2, 4, 6, 8, 9, 11, 12, 13, 14, 24, 27, 31, 32, and 43. The first four belong to KS family, six of them are the maximum subregion descriptors, four represent the SFTA, and only one represents the Haralick feature.

The third selection method investigated in this work is the fast correlation-based filter (FCBF), exploiting the correlation measure based on the information-theoretical concept of entropy, defined for the variable x and for variable x after observing the variable y . The task is to select the features which are important to the class recognition but at the same time not redundant to any of the other relevant features.

The relevance of the feature x to class c is decided by calculating the symmetrical uncertainty $SU(x,c)$ measure between each feature and the class and also the values of $SU(x_i,x_j)$ for pairwise correlations [27]. By assuming proper threshold values for both measures, we eliminate the features below the threshold.

In practical application of this algorithm, we have assumed the threshold $SU(x,c)$ equal to 0.68 and $SU(x_i,x_j) = 0.50$. As a result, we have selected only six features treated as the most important for the class recognition. This set included one feature representing KS

family (feature 4), the next two representing the maximum subregion descriptors (features 9 and 12), two of the SFTA (features 24 and 31), and one of the Haralick family (feature 43).

3.4 Classification systems

The selected features are used as the input attributes to the classifier system. To get the best possible solution, we compared the performance of two classifiers: support vector machine (SVM) and random forest (RF), both having the opinion of the best. All of them have been implemented in Matlab [28].

The SVM [29, 30] is a linear machine, working in the high-dimensional feature space formed by the non-linear mapping of the N -dimensional input vector \mathbf{x} into an L -dimensional feature space ($L > N$) through the use of a kernel function $K(\mathbf{x},\mathbf{x}_i)$. The learning problem of SVM is formulated as the task of separating the learning vectors \mathbf{x}_i into two classes of the destination values either $d_i = 1$ (one class) or $d_i = -1$ (the opposite class), with the maximal separation margin. The SVM of the Gaussian kernel has been used in our application. The hyperparameters (the regularization constant C and Gaussian kernel width) have been adjusted by repeating the learning experiments for the set of their predefined values and choosing the best one for the validation data set.

The Breiman random forest (RF) is an ensemble of decision trees for classification [31]. It operates by constructing many decision trees at training time and outputting the class that is the mode of the classes output by individual trees. The generalization property is improved by applying randomness in selecting the learning data and using the limited set of decision

variables chosen randomly in each node of the tree. Random forest has the reputation of very high efficiency classification system.

4 Results of numerical experiments

The abovementioned classifier systems have been associated with different sets of selected features obtained in the first phase of processing and used in melanoma recognition. The system resulting in the best accuracy is treated as the final solution. The statistical results of accuracy related to our base of 175 images are depicted in Table 3. They refer to the testing data not taking part in learning in tenfold cross-validation mode and correspond to the application of two abovementioned classifiers associated with different sets of input attributes (without selection and selection made by FCBF, CFS, and Fisher). Different feature selection methods have resulted in changing average accuracy of the class recognition. SVM was found the best in three configurations of features (CFS, Fisher, and also for the set of all features without selection). RF was found the best in association with the FCBF selection method. The accuracy of melanoma recognition differs significantly, from the worst 78.63 % to the best of 93.80 % (Fisher selection and SVM).

In the case of medical application, the important is the confusion matrix of classification and the measures associated with it, defined in the form of sensitivity and specificity [29]. Sensitivity refers to the ability of the classifier to correctly detect a melanoma among all cases of melanoma. Specificity determines the ability of correctly excluding a melanoma. Given the preventive applications of these kinds of systems, it is more critical detecting correctly a melanoma than making fewer mistakes when determining that an image is not a melanoma. Table 4 presents the confusion matrix corresponding to the best result of classification.

The confusion matrix illustrates how the cases belonging to two classes (class of melanoma and the class of other skin lesions) have been classified by our system. The columns represent the actual outputs of our system and the rows—the targets. The number in each entry of the 2×2 matrix is the total number of the actually recognized classes in testing mode, calculated in five runs of cross-validation experiments. The diagonal entries of this matrix represent the quantity of the properly recognized cases. Each entry outside the diagonal means the

Table 3 Statistical results of accuracy of melanoma recognition in tenfold cross validation

Classifier	All features	FCBF	CFS	Fisher
SVM	85.47 %	78.62 %	89.51 %	93.76 %
RF	85.34 %	83.15 %	89.50 %	91.5 %

Table 4 The confusion matrix of melanoma recognition for our base of images

	Melanoma	Non-melanoma
Melanoma	80	4
Non-melanoma	7	85

The rows represent the target class and columns the output class

number of misclassifications. The entry in the (i,j) th position of the matrix for $i \neq j$ means false assignment of the case of i th class to the j th one.

The sensitivity is defined as the ratio of the true positive cases of melanoma to the sum of true positive and false negative cases. The specificity represents the ratio of the true negative cases (class of non-melanoma) to the sum of true negative and false positive cases.

The results show that the best classification system (SVM associated with Fisher selection) is able to recognize the melanoma from the other lesions of the skin with the total accuracy of 93.8 %. The sensitivity in recognition of melanoma is equal to 95.2 % and the specificity 92.4 %. The non-zero class recognition errors are due to the non-unique characteristics of the images in the data sets. The melanoma and other skin lesions images inherently vary greatly from patient to patient according to the type and advancement degree of lesions. Special difficulties in recognition follow from the changing colors of the skin lesions taking part in experiments. Among the processed lesions images, we have found brown, skin-colored, pink, red, purple, and even blue.

Figure 8 illustrates the receiver operating characteristic (ROC) curve of the best classification system [32]. The quality of the classifier is assessed on the basis of the area under curve (AUC). The closer this area to value of one the better is the classifier. The AUC of our classifier system is equal to 0.923, and this value is an evidence of high quality of solution.

Additional experiments have been performed using the publically available PH2 database [23]. The testing results of tenfold cross validation at application of SVM are given in Table 5 in the form of the confusion matrix.

The sensitivity of the system for this database was equal to 95.0 % and specificity 88.1 %. The results are slightly better than these presented by authors in [16, 23] for the same database. The best reported measures in these works were given for different variants of solution: sensitivity 93 % and specificity 78 % (global method and texture features), sensitivity 90 % and specificity 89 % (global method and color features), sensitivity 96 % and specificity 80 % (global system splitting image into two subregions), or sensitivity 100 % and specificity 75 % (local system).

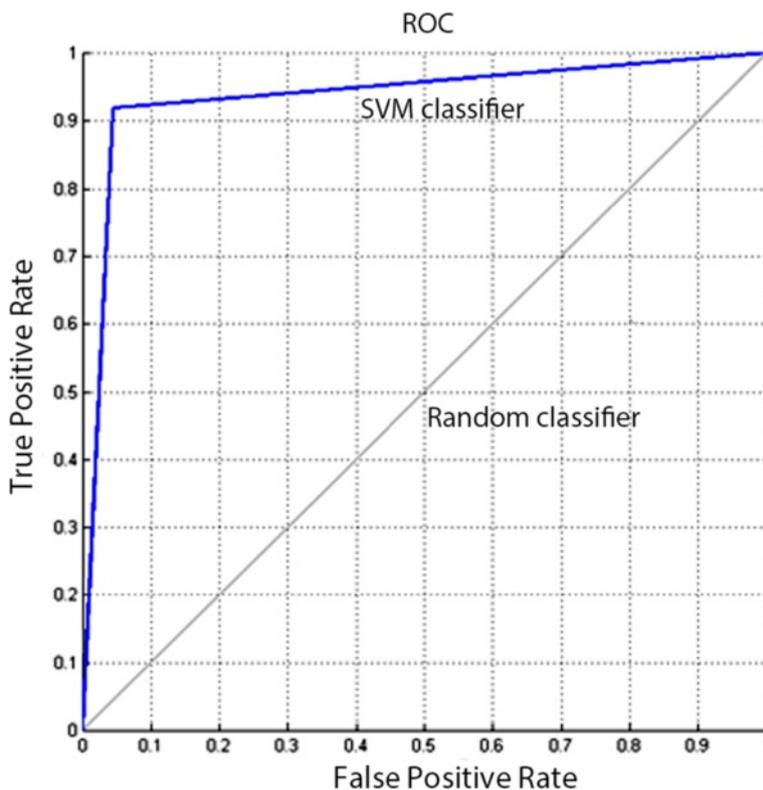


Fig. 8 The ROC curve of the best classification system (SVM associated with attributes selected by the Fisher method)

In general, it is impossible to get at the same time very high sensitivity and specificity. By changing the classification criteria, it is possible to change the balance level between these two quality measures. For example, at application of one-class SVM [30], we have got 100 % sensitivity; however, on the cost of specificity, it has dropped to 79.4 %.

5 Conclusions

The paper has presented the research directed to the automatic recognition of melanoma from the other lesions of the skin on the basis of color image of the nevus. The proposed approach uses extended set of diagnostic features describing the image of the skin lesions combined with different solutions of the classifiers. In our solution, we have resigned from the popular ABCD features, trying to find more powerful descriptors of the image, which are able to increase the accuracy of class recognition (melanoma versus non-melanoma lesions).

Table 5 The confusion matrix of melanoma recognition for PH2 database

	Melanoma	Non-melanoma
Melanoma	38	2
Non-melanoma	19	141

The rows represent the target class and columns the output class

The applied descriptors rely on the Kolmogorov–Smirnov statistics, maximum subregions statistics, percolation theory, fractal texture analysis, and Haralick texture descriptors. To reduce the number of input attributes applied in classification, we have tried three different selection methods of diagnostic features: standard Fisher discrimination measure, correlation feature selection, and fast correlation-based filter. Each of these methods applies different mechanism of selection, which results in various sets of attributes.

These sets have been confronted with three different classification systems. As the classifiers, we have tried support vector machine and random forest of decision trees. The best accuracy of class recognition on the database of Warsaw Memorial Cancer Center (Poland) has been achieved in the system formed by the SVM classifier supplied by the attributes selected using the Fisher method. The results of numerical experiments show that this classification system is able to recognize the melanoma from the other lesions of the skin with the total accuracy of 93.8 %. The sensitivity in recognition of melanoma is equal to 95.2 % and the specificity 92.4 %.

Additional experiments performed on the publically available PH2 database of Hospital Pedro Hispano (Matosinhos, Portugal) have also shown the superiority of this approach. In this case, the sensitivity in

recognition of melanoma was equal to 95.0 % and the specificity 88.1 %. They are of slightly higher quality than the results reported for this database in [23].

These experimental results obtained on these two data bases confirm that an automatic system applying extended set of image descriptors can reach the efficiency close to the dermatologist' expert results.

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

This work was supported by The National Centre for Research and Development of Poland under grant TANGO1/266877/NCBR/2015 which is being realized in the years 2015–2018.

Author details

¹Warsaw University of Life Sciences, Warsaw, Poland. ²Warsaw University of Technology, Warsaw, Poland. ³Military University of Technology, Warsaw, Poland. ⁴Central Clinical Hospital Ministry of Interior Affairs, Warsaw, Poland.

Received: 1 May 2015 Accepted: 23 November 2015

Published online: 14 December 2015

References

- N. Howlader, A.M. Noone, M. Krapcho, J. Garshell, N. Neyman, S.F. Altekruse, C.L. Kosary, M. Yu, J. Ruhl, Z. Tatalovich, H. Cho, A. Mariotto, D.R. Lewis, H.S. Chen, E.J. Feuer, K.A. Cronin, SEER Cancer Statistics Review, 1975-2010, National Cancer Institute. Bethesda, MD, http://seer.cancer.gov/archive/csr/1975_2010/, based on November 2012 SEER data submission, posted to the SEER web site, April 2013.
- SE Yagerman, A Marghoob, The ABCDs and beyond. *Skin Canc. Found. J.* **31**, 61–3 (2013)
- SE Yagerman, L Chen, N Jaimes, SW Dusza, AC Halpern, A Marghoob, 'Do UC the melanoma?' Recognizing the importance of different lesions displaying unevenness or having a history of change for early melanoma detection. *Australas. J. Dermatol.* **55**, 119–24 (2014)
- SM Goldsmith, A unifying approach to the clinical diagnosis of melanoma including "D" for "Dark" in the ABCDE criteria. *Dermatol. Pract. Concept.* **4**(4), 75–78 (2014)
- R Garnavi, M Aldeen, J Bailey, Computer-aided diagnosis of melanoma using border and wavelet-based texture analysis. *IEEE Trans. Inf. Technol. Biomed.* **16**(6), 1–13 (2012)
- H Ganster, A Pinz, E Wildling, M Binder, H Kittler, Automated melanoma recognition. *IEEE Trans. Med. Imag.* **20**(3), 233–239 (2001)
- C Grana, G Pellacani, R Cucchiara, S Seidenari, A new algorithm for border description of polarized light surface microscopic images of pigmented skin lesions. *IEEE Trans. Med. Imag.* **22**(8), 1235–1247 (2003)
- M Duarte, T Matthews, WS Warren, Calderbank, *Melanoma Classification from Hidden Markov Tree Features. Int. Conf. ICASSP*, 2012, pp. 6865–688
- AG Manousaki, AG Manios, EI Tsompanaki, AD Tosca, Use of color texture in determining the nature of melanocytic skin lesions—a qualitative and quantitative approach. *Comput. Biol. Med.* **36**, 419–427 (2006)
- M Silveira, JC Nascimento, JS Marques, AR Marcal, T Mendarca, S Yamauchi, J Maeda, J Rozeira, Comparison of segmentation methods for melanoma diagnosis in dermoscopy images. *IEEE J. Sel. Top. Sign. Proces.* **3**(1), 35–45 (2009)
- P Sabouri, HH Gholam, T Larsson, J Collins, *A Cascade Classifier for Diagnosis of Melanoma in Clinical Images. Engineering in Medicine and Biology Society (EMBC) 36th Annual Intern. Conf. of the IEEE, Chicago*, 2014
- M Celebi, HA Kingravi, B Uddin, H Iyatomi, Y Aslandogan, W Stoecker, R Moss, A methodological approach to the classification of dermoscopy images. *Comput. Med. Imaging Graph.* **32**(6), 362–373 (2007)
- E Zagrouba, W Barhoumi, A preliminary approach for the automated recognition for malignant melanoma. *Image Anal. Stereol.* **23**(2), 121–135 (2004)
- AG Isasi, BG Zapirain, AM Zorrilla, Melanomas non-invasive diagnosis application based on the ABCD rule and pattern recognition image processing algorithms. *Comput. Biol. Med.* **41**, 742–755 (2011)
- B Salah, M Alshraideh, R Beidas, F Hayajneh, Skin cancer recognition by using a neuro-fuzzy system. *Cancer Informat.* **10**, 1–11 (2011). doi:10.4137/CIN.S5950
- C Barata, M Ruela, M Francisco, T Mendonça, J Marques, Two systems for the detection of melanomas in dermoscopy images using texture and color features. *IEEE Syst. J.* **8**(3), 965–979 (2013)
- R Haralick, L Shapiro, Image segmentation techniques. *Computer Vision Graphics Image Process.* **29**, 100–132 (1985)
- F Costa, GE Humpire-Mamani, AJ Traina, An efficient algorithm for fractal analysis of textures, in *SIBGRAPI - XXV Conf. Graphics, Patterns and Images, Ouro Preto, Brazil*, 2012, pp. 39–46
- M Schroeder, *Fractals, Chaos, Power Laws* (W.H. Freeman and Company, New York, 2006)
- GW Corder, DI Foreman, *Nonparametric Statistics for Non-Statisticians: a Step-by-step Approach* (Wiley, New York, 2009)
- Swiderski, S Osowski, M Kruk, J Kurek, Texture characterization based on the Kolmogorov-Smirnov distance. *Expert Syst. Appl.* **42**(1), 503–509 (2015)
- D Stauffer, *Introduction to Percolation Theory* (Taylor & Francis, London, 1985)
- T Mendonca, P Ferreira, J Marques, AR Marcal, J Rozeira, *PH2—a Dermoscopic Image Database for Research and Benchmarking, 35th Int. Conf. IEEE Engineering in Medicine and Biology Society, Osaka*, 2013. <http://www.fc.up.pt/addi/ph2%20database.html>
- K Kimia, RS Ahmad, E-shaver: an improved DullRazor for digitally removing dark and light-colored hairs in dermoscopic images. *Comput. Biol. Med.* **41**, 139–145 (2011)
- RO Duda, PE Hart, P Stork, *Pattern Classification and Scene Analysis* (Wiley, New York, 2003)
- M Hall, *Correlation-Based Feature Selection for Discrete and Numeric Class Machine Learning. Proc. 17th Intern. Conf. Machine Learning Morgan Kaufmann Publishers, San Francisco*, 2000, pp. 359–366
- H Liu, L Yu, *Feature Selection for High-Dimensional Data: a Fast Correlation-Based Filter Solution. Proc. 20th Intern. Conf. Machine Learning (ICML-03), Washington, D.C.*, 2003, pp. 856–863
- MATLAB (R2012a) (2012) MATLAB user manual, Release 7.14.0. The Math Works, USA
- S Haykin, *Neural Networks, a Comprehensive Foundation* (Macmillan College Publishing Company, New York, 2000)
- B Schölkopf, A Smola, *Learning with Kernels* (MIT Press, Cambridge MA, 2002)
- L Breiman, Random forests. *Mach. Learn.* **45**(11), 5–32 (2001)
- PN Tan, M Steinbach, V Kumar, *Introduction to Data Mining* (Pearson Education Inc, Boston, 2006)

Submit your manuscript to a SpringerOpen® journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com