Computational Diagnosis of Skin Lesions from Dermoscopic Images using Combined Features

Roberta B. Oliveira[a], Aledir S. Pereira[b] and João Manuel R. S. Tavares[a,*]

[a] Instituto de Ciência e Inovação em Engenharia Mecânica e Engenharia Industrial, Departamento de Engenharia Mecânica, Faculdade de Engenharia, Universidade do Porto, Rua Dr. Roberto Frias, 4200-465 Porto, Portugal

[b] Departamento de Ciências de Computação e Estatística, Instituto de Biociências, Letras e Ciências Exatas, Universidade Estadual Paulista, Rua Cristóvão Colombo, 2265, 15054-000, São José do Rio Preto, SP, Brazil

Abstract

There has been an alarming increase in the number of skin cancer cases worldwide in recent years, which has raised interest in computational systems for automatic diagnosis to assist early diagnosis and prevention. Feature extraction to describe skin lesions is a challenging research area due to the difficulty in selecting meaningful features. The main objective of this work is the find the best combination of features, based on shape properties, colour variation and texture analysis, to be extracted using various feature extraction methods. Several colour spaces are used for the extraction of both colour- and texture-related features. Different categories of classifiers were adopted to evaluate the proposed feature extraction step and several feature selection algorithms were compared for the classification of skin lesions. The developed skin lesion computational diagnosis system was applied to a set of 1104 dermoscopic images using a cross-validation procedure. The best results were obtained by an optimum-path forest classifier with very promising results. The proposed system achieved an accuracy of 92.3%, sensitivity of 87.5% and specificity of 97.1% when the full set of features was used. Furthermore, it achieved an accuracy of 91.6%, sensitivity of 87% and specificity of 96.2%, when 50 features were selected using a correlation-based feature selection algorithm.

Keywords: Feature extraction and selection; Fractal dimension analysis; Discrete wavelet transform; Co-occurrence matrix.

1. Introduction

Dermoscopic images are widely applied for automated diagnosis of pigmented skin lesions. Such images can be acquired from dermatoscopes or specific cameras to provide a better visualization of the pigmentation pattern on the skin surface. Several computational systems have been proposed to assist dermatologists in obtaining an effective diagnosis [1-3]. These systems can be used to monitor benign skin lesions, and malignant lesions may be diagnosed at an early stage, so that the patient has a higher

---

* Corresponding author. Tel.: +351 220413472; Fax: +351 225081445 (João Manuel R. S. Tavares).
  Email addresses: roberta.oliveira@fe.up.pt (Roberta B. Oliveira), aledir@ibilce.unesp.br (Aledir S. Pereira), tavares@fe.up.pt (João Manuel R. S. Tavares).

probability of being cured with less aggressive therapies. The ABCD dermoscopy rule is usually taken into account for skin lesion diagnoses and when designing feature extraction methods; therefore, such diagnoses are based on the analysis of asymmetry, border, colour and differential structures, A, B, C and D, respectively. The asymmetry criterion can be defined by the asymmetry analysis of the skin lesion border, its colour or structures. The border criterion analyses the abrupt cut-off of the network at the lesion border, and the colour criterion identifies the presence of possible basic colours, such as white, red, light-brown, dark-brown, blue-grey and black. The differential structures criterion is characterized by the presence of pigment networks, vascularization, regression structures, streaks and dots/globules [4]; nevertheless, the identification of these structures are rarely used for automated diagnosis of skin lesions, mainly due to their complexity [5].

The features extracted from skin lesion images must represent their class, e.g., benign or malignant. Several methods to extract shape-, colour- and texture-related features for the automated diagnosis have been proposed in the literature [6-11]. Such features are based on the ABCD rule and they can characterize skin lesion properties adequately. Equivalent diameter, solidity, rectangularity, aspect ratio and eccentricity, are some examples of the shape features used, which represent both the A and B criteria of the ABCD rule. Statistical measures in several colour spaces are used to represent colour features based on this rule, and texture analysis methods, e.g. grey-level co-occurrence matrix are commonly used to represent the D criterion [12,7,5]. Nevertheless, few of the systems that have been proposed combine different methods to extract features in a similar category, e.g., texture analysis. Texture analysis methods are usually categorized as structural, statistical, model-based and transform. Although the structural approach provides a good symbolic description, some extracted features can be more useful for synthesis tasks rather than analysis tasks [13]. Among the various statistical methods that have been proposed, the co-occurrence matrix has shown potential for effective texture discrimination with dermoscopic images [14,5,15]. Fractal dimension is a model-based method that is also potentially useful for texture analysis in skin lesion images [16]. Fourier [17], Gabor [18] and wavelet [7] transforms have been also applied to extract texture features in skin lesion images.

The assessment of classifiers is an important issue for pattern recognition processes [19,20]. The most commonly used classifiers in skin lesion pattern recognition [24] include the nearest neighbours [12,21], Bayes networks [7,5], decision tree [17,7], artificial neural network [2,22] and support vector machine [7,6]. Other difficulties for pattern recognition processes involve defining which features are meaningful to describe the skin lesions, including the presence of highly correlated, redundant and irrelevant features. Some studies have proposed feature selection methods [23] to overcome these difficulties, such as feature selections based on correlation, gain information and relief-F [7,6]. An overview of the computational methods for pigmented skin lesion classification in images, which addresses the feature extraction and selection, and the classification steps, is presented in Oliveira et al. [24].

The aim of the present study was to evaluate and propose the most relevant features for skin lesion computational diagnosis based on the ABCD rule, including shape properties, colour variation and texture analysis using several different methods. The main contributions of this study were expected to be the texture analysis based on four colour spaces and the combination of different texture extraction methods, since texture features are usually extracted from grey-level images or from a few colour channels, and using only one texture extraction method [25,7]. In addition, good classification results were also expected when these features were combined with shape and colour features.

This article is organized as follows: The proposed feature extraction system, based on shape, colour and texture properties, is explained in Section 2. The algorithms used for selecting features and classifying skin lesions in dermoscopic images are detailed in Section 3. The experimental results are presented in Section 4. A discussion about the results obtained with the skin lesion classification is presented in Section 5. Finally, the conclusions drawn and proposals for future studies are presented in Section 6.

2. Proposed feature extraction

In this section, a combination of features to represent the skin lesion images is proposed. These features are based on the ABCD rule of dermoscopy, which is commonly used by dermatologists when diagnosing skin lesions. Various approaches have been proposed in the literature for skin lesion diagnosis in dermoscopic images [24]. Here, the feature extraction step is based on the intensities of the pixels in the regions of interest (ROIs) defined by specialists, i.e., binary masks, where the non-zero pixels belong to the lesion, and the others to the background skin. The binary masks were used in order to obtain trustworthy classification results and conclusions. Fig. 1 provides an overview of the approach proposed in this study. The features were categorized into shape properties, colour variation and texture analysis as described in Table 1. The extracted features were combined in a pool in the following sequence: shape, colour and texture. A dataset was built from this pool of features with a number of samples (x_i), according to the number of images n for a given classification problem, $i = 1,2,...,n$. Each sample (x_i) was composed of m features (x_{im}) and the class to which it belongs (y_i). Such a dataset was used in the image classification process of benign or malignant lesions using different classifiers and feature selection algorithms to evaluate the proposed approach.
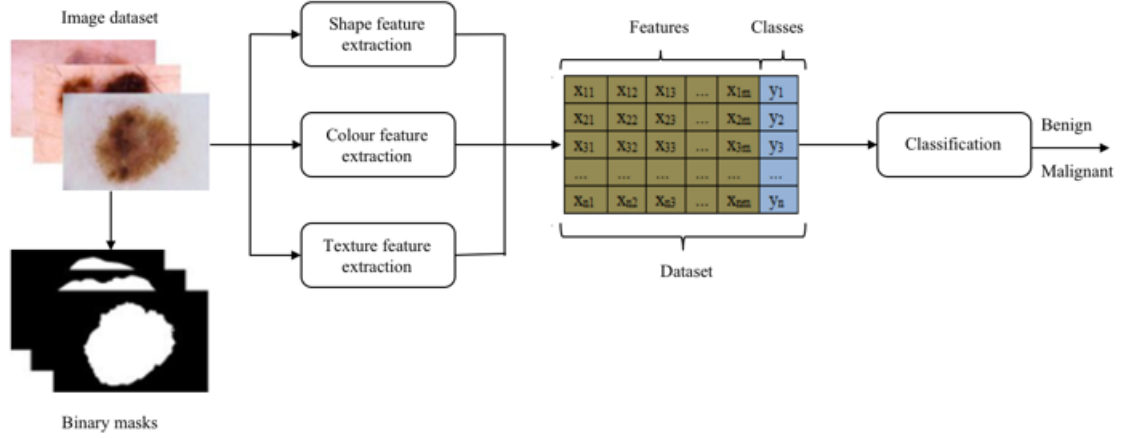
Fig. 1 Overview of the proposed approach for the skin lesion computational diagnosis.

Table 1 Features extracted from skin lesion images based on shape properties, colour variation and texture analysis.

Skin lesion features	Denotation	Number of features (channels $c = 1,2,...,12$)
Shape properties		
Geometrical properties	$A, P, ED, CO, CI, S, R, AR, e$	9
Lesion asymmetry	μ_s, s_s^2, s_s	3
Border irregularity	$p_S, v_S, l_S, p_L, v_L, l_L$	6
Colour variation		
Colour average, variance and standard deviation	μ_c, s_c^2, s_c	3 ($\times$ 12)
Minimum and maximum colours	min_c, max_c	2 ($\times$ 12)
Colour skewness	SK_c	1 ($\times$ 12)
Texture analysis		
Fractal dimension analysis	D_c^2	1 ($\times$ 12)
Discrete wavelet transform	$E(Sb)_c, H(Sb)_c; Sb = 1,2,...,10$	20 ($\times$ 12)
Co-occurrence matrix	$ASM_c, C_c, CRL_c, VAR_c, IDM_c, SA_c, SV_c,$ $SH_c, H_c, DV_c, DH_c, CRL1_c, CRL2_c, MCC_c$	14 ($\times$ 12)

2.1. Shape properties

Shape properties provide measures of the lesions based on their geometrical properties, their asymmetry or irregularity of their borders. These features are important for skin lesion diagnosis, as an asymmetric shape, border irregularity or ill-defined structure can characterize malignant lesions. Other geometrical properties of the lesion area which are commonly computed include the number of pixels inside the lesion region, aspect ratio, compactness, perimeter, greatest and shortest diameters, equivalent diameter, eccentricity, solidity, rectangularity, and circularity [7,14,6].The lesion asymmetry can be evaluated by dividing the region of the lesion under analysis into two sub-regions using an axis of symmetry, and thereby analyse the similarity of the area by overlapping the two sub-regions of the lesion along the axis. In some studies, the axis of symmetry is based both major and minor axes [7,6]. Features extracted from a wavelet transform [27,7], Fourier transform [28], fractal dimension [29], and irregularity index [7] have also been used to assess border irregularity. More details about shape classification and analysis can be found in [26]. In this study, 18 shape features of lesion were extracted

fron   of the standard features previously

men   ].

2.1.

T   ion by comparing the shape of the

lesi   r, some of these features depend on

the   re different as they may have been

acqu   ent resolutions. Consequently, a

norn   following sections.

a)   er of pixels within the lesion border,

                                     ion border.

b)   asures are based on a circle. The

                                     same as the lesion area A, which is

                                     io of the lesion area to a circle with

                                     n the perimeter can be calculated as

                                     ameter MD of the lesion [6], $CO =$

                                     ie lesion area approaches that of a

c)   convex hull (it checks a curve for

                                     from the lesion area. The solidity S

                                     CH, $S = A/CH$. Rectangularity R

                                     l, i.e., a bounding-box, $R = A/BA$,

d)   the structure of moments, up to the

                                     ied by the ratio of the length of the

                                     where A_1 and A_2 are given by:

$$|^2 \quad mu_{11}]^{1/2}\}\}^{1/2}, \tag{1}$$

or   of central moments of the lesion shape. The

oi   ven by: $c_x = m_{10}/m_{00}$ and $c_y = m_{01}/m_{00}$,

m_{ij}, given by Eq. (3).

$$\cdot (\tag{2}$$

$$m_{ij} = \qquad i \cdot y \tag{3}$$

The ec   a m   of   igation of the lesion region, which can be computed

as:

$$e = [(mu_{02} - mu_{20})^2 4mu_{11}]/(mu_{02} + mu_{20})^2, \tag{4}$$

where mu_{ij} is the central moments defined in Eq. (2).

2.1.2. Lesion asymmetry

In order to extract features based on the asymmetry properties, adapted from Oliveira et al. [16], the region of the lesion under analysis is divided into two sub-regions (R_1, R_2) by an axis, according to the longest diagonal, d, defined by the Euclidian distance: $D_{(p,q)} = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$, where (x_1, y_1) and (x_2, y_2) are the coordinates of the border pixels p and q. All the border pixels are analysed in order to find which pair has the greatest distance $D_{(p,q)}$. Perpendicular lines S_i from the pixels of the longest diagonal d are computed to analyse the similarity between two sub-regions of the lesion. Afterwards, two semi-lines are determined from each perpendicular line of the set S_i, one semi-line represents the sub-region R_1, and the other represents the sub-region R_2.

The distance $D_{(p,q)}$ of the semi-line for both sub-regions (R_1, R_2) is computed for each perpendicular, where p is a pixel of the diagonal d and q is a pixel of the border. The ratio between the shortest and longest distances based on the semi-lines (R_1, R_2) from each perpendicular line of set S_i are computed. The ratio between the two semi-lines can determine whether the lesion area is more symmetric or more asymmetric to a particular pixel of the longest diagonal. Three features are extracted to represent the lesion asymmetry: average μ_s, variance s_s^2 and the standard deviation s_s from the ratios between the two semi-lines based on all perpendicular lines of set S_i.

2.1.3. Border irregularity

The border is represented by pixels that make up the lesion boundary. A one-dimensional border of the lesion under analysis is defined to extract features based on this property. The number of peaks, valleys and straight lines of the border is computed using the vector product and inflexion point descriptors from the one-dimensional border, according to Oliveira et al. [16]. The inflexion point descriptor aims to analyse border pixels P_i to define which pixels show a change of direction. On the other hand, the vector product descriptor aims to analyse the border pixels to identify peaks and valleys with substantial irregularities. Six features are extracted to represent border irregularities: 1) the number of peaks p_S, valleys v_S and straight lines l_S based on small irregularities of the border using the inflexion point descriptor; and 2) the number of peaks p_L, valleys v_L and straight lines l_L based on large irregularities of the border using the vector product descriptor.

2.2. Colour spaces

Several colour spaces, described in the literature, are used to obtain more specific information about the colours of a lesion [24]. Some studies were focused on using only RGB images, and most of them only used the red channel as it is suitable to characterise skin lesions due to the dark colour of malignant lesions and the reddish colour of benign lesions [30]. Other studies used the RGB space combined with other colour spaces to describe the colours of skin lesions, such as the *HSV*, *CIE Lab* and *CIE Luv* spaces

that represent colours based on human perception [14,12,6,5]. Furthermore, *CIE Lab* and *CIE Luv* spaces are approximately perceptually uniform colour spaces which can facilitate the human perception of the colour properties [31]. Here, for the extraction of colour and texture features, four colour spaces were used: *RGB*, *HSV*, *CIE Lab* and *CIE Luv*, which correspond to the defined sequence of the channels $c = 1, 2, ..., n$, where n is the number of channels ($n = 12$), in order to explore the potential of each of them as already mentioned.

a) *RGB* colour space: This colour space represents the numerical values of the red, green and blue channels, and is widely used, since the images are originally obtained with this colour space. Moreover, the original *RGB* colour image can be used for conversion to other colour spaces. Although this colour space presents some disadvantages such as high correlation between the channels and no perceptual uniformity [32], several studies have achieved good results from it [6,14].

b) *HSV* colour space: This colour space represents the hue, saturation and value channels, which define the perceived colour of an area, the purity of colour and the brightness of colour, respectively. The conversion from the *RGB* colour space to the *HSV* colour spaces is given by:

$$V = \max(R, G, B),$$

$$S = \begin{cases} [V - \min(R, G, B)]/V, & if\ V \neq 0 \\ 0, & if\ V = 0 \end{cases},$$

$$H = \begin{cases} 60(G - B)/[V - \min(R, G, B)], & if\ V = R \\ 120 + 60(B - R)/[V - \min(R, G, B)], & if\ V = G \\ 240 + 60(R - G)/[V - \min(R, G, B)], & if\ V = B \end{cases}.$$

$$H = H + 360,\ if\ H < 0, \tag{5}$$

where $0 \leq H \leq 360$, $0 \leq S \leq 1$ and $0 \leq V \leq 1$, and the separation of each channel corresponds to $H = H/2$, $S = 255S$ and $V = 255V$.

c) *CIE Lab* and *CIE Luv* colour spaces: These colour spaces were proposed by the International Commission on Illumination (*CIE*, in French), whose main goal was to provide a uniform colour space. This means that the distance between two colours in such a colour space is strongly correlated with the human visual perception. Another advantage of these colour spaces is the separation of the luminance component L from the chrominance channels (a, b) and (u, v). A difference between these two colour spaces is that the *CIE Lab* colour space normalizes the values by division with the white colour point of the *CIE XYZ* colour space, whereas the *CIE Luv* colour space normalizes the values by the subtraction of such a white colour point [32,31]. The conversion from *RGB* colour space to the *CIE Lab* and *CIE Luv* colour spaces is based on the *CIE XYZ* colour space. Considering the values X_n, Y_n, and Z_n as being the white colour points, the *CIE Lab* colour space is computed by the following equations:

$$L = \begin{cases} 116(Y/Y_n)^{1/3} - 16, & for\ Y > 0.008856 \\ 903.3\ Y/Y_n, & for\ Y \leq 0.008856 \end{cases},$$

$$a = 500[(X/X_n)^{1/3} - (Y/Y_n)^{1/3}],$$

$$b = 200[(Y/Y_n)^{1/3} - (Z/Z_n)^{1/3}], \tag{6}$$

where $0 \le L \le 100$, $-127 \le a \le 127$ and $-127 \le b \le 127$, and the separation of each channel corresponds to $L = L * 255/100$, $a = a + 128$ and $b = b + 128$. And finally the *CIE Luv* colour space is computed by the following equations:

$$L = \begin{cases} 116(Y/Y_n)^{1/3} - 16, & for\ Y > 0.008856 \\ 903.3\ Y/Y_n, & for\ Y \le 0.008856 \end{cases},$$

$$u = 13L(u' - u_n), v = 13L(v' - v_n),$$

$$u' = 4X/X + 15Y + 3Z, v' = 9Y/X + 15Y + 3Z,$$

$$u_n = 4X_n/X_n + 15Y_n + 3Z_n, v_n = 9Y_n/X_n + 15Y_n + 3Z_n, \tag{7}$$

where $0 \le L \le 100$, $-134 \le u \le 220$ and $-140 \le v \le 122$, and the separation of each channel corresponds to $L = L * 255/100$, $u = 255/354(u + 134)$ and $v = 255/262(v + 140)$.

2.3. Colour variation

Statistical measures based on several colour spaces are commonly applied to the feature extraction from the lesion region [14,5,6]. Furthermore, these measures are also applied to other regions associated with the lesion border. The background skin [14] and surrounding skin (inner or outer peripheral regions) [6] are examples of such regions that are considered for feature extraction. Skin lesion features based on relative colours have been proposed [14,6] in order to assess colour features from the different regions associated with the lesion. Basic colours in the skin lesions have also been considered and computed [33].

In order to analyse the colour variation, six statistical measures are computed for each colour channel c of the lesion region using the four-colour spaces as defined earlier, with $c = 1,2, \dots, n$, where n is the number of channels used for the colour feature extraction.

a) Colour average, variance and standard deviation: These measures evaluate the average and the variation of a set of lesion intensity values I_p, of each colour channel c. The average μ_c, variance s_c^2, and standard deviation s_c are computed by the following equations:

$$\mu_c = \frac{1}{N}\sum_{p=1}^{N}(I_p), \tag{8}$$

$$s_c^2 = \frac{1}{N-1}\sum_{p=1}^{N}(I_p - \mu_c)^2, \tag{9}$$

$$s_c = \sqrt{s_c^2}, \tag{10}$$

where N is the number of pixels of the ROI in the image.

b) Minimum and maximum colours: These measures define the minimum value, $min_c = min(I_p)$, and the maximum value, $max_c = max(I_p)$ of the set of lesion intensity values I_p, of each colour channel c.

c) Colour skewness: This measure computes the asymmetry SK_c of the data around the set of lesion intensity values I_p:

$$SK_c = \left[\frac{1}{N}\sum_{p=1}^{N}\left(I_p - \mu_c\right)^3\right]/s_c^3, \tag{11}$$

where μ, s are the average and the standard deviation of the set of lesion intensity values I_p, and N is the number of pixels of the ROI in the image.

2.4. Texture analysis

The best features to represent the skin lesion texture were acquired by using three texture analysis methods. The texture features are computed for each colour channel using the four-colour spaces as defined earlier. Thus, a total of 420 texture features are extracted: 12 features from the fractal dimension analysis [34], 240 features from the discrete wavelet transform [35] and 168 features from the single-channel co-occurrence matrix [36].

2.4.1. Colour image-based fractal dimensional analysis

In order to extract the texture properties of the skin lesions, fractal dimensions are computed from the image under study using a box-counting method (BCM), since it is simple and effective for skin lesion analysis [16]. A fractal dimension [34] is a procedure for splitting the input image into several quadrants to quantify the irregularity level or self-similarity of the image fractals, according to $D = \log(P)/\log(1/T)$, where P represents the number of elements of the self-similar parts that reconstruct the original image, and T is the number of quadrants corresponding to a fraction of its previous size. BCM projects a grid over the image; i.e., it divides the image into several squares. The process is iterative, in which the size of each square decreases and the number of squares that covered the fractal is counted at each iteration.

The bi-dimensional fractal dimension D_c^2, which is computed individually for each channel c of the colour spaces, is defined as:

$$D_c^2 = \frac{1}{N}\left(\sum_{i=1}^{rows}\sum_{j=1}^{cols} D_{i,j}\right) + 1, \text{ with } c = 1,2,\dots,n, \tag{12}$$

where $D_{i,j}$ is the fractal dimension obtained at each iteration, i.e., it is computed individually for each row i and column j of the image, N is the total number of fractal dimensions, and n is the number of channels used for the texture feature extraction.

2.4.2. Colour image-based wavelet transform

There are several transform methods that have been applied to diagnose skin lesions based on texture feature analysis, including the Fourier [17], Gabor [18] and wavelet [7] transforms. Texture analysis methods based on the Fourier transform may present poor performance due to its lack of spatial localization, whereas a Gabor filter allows a superior spatial localization. However, the wavelet transform presents several advantages compared to the Gabor transform; for example, the variation of

the spatial resolution allows it to represent textures using a more suitable scale. There are several scales available to the wavelet function and therefore it can chose the best one for a given application [13]. In this work, a discrete wavelet transform (DWT) was adopted to extract texture features from images, since it provides a representation that is easy to interpret [35], and that can be efficiently implemented with a pyramidal structure using quadrature mirror filters for texture discrimination [37].

A bi-dimensional wavelet transform is used to decompose a 2-D image, to which one-dimensional transformations are applied individually along the horizontal and vertical directions of an image [35]. The decomposition of a one-dimensional signal $f(t)$ is based on a family of wavelet functions that usually is defined as complete and with an orthogonal base:

$$W_{a,b} = \int_{-\infty}^{\infty} f(t)\psi_{a,b}(t)dt. \tag{13}$$

This family is obtained by dilating and translating a single function defined as the mother wavelet ψ:

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right), \tag{14}$$

where a and b are the parameters of dilating and translating, respectively. When a and b are defined for discrete signals, a DWT is obtained.

The DWT, based on a multi-resolution, decomposes an input signal in two new signals with different frequencies using quadrature mirror filters. Such signals correspond to low- and high-pass filters that represent the wavelet functions (mother wavelet) $\psi(t)$ and scaling functions (father wavelet) $\phi(t)$, respectively. The low-pass filter corresponds to approximation coefficients, whereas the high-pass filter corresponds to detail coefficients.

The decomposition of a bi-dimensional signal using DWT yields a sub-sample with four sub-bands for one-level of decomposition that are: LL, LH, HL, HH. The sub-band LL corresponds to the clustering of low-pass filtering in the lines and columns. The sub-band LH corresponds to the clustering of low-pass filtering in the lines and high-pass filtering in the columns. The sub-band HL corresponds to the clustering of high-pass filtering in the lines and low-pass filtering in the columns. The sub-band HH corresponds to the clustering of high-pass filtering in the lines and columns. These sub-bands have an equal number of pixels as the original image. A multi-level decomposition can be considered, when the decomposition is applied recursively to the LL sub-band. The result of such decomposition is a standard pyramidal wavelet transform.

A problem in this wavelet decomposition approach is the large number of features that can be obtained depending on the number of levels used and it can give the classification a high computational cost. In addition, the resolution of the images decreases at each level decomposition and smaller details can gradually disappear [37]. Therefore, a-three-level decomposition was used to decompose the images

based on experiments performed by Mallat [37] who illustrated the numerical stability of this level for the decomposition and reconstruction processes with good quality. Based on this, the number of sub-bands ns was defined as 10 for each channel of the colour spaces. A Haar wavelet filter was used to implement the DWT, with the coefficients defined as $h = (1.0/\sqrt{2}, 1.0/\sqrt{2})$. This filter was used since it is simple and has been previously applied to extract texture from skin lesion images [38].

The energy $E(Sb)_c$ and entropy $H(Sb)_c$ measures for the feature extraction from the coefficients obtained by DWT are computed for each sub-band $Sb = 1,2, \dots, ns$ and each colour channel c:

$$E(Sb)_c = \sqrt{\frac{1}{N}\sum_{i=1}^{rows}\sum_{j=1}^{cols}(Sb_{i,j}^2)}, \tag{15}$$

$$H(Sb)_c = \frac{1}{N}\sum_{i=1}^{rows}\sum_{j=1}^{cols}[Sb_{i,j}^2 \times \log(Sb_{i,j}^2)], \tag{16}$$

where $Sb_{i,j}$ corresponds to the sub-band coefficient for the pixel i,j and N is the total number of pixels in the sub-band. These measures are commonly used to represent the texture of skin lesion images [7].

2.4.3. Colour image-based co-occurrence matrices

The grey-level co-occurrence matrices (GLCMs) represent the relationship between the intensities of neighbouring pixels to characterize the texture of an image [36]. Such a matrix $m(i,j,d,\theta)$ is obtained by the joint probability of occurrence of grey-levels considering each pair of neighbour pixels i,j of an image, where these pixels are separated by a distance d and in a specific direction θ.

In this study, co-occurrence matrices (CMs) were used for the colour channels. The single-channel co-occurrence matrices (SCMs) were applied separately to each colour channel, with $c = 1,2, \dots, n$, where n is the number of colour channels. The parameters used to set up the matrices are based on Haralick et al. [36]. The intensities of each channel are quantized by an equal probability quantizing algorithm, with $q = 16$. The distance d between one pixel and its neighbours is $d = 1$, and four orientations θ are considered $\theta = (0°, 45°, 90°, 135°)$. In order to extract rotation invariant features, a normalized SCM is obtained from the SCMs corresponding to the four orientations.

From the normalized SCM, 14 statistical measures based on Haralick's texture features [36] were extracted from the image: angular second moment ASM_c, contrast C_c, correlation CRL_c, variance VAR_c, inverse difference moment IDM_c, sum average SA_c, sum variance SV_c, sum entropy SH_c, entropy H_c, difference variance DV_c, difference entropy DH_c, information measure of correlation 1 $CRL1_c$, information measure of correlation 2 $CRL2_c$, and maximal correlation coefficient MCC_c. These features are expressed in Eq. (17)-(30), where $m_{i,j}$ is the entry value in the position i,j of the normalized matrix, and N is the number of different intensities contained in the quantized image:

$$ASM_c = \sum_{i=1}^{N}\sum_{j=1}^{N}(m_{i,j})^2, \tag{17}$$

$$C_c = \sum_{i=1}^{N} \sum_{j=1}^{N} [m_{i,j}(i-j)^2], \tag{18}$$

$$CRL_c = [\sum_{i=1}^{N} \sum_{j=1}^{N} (i \times j \times m_{i,j}) - \mu_x \mu_y]/\sigma_x \sigma_y, \tag{19}$$

where μ_x, μ_y, σ_x and σ_y are the averages and standard deviations of $m_x = \sum_{j=1}^{N}(m_{i,j})$ and $m_y = \sum_{i=1}^{N}(m_{i,j})$; and:

$$VAR_c = \sum_{i=1}^{N} \sum_{j=1}^{N} [(i-\mu)^2 m_{i,j}], \tag{20}$$

$$IDM_c = \sum_{i=1}^{N} \sum_{j=1}^{N} [m_{i,j}/1 + (i-j)^2], \tag{21}$$

$$SA_c = \sum_{i=2}^{2N} (i \times m_{x+y(i)}), \tag{22}$$

$$SV_c = \sum_{i=2}^{2N} [(i - SE_{ch})^2 m_{x+y(i)}], \tag{23}$$

$$SH_c = -\sum_{i=2}^{2N} [m_{x+y(i)} \log(m_{x+y(i)})], \tag{24}$$

$$H_c = \sum_{i=1}^{N} \sum_{j=1}^{N} [m_{i,j} \log(m_{i,j})], \tag{25}$$

$$DV_c = variance(m_{x-y}), \tag{26}$$

$$DH_c = -\sum_{i=0}^{N-1} [m_{x-y(i)} \log(m_{x-y(i)})], \tag{27}$$

where $m_{x+y(k)} = \sum_{i=1}^{N} \sum_{j=1}^{N}(m_{i,j})$, with $k = 2,3,\dots,2N$, $i + j = k$, and $m_{x-y(k)} = \sum_{i=1}^{N} \sum_{j=1}^{N}(m_{i,j})$, with $k = 0,1,\dots,N-1$, $|i-j| = k$; with:

$$CRL1_c = (HXY - HXY1)/max(HX, HY), \tag{28}$$

$$CRL2_c = (1 - exp[-2.0(HXY2 - HXY)])^{1/2}, \tag{29}$$

where HX and HY are entropies of $m_{x(i)}$ and $m_{y(j)}$, $HXY = -\sum_{i=1}^{N} \sum_{j=1}^{N} [m_{i,j} \log(m_{i,j})]$, $HXY1 = -\sum_{i=1}^{N} \sum_{j=1}^{N} [m_{i,j} \log(m_{x(i)} m_{y(j)})]$, and $HXY2 = -\sum_{i=1}^{N} \sum_{j=1}^{N} [m_{x(i)} m_{y(j)} \log(m_{x(i)} m_{y(j)})]$, and:

$$MCC_c = (second\ largest\ eigen\ value\ of\ Q)^{1/2}, \tag{30}$$

where $Q_{i,j} = \sum_{k}^{N} [(m_{i,k} m_{j,k})/(m_{x(i)} m_{y(k)})]$.

3. Skin lesion classification

Here, first the set of features for skin lesion diagnosis are constructed, and then classified. The classification process must be accurate, since it is used to assist dermatologists in their diagnosis; however, the accuracy of the classification depends on several factors, such as a reliable dataset. The pre-processing step in this study included data normalization, dataset balancing and feature selection. The classification was carried out using the Weka library [39].

3.1. Data pre-processing

The data pre-processing step, which precedes the classification process, normalizes the dataset values from the feature extraction process as they contain different ranges, and some classifiers cannot handle such differences. The normalization procedure scales all numeric values in the dataset to within the same interval [0,1] by computing:

$$xn_{im} = [x_{im} - \min{(x_{im})}]/[\max(x_{im}) - \min{(x_{im})}], \tag{31}$$

where x_{im} is the actual value of the feature m in the sample i, with the minimum and maximum values of features of all the samples, and xn_{im} is the normalized value of the same feature m in the same sample i.

Unbalanced datasets can affect the performance of classifiers. For example, here the dataset was composed of 916 samples of benign lesions and 188 samples of malignant lesions. This unbalanced dataset, i.e., with different numbers of samples in each class, can decrease the accuracy of the evaluation result, since classifiers tend to prioritize classes with a higher number of samples. Sampling methods have effective strategies to overcome such a problem and are commonly used [40]. In this work a combined resampling strategy was applied to the dataset [39], considering the random under-sampling and random over-sampling methods that are the two basic methods used for balancing classes. The random under-sampling removes samples randomly in the majority class, i.e., samples of benign lesions, while the random over-sampling replicates samples randomly in the minority class, i.e., samples of malignant lesions. This strategy produced a random subsample of the original dataset using sampling with replacement, where the samples are replicated or removed in the minority or majority class until a uniform distribution of the samples is reached. This strategy was adopted because it ensured a uniform distribution of the samples without removing to many samples from the majority class and without replacing to many samples in the minority class. This process established 552 samples of benign lesions and 552 samples of malignant lesions.

Another problem that also affects the performance of classifiers is the choice of meaningful features to represent the input images. Therefore, feature selection algorithms are used to define the best features to solve such a problem [41]. Feature selection consists of finding the best features through an evaluation process according to either ranking or search strategies. The ranking strategy produces a ranked list of features based on the evaluation process. On the other hand, the search strategy influences the search direction and execution time of the selection process depending on the search strategy adopted, which can be complete, sequential, or random [42]. The sequential search strategy is usually used for skin lesion feature selection and it can be by the forward, backward or bi-direction selection depending on the search method used. The forward selection process starts with an empty set, and the best features are gradually added to the set, according to the performance obtained from the evaluation method, whereas the backward selection process starts with all features and the worst features are removed at each iteration. The bi-direction selection combines both the forward and backward searches.

The evaluation process using filters allows for assessing the quality of selected features without using any classification algorithms. Each candidate subset is evaluated by applying an independent criterion, which can be based on several measures to compare it with the best current subset previously established. If the new evaluated subset is considered better then it becomes the best current subset. These measures can be defined as [43]:

- Distance measures that try to find the feature that can separate the classes as far as possible from each other;
- Information measures that establish the information gain from a feature and the feature with the most information is preferred; and
- Dependency measures that are also known as correlation measures applied to evaluate the ability to predict the value of one feature from the value of another, or how strongly a feature is in regard to the class.

In this study, six feature selection algorithms, based on the measures discussed above and on a feature transformation algorithm, were used to generate different subsets of features. These six algorithms are commonly used for the selection of skin lesion features [24], since they present several advantages over others, such as computationally efficient, simpler and faster algorithms, independent evaluation criteria, and ability to overcome over-fitting.

a) Relief-F feature selection [44]: This algorithm is an extension of the relief algorithm to deal with noise and multi-class problems. The dataset samples are randomly defined. For each sample that is defined, the closest samples of the same and different classes are selected using a nearest-neighbour algorithm [45]. The quality of each feature is estimated, according to its value in regard to these closest samples.

b) Information gain-based feature selection [41]: This algorithm estimates the quality of a feature, according to its information gain in regard to the class. The information gain between each feature F and the class C is measured by the entropy H, according to the information theory criteria [46]. Therefore, the features that have high information gain $Ig_{(C,F)}$ are considered the most relevant, where $Ig_{(C,F)} = H(C) - H(C|F)$.

c) Gain ratio-based feature selection (GRFS) [39]: This algorithm is also based on the entropy H and it estimates the quality of a feature F, according to its gain ratio in regard to the class C. Therefore, the features that have high gain ratio $Gr_{(C,F)}$ are considered the most relevant, where $Gr_{(C,F)} = [H(C) - H(C|F)]/H(F)$.

d) Correlation coefficient-based feature selection [41]: This algorithm estimates the quality of a feature, according to its Pearson's correlation coefficient in regard to the class. The correlation coefficient is computed by a covariance and variance between the features and the class.

e) Correlation-based feature selection (CFS) [47]: This algorithm tries to find a set of features that are highly correlated with a class and with low inter-correlation between them. The degree of correlation between the features is computed by a symmetrical uncertainty, which is a modified version of the information gain measure.

f) Principal-component analysis (PCA) [48]: Here the features are transformed to a PC based on a correlation matrix, where eigenvectors (vectors of features) are defined, according to some percentage of the variance in the original data. The worst eigenvectors are removed and the new features are ranked, according to the best eigenvalues.

All feature selection algorithms discussed above are single-feature evaluators, with the exception of CFS that is a feature subset evaluator. The single-feature evaluators are used with a ranking strategy, where the features are ranked individually according to their evaluation, i.e., the most relevant [39]. Here, in order to study different stopping criteria for the ranking strategy, the numbers of features to be retained (N) were empirically defined: 25, 50 and 75. On the other hand, the feature subset evaluator, i.e., CFS, measures the quality of a subset of features and returns a value that is used in the search [39]. In this study, the greedy stepwise and best first search methods were compared for use with the CFS algorithm. The greedy stepwise method searches for feature subsets in either the forward or backward directions in a greedy way [39]. The selection process using the greedy stepwise method and the CFS algorithm must stop when the addition or removal of any feature worsens the quality of the best-found subset, i.e., when the evaluation of the current subset presents a lower quality than the evaluation of the subset of the previous iteration. The best first method searches the feature subsets by greedy hill-climbing, and the search direction can be forward, backward or bi-direction [39]. The stopping criterion for the best first method and CFS algorithm was to stop after five successive iterations that did not improve the previous result.

3.2. Classification

In this study, the focus is on models with a single classifier that can choose the best classification using different datasets, e.g., using a stratified k-fold cross-validation procedure [39]. This approach splits the training set in k subsets of equal size and the procedure is repeated k times. In each procedure, one subset is employed as a test set while the others are used as the training set. The best model is chosen, according to its performance, which is measured by averaging the accuracy obtained from each trial. This procedure can be applied to avoid over-fitting while testing the capacity of the classifier to generalize. In addition, this approach has shown good results compared with other procedures [49].

Six different categories of classifier were applied in this work to evaluate the dataset from the extracted features: the k-nearest neighbours (KNN) [45], Bayes networks (Bayes Net) [50], C4.5 decision tree [51], multilayer perceptron (MLP) [52], and support vector machine (SVM) [53] were the most commonly used classifiers, according to the categories presented by Oliveira et al. [24]. In addition,

the optimum-path forest (OPF) classifier [22] was also used in this study. To the best our knowledge, no previous study has used this later classifier to identify skin lesions in images.

a) kNN: Here, a search algorithm and a distance function are used to assess which sample of the training set is closest to an unknown sample and then assigning the unknown sample to the class with the majority of k-nearest neighbours. The main advantages of these classifiers are their simplicity to implement and the possibility to add new samples to the training set at any time.

b) Bayes Net: This is a Bayesian learning-based algorithm [50] that computes the probability of a given set of features to belong to each class, assuming that the features are independent. The Bayes Net learning uses search algorithms and quality measures, which provide a network structure and conditional probability distributions.

c) C4.5: This algorithm is used to create a decision tree [54] that has a structure similar to a flowchart, in which each internal node (non-leaf) represents a test of a feature, each branch represents the result of the test, and each external node (leaf) indicates a prediction of the class. A complete decision tree can contain unnecessary structures, and strategies of pre-pruning and post-pruning can be performed to simplify its structure. Pre-pruning involves decision making during the tree building process, whereas in the post-pruning this is done afterwards. The C4.5 algorithm divides the features at the nodes based on information gain. It prevents over-fitting which is also a form of pre-pruning. The post-pruning in C4.5 yields a dense decision tree very quickly. It can also deal with situations in which two features that individually present no contribution, but are powerful predictors when combined [39].

d) MPL: This algorithm is one of the most commonly used architectures of artificial neural network (ANNs) [52] that are parallel distributed systems composed of layers of input and output elements linked by weighted connections. During the learning phase, the weights are adjusted to predict the correct class based on the input samples. The MPL can include one or more layers of processing, also called hidden layers, placed between the input and output layers. Back-propagation is a supervised learning method widely used in the MLP architecture, which consists of forward and backward processes applied to adjust the weight values of the connections. The MLP algorithm has good capability and flexibility to overcome various non-separable problems.

e) SVM: This classifier is used to build a hyper-plane to separate data, according to the defined classes. This kind of classifier has been commonly applied to classify skin lesions due to its good overall properties. Furthermore, kernel functions simplify the process of separating the non-linear data using a simple hyper-plane in a high dimension feature space. The radial basis function (RBF) and polynomial kernels have been frequently used in several different studies [24]. For the SVM classifier, Platt's [55] sequential minimal optimization algorithm was used.

f) OPF: This is applied to solving pattern recognition problems as a graph based on prototypes to represent each class by one or more optimum-path trees, considering some key samples. The training

16

samples are nodes of a complete graph; whose arcs are the link of all pairs of nodes. The arcs are weighted by the distances between the feature vectors of their corresponding nodes. The classification of a new sample is defined, according to the strong connectivity of the path between the sample and the prototype. Therefore, the path with minimum-cost, among all paths, is considered the optimum one. The OPF classifier shows some interesting properties, such as speed, simplicity, ability to deal with multi-class classification and overlapping between classes, parameter independence and no assumptions are based on the shape of the classes. For the application of the OPF classifier, it was used the Weka library based on LibOPF [22] as proposed by Amorim et al. [56].

The performance of the classification was evaluated using accuracy (ACC), sensitivity (SE) and specificity (SP) measures, which are based on outcomes of classifiers, according to the predicted class and known class. These outcomes represent the number of correct (true) and incorrect (false) classification for each class, positive and negative. These measures are commonly used according to [24] and they are defined as:

$$ACC = \frac{TP+TN}{P+N} \times 100\%, \tag{32}$$

$$SE = \frac{TP}{TP+FN} \times 100\%, \tag{33}$$

$$SP = \frac{TN}{TN+FP} \times 100\%, \tag{34}$$

where P is the number of positive samples and N is the number of negative samples of the dataset. Here, the positive samples represent the benign lesions and the negative samples the malignant lesions. Therefore, TP (true positive) is the number of correctly classified benign lesions, TN (true negative) is the number of correctly classified malignant lesions, FP (false positive) is the number of incorrectly classified malignant lesions, and FN (false negative) is the number of incorrectly classified benign lesions.

A cost function C adopted from Barata et al. [12] is used to deal with the trade-off between SE and SP, which is defined as:

$$C = \frac{c_{10}(1-SE)+c_{01}(1-SP)}{c_{10}+c_{01}}, \tag{35}$$

where c_{10} is the cost of an incorrectly classified benign lesion, and c_{01} is the cost of an incorrectly classified malignant lesion. The costs used to evaluate the classification were defined according to Barata et al. [12], where $c_{10} = 1$ and $c_{01} = 1.5$. These authors chose a higher cost for c_{01}, since an incorrect classification of a malignant lesion is more critical. The lower the value of cost C, the better the classification performance is.

4. Experimental results

In order to evaluate the proposed feature extraction in the classification of benign and malignant skin lesions, two experiments were performed. First, the experiments for the skin lesion classifications using all features of the dataset are presented. Second, the experiments for the feature selection of skin lesions are presented as well as these for the lesion classification. In this section, classification results are described and discussed. In addition, the image dataset used to evaluate the results is presented, as well as the computational time of the system.

4.1. Dermoscopic image dataset

The dermoscopic images of pigmented skin lesions used to evaluate the extraction of features were collected from the International Skin Imaging Collaboration (ISIC) dataset [57]. Examples of these images are shown in Fig. 2. In addition, the images are paired with the expert manual that contains the skin lesion diagnoses, as well as the ground truth lesion segmentations in the form of binary masks. In this study, a feature extraction approach, based on shape properties, colour variation and texture analysis, is proposed. Moreover, since the shape properties are obtained from the lesion borders, only the images where the lesion fitted completely within the image frame were selected so that the features could be extracted with greater precision. A total of 1,104 images were selected from the original dataset. Of these, 916 images were benign lesions and 188 images were malignant lesions. The images of the dataset were resized to an average resolution of 400×299 pixels to simplify their processing.
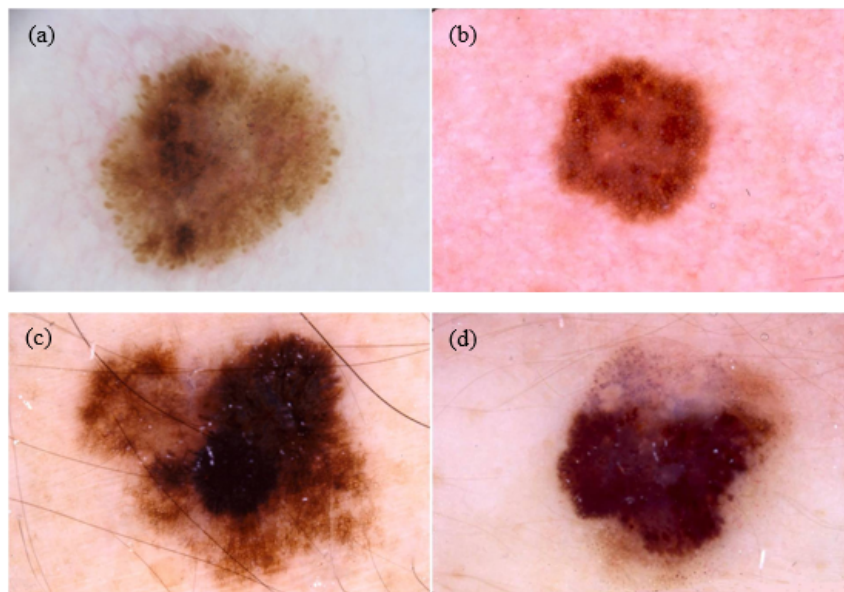

Fig. 2 Four examples of dermoscopic images: (a) and (b) are benign lesions, (c) and (b) are malignant lesions.

4.2. Evaluation of the proposed feature extraction

The performance of the classification using all extracted features was evaluated by different classifiers, which were described in the previous section. Each classifier was used with several different parameters to find the best results with a ten-fold cross-validation procedure. The set of parameters evaluated in this study was defined based on previous studies that had used these classifiers for skin

lesion classifications [12,5,58,59,21]. The kNN classifier used a linear nearest neighbour search algorithm and three distance functions were compared, i.e., Euclidean, Chebyshev and Manhattan, to find the nearest neighbours. Different values of k were applied for each distance function and the number of neighbours used was $k = \{5,7, \ldots ,25\}$. The Bayes net classifier used a hill-climbing search algorithm to find the network structures, and a simple estimator to estimate the conditional probabilities of a network. The parameter alpha for the simple estimator was settled with the following values: $A = \{0.1,0.2, \ldots ,0.9\}$. The C4.5 classifier used two sets to define the minimum number of samples per leaf, $M_1 = \{2,4, \ldots ,20\}$ and $M_2 = \{82,84, \ldots ,100\}$, and the values of the confidence factor used for pruning were $CF = \{0.1,0.2, \ldots ,0.9\}$.

The MPL classifier analysed two values: one hidden layer of the neural network, with $H_1 = (features + classes)/2$ and the other $H_2 = classes$. The learning rate $L = 0.3$ is the number of the weights that were updated, and the momentum $M = 0.2$ was applied to the weights when updating. The SVM classifier analysed two kernels: the polynomial and RBF kernels. In the RBF kernel, the parameter gamma was carried out with different values of $G = \{0.001,0.002, \ldots ,0.1\}$, and the complexity parameter $C = \{1,2, \ldots ,10\}$ was applied to both kernels. And finally the OPF classifier compared three distance functions: Euclidean, Chebyshev and Manhattan, in order to find the distances between the feature vectors.

As aforementioned, the best parameters for each classifier were defined based on initial experiments. Table 2 indicates the values of the parameters used in the following experiments performed in this study. Table 3 shows that good results were achieved using these parameters and the proposed extracted features, mainly for the specificity of the malignant lesion classification (SP).

Table 2 Best parameters achieved by each classifier.

Classifier	Parameters
k-nearest neighbours	k:5 search algorithm: Linear NN Search (distance function: Manhattan)
Bayes networks	estimator: Simple Estimator (alpha: 0.1) search algorithm: hill-climbing
C4.5 decision tree	confidence factor: 0.3 minimum number: 2
Multilayer perceptron	one hidden layer: $(features + classes)/2$ learning rate: 0.3 momentum: 0.2
Support vector machine	complexity parameter: 10 kernel: RBF (gamma: 0.1)
Optimum-path forest	distance function: Euclidean

Table 3 Performance results for each classifier using all features.

Classifier	ACC	SE	SP	C
kNN	75.8%	69.4%	82.2%	0.229
Bayes Net	68.2%	54.0%	82.4%	0.290
C4.5	86.9%	82.2%	91.5%	0.122

19

MLP	74.5%	61.2%	87.7%	0.229
SVM	**91.7%**	**87.1%**	**96.2%**	**0.074**
OPF	**92.3%**	**87.5%**	**97.1%**	**0.067**

4.3. Performance evaluation using feature selection

The best results were obtained by the OPF and SVM classifiers as shown in Table 3 (in bold), where both classifiers achieved a good generalization between the classes. Despite the fast training of the Bayes Net classifier, the classification results were not so expressive, as this classifier is sensitive to redundant features as it assumes that the features should be independent. The kNN classifier did not make a good distinction between the benign and malignant classes. This classifier is sensitive to the existence of irrelevant features, which explains these results. Although the MLP classifier is competent to solve several non-separable problems, it was not able to make a good distinction between the classes. Furthermore, this type of classifier needs a long training time for the size of the feature set. The C4.5 classifier, on the other hand, resulted in a more balanced classification result between the two classes. However, this classifier can have difficulties in dealing with correlated features. All these classifiers can achieve superior results using feature selection algorithms.

In order to improve the classification results and to avoid over-fitting caused by a large number of features, several different feature selection algorithms were used to find the best features for the classification process. These algorithms considered two types of evaluators as mentioned earlier. The single-feature evaluators that use a ranking method, i.e., the correlation coefficient, GRFS, information gain, relief-F and PCA, were applied until a certain number of features are selected, which correspond to the stopping criterion belonging to the set $N = \{25,50,75\}$, with the exception of the PCA algorithm that chooses enough eigenvalues to rank the new transformed features. The maximum number of features $F = 5$ was used for the PCA algorithm in order to include this number of features in each transformed feature, and the proportion of variance $V = 0.95$ was used to retain a sufficient number of PC features. Accordingly, 31 eigenvalues were selected by the PCA algorithm to represent the vector with the new features. The number of nearest neighbours for the relief-F was defined as $k = 10$ for the feature estimation.

In the case of the feature subset evaluator, i.e., CFS, the greedy stepwise search method, in either forward or backward directions, was applied until the addition or removal of any feature in the subset caused a lower evaluation, i.e., low correlation to the class and high correlation with one or more of the other features relative to the previous evaluation. This resulted in 37 features selected by the forward direction and 50 by the backward direction. The best first search method was also performed in the directions: forward, backward or bi-direction. However, experimental results, using the classifiers discussed in the previous section, showed that this second method did not improve the classification

performance over that obtained using the stepwise search method alone. Therefore, only the stepwise method was used with CFS for comparison with the other feature selection algorithms.

Fig. 3 shows the percentage of selected features for each feature selection algorithm. The features were divided into five categories: shape, colour, fractal texture, wavelet texture and Haralick's texture; the percentage was computed individually for each category. Only the best configurations from the classification results were used for each feature selection algorithm and the features selected were: the first 75 ranked features from the correlation coefficient, GRFS, information gain and relief-F algorithms, the first 31 new features ranked by the PCA algorithm, and a subset of 50 features defined by the CFS algorithm.
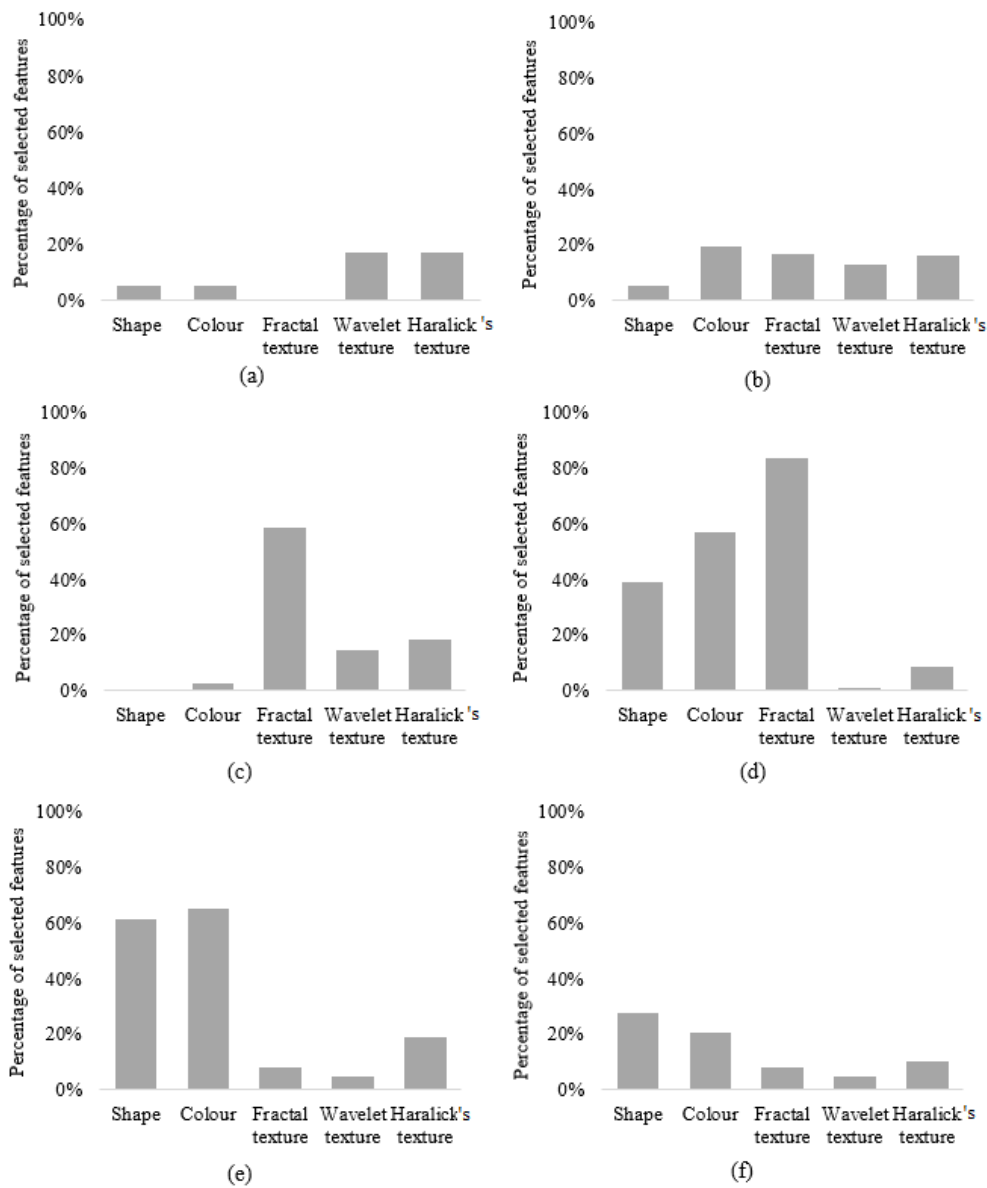

Fig. 3 Percentage of selected features after applying feature selection algorithms: (a) correlation coefficient, (b) GRFS, (c) information gain, (d) relief-F, (e) PCA, and (f) CFS.

Figure 3 shows that there were large differences between the feature selection algorithms. The correlation coefficient and information gain were the only algorithms that did not select features from all the categories. The PCA algorithm selected the greatest percentage of features from the shape and colour categories, whereas the information gain algorithm selected the greatest percentage of texture features. The relief-F algorithm selected over 80% of the fractal texture, but it did not select the wavelet and Haralick's texture features proportionally. On the other hand, the GRFS and CFS algorithms selected features from amongst all the categories in a more uniform way. The results of this feature selection process were evaluated using several different classifiers. The objective of this evaluation was to analyse which feature selection algorithms achieved the best classification results. The algorithms that select features from all the categories were expected to obtain the best classification results, according to the objective proposed in this study.

Table 4 shows the best classification results using the feature selection algorithms. These results, show that the OPF classifier with the features selected by the CFS algorithm and the MPL classifier with the features selected by the GRFS algorithm achieved superior results compared to the others, as presented in Table 4 (in bold). In addition, the features selected by the CFS and GRFS algorithms obtained better results for the classifiers than the other algorithms. As mentioned earlier, these algorithms selected the features of all the categories more uniformly (Fig. 3), which explains these results. The features selected by the PCA algorithm also obtained good results among the classifiers, despite the fact that it did not select the features uniformly; also the C4.5 classifier had a high SP result. However, this classifier did not stand out as much as the OPF and MPL classifiers, i.e., the C4.5 classifier had a higher classification cost.

Table 4 The best classification results using feature selection algorithms.

Classifier	FS algorithm (Search)	Features	ACC	SE	SP	C
kNN	CFS (Backward stepwise)	50	75.8%	67.8%	83.9%	0.225
Bayes Net	CFS (Forward stepwise)	37	74.4%	64.3%	84.4%	0.236
C4.5	PCA (Ranker)	31	89.7%	83.5%	**95.8%**	0.091
MLP	GRFS (Ranker)	75	**90.6%**	**86.6%**	94.6%	**0.086**
SVM	Relief-F (Ranker)	75	80.1%	76.1%	84.1%	0.191
OPF	CFS (Backward stepwise)	50	**91.6%**	**87.0%**	**96.2%**	**0.075**

The classification results are presented in more details in Fig. 4, where it is possible to analyse the variation of the accuracy, sensitivity and specificity, according to the number of ranked features defined by the correlation coefficient, GRFS, information gain and relief-F algorithms. Fig. 5 shows the variation of the results for the features selected by the PCA and CFS algorithms. In addition, the classification results for each feature selection are compared with the results using the entire set of features. From the feature selection, the OPF and kNN classifiers maintained their results, but they did not achieve better results. The MPL, C4.5 and Bayes Net classifiers had better results with the feature selection, whereas the SVM classifier achieved much better results with the entire set of features.
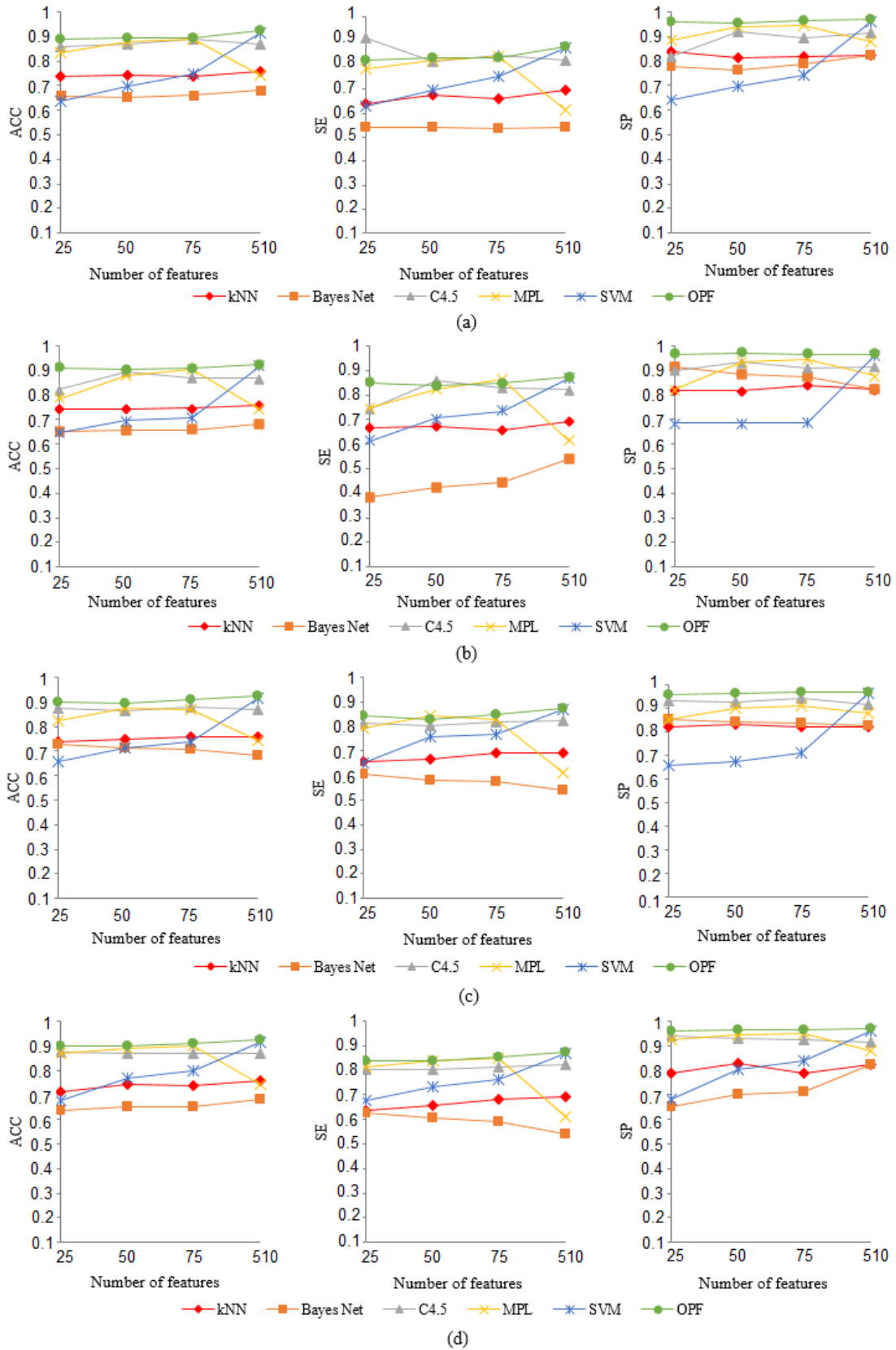
Fig. 4 Variation of the classification measures, according to the number of features defined by the ranker of each feature selection algorithm for all features of the dataset: (a) correlation coefficient, (b) GRFS, (c) information gain, and (d) relief-F.
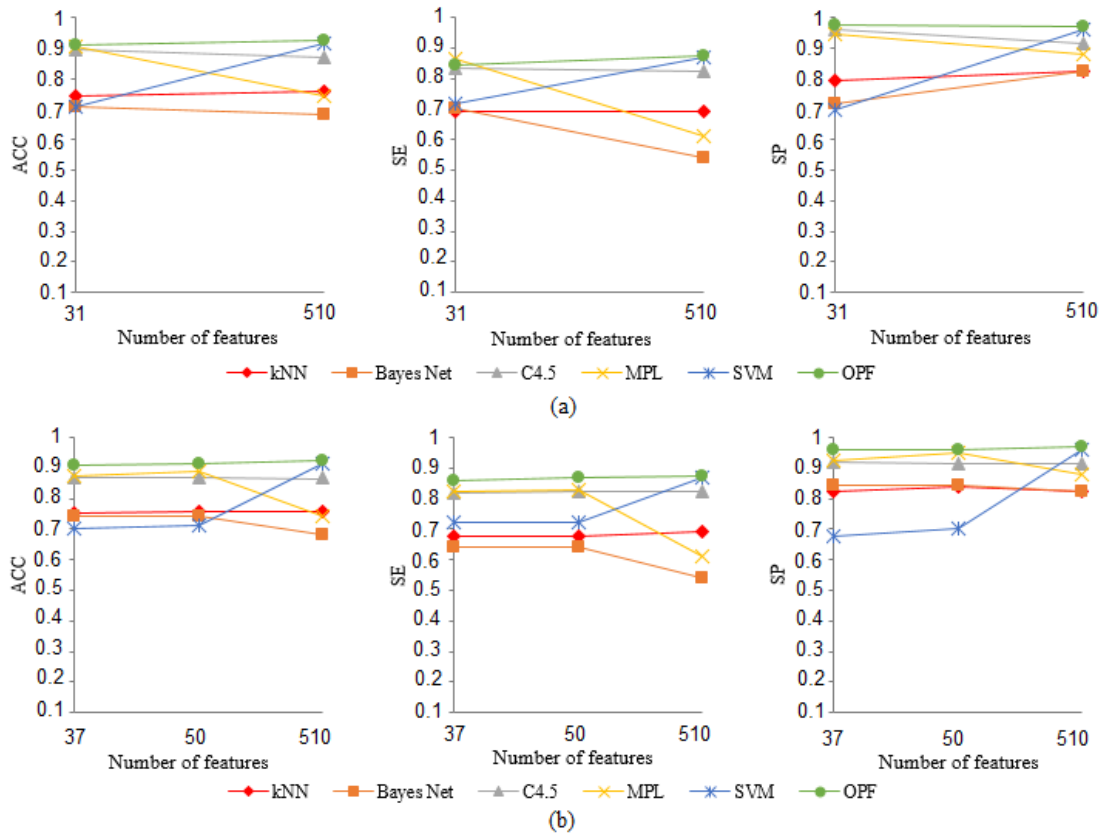
Fig. 5 Variation of the classification measures, according to the automatic number of features established by the feature selection algorithms for all features of the dataset: (a) PCA and (b) CFS.

In order to evaluate the/a combination of features (fractal texture, wavelet texture, and Haralick's texture categories combined with shape and colour features), as proposed in this study, some experiments considering feature subsets for each category individually and the best classifier achieved (OPF) were also performed. A texture subset, i.e., with the combination of all features of the texture categories achieved better results (ACC = 91.6%, SE = 86.8%, SP = 96.4%, C = 0.074) than using each category individual, i.e., fractal texture (ACC = 89.7%, SE = 84.1%, SP = 95.7%, C = 0.089), wavelet texture (ACC = 90.7%, SE = 85%, SP = 96.4%, C = 0.082), and Haralick's texture (ACC = 88.3%, SE = 80.1%, SP = 96.6%, C = 0.100). The extracted texture features combined with shape and colour features obtained superior results for skin lesion diagnosis (ACC = 92.3%, SE = 87.5%, SP = 97.1%, C = 0.067) than when only shape and colour features were used (ACC = 90.5%, SE = 85%, SP = 96%, C = 0.084).

4.4. Computational time

The proposed approach was developed using: 1) Visual Studio Express 2012 environment, C/C++ and OpenCV 2.4.9 library for the feature extraction algorithms; and 2) Eclipse IDE 4.6.1 environment, java 1.8.0_111, and Weka 3.8 library for the classification algorithms. Table 5 shows the computational time of the processing of all images for each task, which includes feature extraction, and classification with and without feature selection using the best classification model. All algorithms were performed on

24

an Intel(R) Core(TM) i5 CPU 650 @ 3.20 GHz with 8 GB of RAM, running Microsoft Windows 7 Professional 64-bits.

Table 5 Computational time for the feature extraction and classification tasks considering all images.

Task	Features	Time
Shape feature extraction	18	10.26 min
Colour feature extraction	72	10.12 min
Fractal feature extraction	12	26.79 min
Wavelet feature extraction	240	34.37 min
Haralick's feature extraction	168	29.48 min
Classification (without feature selection)	510	8.01 sec
Classification (with feature selection)	50	5.91 sec

The values in Table 5 indicate that the feature extraction step was the most time-consuming; however, the computation time required by this step can be considerably decreased using optimized C/C++ implementations. To find the lesion asymmetry, the proposed algorithm will take $O(n^2)$ time where n is the number of boundary points; however, the rotating callipers method [63] can be used to reduce the complexity to $O(n \log n)$.

5. Discussion

The main objective of this study was to evaluate and propose a set of features based on shape properties, colour variation and texture analysis, using several different methods, to diagnose skin cancer with a dataset of 1104 dermoscopic images. The full set of features (Table 1) achieved ACC = 92.3%, SE = 87.5% and SP = 97.1% using the OPF classifier. The best set of features from the selection process was obtained using the CFS algorithm and the OPF classifier that obtained ACC = 91.6%, SE = 87% and SP = 96.2%. This set was defined with the following features (Table 1): CO, CI, AR, s_s^2, s_s, μ_2, s_2^2, s_2, max_3, min_4, s_5^2, μ_6, s_6^2, SK_6, s_8^2, s_8, SK_8, max_9, s_{11}^2, s_{11}, D_3^2, $E(4)_2$, $E(3)_3$, $H(8)_3$, $E(8)_5$, $H(5)_5$, $H(6)_5$, $H(2)_9$, $H(3)_{10}$, $E(7)_{11}$, $H(2)_{12}$, $H(4)_{12}$, $H(7)_{12}$, VAR_2, SA_3, MCC$_3$, SV_4, CRL1$_4$, MCC$_4$, VAR_5, MCC$_5$, VAR_6, CRL1$_6$, IDM_8, DV$_8$, DH$_8$, SA_9, CRL1$_9$, SV_{11}, CRL1$_{11}$. The selected features were from all of the proposed categories, i.e., shape, colour, fractal texture, wavelet texture and Haralick's texture. In addition, the four colour spaces were considered by the automatic selection of the colour and texture features. Although the feature selection results reduced the number of features, i.e., removed the redundant and irrelevant features, the full set of features presented the best results, since the OPF classifier deals very well with redundant and irrelevant features.

There are some important issues to be analysed in this study regarding the extracted features. One of the texture extraction methods adopted in this article was based on DWT. There are also several other effective methods based on transform, such as discrete cosine transform (DCT) and wavelet packet decomposition (WPD) also known as tree-structured wavelet, which have been used for texture analysis in images [64,65]. Therefore, comparing the results of the combination of features proposed in this article using other transform methods would be very interesting in order to improve the findings of this

study. Since the extracted features in this study are all represented in one pool in sequence as mentioned earlier, the feature selection process using a sequential search strategy can select different features if the feature extraction considers another representation, e.g., randomly. However, this representation did not affect significantly the results of any of the studied classifiers. For example, only two different features were selected by the CFS algorithm, probably redundant features from the features defined before, because the OPF classifier achieved the same results and thus, the random representation did not influence its generalization.

One limitation with the research described in this article is that the experiments were based on only one strategy to reduce the unbalance of the classes, i.e., a combination between the under-sampling and over-sampling methods. Although this combination overcame the problem of the unbalanced classes, there are several other effective methods that can be used to deal with such a problem. For example, the synthetic minority over-sampling technique (SMOTE) [66], which is an over-sampling method for overcoming the over-fitting and expand the decision region for the minority class samples. Sampling methods can also be combined with ensemble methods for addressing unbalanced classes and they can present effective results [67]. The lack of a lesion segmentation process may be considered another limitation of the present study; however, ground-truth lesion segmentation masks were used in order to obtain a more accurate computational system. For example, the segmentation approach presented by Ma and Tavares [61] can be used to evaluate the effectiveness of the proposed classification model in the segmented images. On other hand, since the study did not use all the images of the original dataset as mentioned earlier, the results cannot be compared with the results obtained in the studies using the same dataset and the ground-truth lesion segmentation masks presented in Gutman et al. [57]. These studies considered a set of 1279 images partitioned into training and test sets. The best results were achieved by Lequan et al. [62] (with ACC = 0.855, SE = 0.547 and SP = 0.931), who proposed a novel method for melanoma recognition by leveraging very deep convolutional neural networks.

Several automatic diagnosis systems have been proposed using models with a single classifier for skin lesion classification, as was used in this study. In Celebi et al. [6] the proposed classification model based on the SVM classifier achieved SE = 93.33% and SP = 92.34% in a dataset of 564 dermoscopic images. The authors extracted 11 shape, 354 colour and 72 texture features. In Abbas et al. [25], the proposed system obtained SE = 88.2% and SP = 91.3% in a dataset of 120 dermoscopic images. These authors applied the SVM classifier to distinguish between benign and malignant lesions using asymmetry, border quantification, colour and differential structure features; however, the number of features used was not mentioned. Zortea et al. [60] proposed a computational system to differentiate benign lesions and melanoma using a discriminant analysis classifier, which achieved SE = 86% and SP = 52% in a dataset of 206 dermoscopic images. The feature extraction in this work used 6 asymmetry, 11 colour, 3 border, 3 geometry and 30 texture features of skin lesions.

Other diagnosis systems that used different feature extraction approaches can also be mentioned. For example, Sharma and Virmani [68] proposed a decision support system for the detection of renal diseases using GLCM statistical features and a SVM classifier from ultrasound images. The authors explored the potential of five texture feature vectors that were obtained in various ways using GLCM statistics exhaustively. The proposed system achieved the highest overall classification result of ACC = 85.7% for the differential diagnosis between normal and MRD images. Wang et al. [69] developed an improved parameter and structure identification of an adaptive neuro-fuzzy inference system (ANFIS) for feature extraction in images. Colour, morphology and texture features were used as inputs and the least-square and k-mean clustering methods were employed as the learning algorithms for such a system. The training errors for the affective values were tested and compared using the International Affective Picture System, which achieved 14% of maximum errors. A new approach of diagnosis by timed automata was proposed in Azzabi et al. [70]. The approach is based on the operating time and is applicable to systems whose dynamic evolution depends on the order of discrete events and on their duration as in industrial processes. The effectiveness of this approach was analysed in a hydraulic system.

Li et al. [71] proposed reliability indices for rule-based for rule-based knowledge presentation by using a back-propagation neural network with a Bayesian regularization algorithm. The proposed method was applied for shoe design in a KANSEI evaluation system, and it achieved superior performance compared to the other algorithms in terms of the performance, gradient, Mu, effective number of parameters, and the sum square parameter in KANSEI support and confidence time series prediction. In Ghosh et al. [72], a classification system for an automated glaucoma diagnosis was proposed. The proposed system is based on both the grid colour moment method as a feature vector to extract the colour feature and a neural network classifier. This system was tested using an open RIM-ONE database to classify both with and without glaucoma retina images and it achieved ACC = 87.47%. An effective method for analysing plantar pressure images in order to obtain the key areas of foot plantar pressure characteristics was proposed by Li et al. [73]. A plantar pressure imaging dataset of diabetic patients was used to evaluate the proposed method. First, the dataset was pre-processed by using watershed transformation to determine the region of interest. Afterward, the convolutional neural network based on k-mean clustering and parameterized manifold learning using an improved isometric mapping algorithm were used to attain segments of the imaging dataset. The experiments achieved an average accuracy of 80% for the clustering result, and the proposed manifold learning method achieved an average accuracy of 87.2%.

6. Conclusion and future works

In this article, a combination of features based on shape properties, colour variation and texture analysis using several different feature extraction methods was presented. Geometrical properties, lesion asymmetry and border irregularity were used for the extraction of the shape properties. Statistical

measures were used to analyse the colour features. The fractal dimension analysis, discrete wavelet transform and co-occurrence matrix methods were applied to obtain the texture features. Four colour spaces, i.e., *RGB*, *HSV*, *CIE Lab* and *CIE Luv*, were used for the extraction of both colour and texture properties. For the evaluation of the proposed feature extraction method, six different categories of classifiers were adopted; namely: kNN, Bayes networks, C4.5 decision tree, MLP, SVM and OPF. Furthermore, the classification performance was also evaluated using six different feature selection algorithms, which were: correlation coefficient, GRFS, information gain, relief-F, PCA and CFS.

Promising results were obtained with the proposed feature extraction for all the models evaluated. The best classification results were from the OPF classifier when all the features were used. The OPF results were: ACC = 92.3%, SE = 87.5% and SP = 97.1%. The OPF classifier also obtained the best classification results using feature selection algorithms for the skin lesion computational diagnosis system and achieved: ACC = 91.6%, SE = 87% and SP = 96.2%, when 50 features were selected using a CFS algorithm. It should be noted that the OPF classifier did not achieve better results by applying the feature selection algorithms, but it maintained the good results obtain when using all features. Moreover, the feature selection step reduced the computational time for the skin lesion classification. Another interesting result is that in most cases, the performance of the classifiers tends to improve when a percentage of features of all categories are select, i.e., shape, colour, fractal texture, wavelet texture and Haralick's texture by feature selection algorithms.

The main contributions of this study were: (1) the texture analysis based on four colour spaces, since the combination of several different colour spaces presented quite good results; skin lesion texture features proposed in the literature are usually extracted using grey-level images or only a few colour channels [25,7,6]; (2) the combination of several methods applied to analyse the skin lesion texture, including fractal dimension, wavelet transform and co-occurrence matrix based on colour image, since the combination presented better results than when only one texture method was used; and (3) the extracted texture features combined with shape and colour features obtained superior results compared to when such features are used seperately.

Future studies regarding the pigmented skin lesion classification of dermoscopic images should involve searching for new methods aiming to develop more efficient and effective systems for better skin lesion diagnoses. However, the classification results can be improved with ensemble methods [39,67,74]. Such methods consist of combining the results of several classification models in order to develop a more robust system that provides more accurate results than using a single classifier. Another solution to improve the classification results would be using deep learning architectures [75], since these architectures have shown that they have the capacity to learn from a large dataset.

Acknowledgments

Conflict of interest statement

The authors report no conflict of interest.

References

1. Scharcanski J, Celebi ME (2013) Computer Vision Techniques for the Diagnosis of Skin Cancer. Springer, Berlin, Heidelberg

2. Iyatomi H, Oka H, Celebi ME, Hashimoto M, Hagiwara M, Tanaka M, Ogawa K (2008) An improved Internet-based melanoma screening system with dermatologist-like tumor area extraction algorithm. Computerized Medical Imaging and Graphics 32 (7):566-579. doi:10.1016/j.compmedimag.2008.06.005

3. Barata C, Celebi ME, Marques JS (2017) Development of a clinically oriented system for melanoma diagnosis. Pattern Recognition 69:270-285. doi: 10.1016/j.patcog.2017.04.023

4. Johr RH (2002) Dermoscopy: alternative melanocytic algorithms-the ABCD rule of dermatoscopy, menzies scoring method, and 7-point checklist. Clinics in Dermatology 20 (3):240-247. doi:10.1016/S0738-081X(02)00236-5

5. Maglogiannis I, Doukas CN (2009) Overview of advanced computer vision systems for skin lesions characterization. IEEE Transactions on Information Technology in Biomedicine 13 (5):721-733. doi:10.1109/titb.2009.2017529

6. Celebi ME, Kingravi HA, Uddin B, Iyatomi H, Aslandogan YA, Stoecker WV, Moss RH (2007) A methodological approach to the classification of dermoscopy images. Computerized Medical Imaging and Graphics 31 (6):362-373. doi:10.1016/j.compmedimag.2007.01.003

7. Garnavi R, Aldeen M, Bailey J (2012) Computer-Aided Diagnosis of Melanoma Using Border- and Wavelet-Based Texture Analysis. IEEE Transactions on Information Technology in Biomedicine 16 (6):1239-1252. doi:10.1109/titb.2012.2212282

8. Celebi ME, Zornberg A (2014) Automated quantification of clinically significant colors in dermoscopy images and its application to skin lesion classification. IEEE Systems Journal 8 (3):980-984. doi:10.1109/JSYST.2014.2313671

9. Shimizu K, Iyatomi H, Celebi ME, Norton K-A, Tanaka M (2015) Four-class classification of skin lesions with task decomposition strategy. IEEE Transactions on Biomedical Engineering 62 (1):274-283. doi:10.1109/TBME.2014.2348323

10. Barata C, Celebi ME, Marques JS, Rozeira J (2016) Clinically inspired analysis of dermoscopy images using a generative model. Computer Vision and Image Understanding 151:124-137. doi:10.1016/j.cviu.2015.09.011

11. Sadri AR, Azarianpour S, Zekri M, Celebi ME, Sadri S (2017) WN-based approach to melanoma diagnosis from dermoscopy images. IET Image Processing 11 (7):475 – 482. doi:10.1049/iet-ipr.2016.0681

12. Barata C, Ruela M, Francisco M, Mendonça T, Marques JS (2013) Two Systems for the Detection of Melanomas in Dermoscopy Images using Texture and Color Features. IEEE Systems Journal 8 (3):965-979. doi:10.1109/JSYST.2013.2271540

13. Materka A, Strzelecki M (1998) Texture analysis methods: a review. COST B11 report, Technical University of Lodz. Brussels

14. Iyatomi H, Norton K, Celebi ME, Schaefer G, Tanaka M, Ogawa K (Year) Classification of melanocytic skin lesions from non-melanocytic lesions. In: Annual International Conference of the

IEEE Engineering in Medicine and Biology Society Buenos Aires, August 31 - September 4 2010. IEEE, pp 5407-5410. doi:10.1109/iembs.2010.5626500

15. Celebi ME, Iyatomi H, Stoecker WV, Moss RH, Rabinovitz HS, Argenziano G, Soyer HP (2008) Automatic detection of blue-white veil and related structures in dermoscopy images. Computerized Medical Imaging and Graphics 32 (8):670-677. doi:10.1016/j.compmedimag.2008.08.003

16. Oliveira RB, Marranghello N, Pereira AS, Tavares JMRS (2016) A computational approach for detecting pigmented skin lesions in macroscopic images. Expert Systems with Applications 61:53-63. doi:10.1016/j.eswa.2016.05.017

17. Leo GD, Paolillo A, Sommella P, Fabbrocini G (Year) Automatic Diagnosis of Melanoma: A Software System Based on the 7-Point Check-List. In: 43rd International Conference on System Sciences, Hawaii January 5-8 2010. IEEE, pp 1-10. doi:10.1109/hicss.2010.76

18. Yuan X, Yang Z, Zouridakis G, Mullani N (Year) SVM-based texture classification and application to early melanoma detection. In: 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, New York, August 30 - September 3 2006. IEEE, pp 4775-4778. doi:10.1109/IEMBS.2006.260056

19. Webb AR (2003) Statistical pattern recognition. 2 edn. John Wiley & Sons, England

20. Japkowicz N, Shah M (2011) Evaluating Learning Algorithms: A Classification Perspective. Cambridge University Press, USA

21. Rahman MM, Bhattacharya P, Desai BC (Year) A multiple expert-based melanoma recognition system for dermoscopic images of pigmented skin lesions. In: 8th IEEE International Conference on International Conference on BioInformatics and BioEngineering, Athens, October 8-10 2008. IEEE, pp 1-6. doi:10.1109/BIBE.2008.4696799

22. Papa JP, Falcao AX, Suzuki CT (2009) Supervised pattern classification based on optimum- path forest. International Journal of Imaging Systems and Technology 19 (2):120-131. doi:10.1002/ima.20188

23. Guyon I, Gunn S, Nikravesh M, Zadeh L (2006) Feature extraction: foundations and applications, vol 207. Studies in Fuzziness and Soft Computing. Springer-Verlag, Berlin, Heidelberg. doi:10.1007/978-3-540-35488-8

24. Oliveira RB, Papa JP, Pereira AS, Tavares JMRS (2016) Computational Methods for Pigmented Skin Lesion Classification in Images: Review and Future Trends. Neural Computing and Applications 27:1-24. doi:10.1007/s00521-016-2482-6

25. Abbas Q, Celebi ME, Garcia IF, Ahmad W (2013) Melanoma recognition framework based on expert definition of ABCD for dermoscopic images. Skin Research and Technology 19 (1):e93-e102. doi:10.1111/j.1600-0846.2012.00614.x

26. Costa LdF, Cesar Junior RM (2009) Shape classification and analysis: theory and practice. 2 edn. Crc Press, USA

27. Clawson KM, Morrow P, Scotney B, McKenna J, Dolan O (Year) Analysis of pigmented skin lesion border irregularity using the harmonic wavelet transform. In: 13th International Machine Vision and Image Processing Conference Dublin, September 2-4 2009. IEEE, pp 18-23

28. Zhou Y, Smith M, Smith L, Warr R (2010) A new method describing border irregularity of pigmented lesions Skin Research and Technology 16:66-76. doi:10.1111/j.1600-0846.2009.00403.x

29. Lee TK, McLean DI, Atkins MS (2003) Irregularity index: a new border irregularity measure for cutaneous melanocytic lesions. Medical image analysis 7 (1):47-64. doi:10.1016/S1361-8415(02)00090-7

30. Celebi ME, Iyatomi H, Schaefer G, Stoecker WV (2009) Lesion border detection in dermoscopy images. Computerized medical imaging and graphics 33 (2):148-153. doi:10.1016/j.compmedimag.2008.11.002

31. Lissner I, Urban P (2012) Toward a unified color space for perception-based image processing. IEEE Transactions on Image Processing 21 (3):1153-1168. doi:10.1109/TIP.2011.2163522

32. Tkalcic M, Tasic JF (Year) Colour spaces: perceptual, historical and applicational background. In: Proceedings in the IEEE Region 8 EUROCON 2003: Computer as a Tool Ljubljana, September 22-24 2003. IEEE, pp 304-308. doi:10.1109/EURCON.2003.1248032

33. Silva CS, Marcal AR (2013) Colour-based dermoscopy classification of cutaneous lesions: an alternative approach. Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization 1 (4):211-224. doi:10.1080/21681163.2013.803683

34. Al-Akaidi M (2004) Fractal speech processing. Cambridge university press, New York

35. Scheunders P, Livens S, Van de Wouwer G, Vautrot P, Van Dyck D (1998) Wavelet-based texture analysis. International Journal on Computer Science and Information Management 1 (2):22-34

36. Haralick RM, Shanmugam K, Dinstein IH (1973) Textural features for image classification. IEEE Transactions on Systems, Man and Cybernetics SMC-3 (6):610-621. doi:10.1109/TSMC.1973.4309314

37. Mallat SG (1987) A theory for multiresolution signal decomposition: the wavelet representation. IEEE Transactions on Pattern Analysis and Machine Intelligence 11 (7):674-693. doi: 10.1109/34.192463

38. Abedini M, Chen Q, Codella NCF, Garnavi R, Sun X (2015) Accurate and Scalable System for Automatic Detection of Malignant Melanoma. In: Celebi ME, Mendonca T, Marques JS (eds) Dermoscopy Image Analysis. CRC Press, Boca Raton, pp 293-343. doi:10.1201/b19107-11

39. Witten IH, Frank E, Hall MA (2016) Data Mining: Practical machine learning tools and techniques. 4 edn. Morgan Kaufmann, San Francisco

40. Chawla NV (2005) Data mining for imbalanced datasets: An overview. In: Maimon O, Rokach L (eds) Data mining and knowledge discovery handbook. Springer, New York, pp 853-867. doi:10.1007/0-387-25465-X_40

41. Guyon I, Elisseeff A (2003) An introduction to variable and feature selection. The Journal of Machine Learning Research 3:1157-1182

42. Dash M, Liu H (1997) Feature selection for classification. Intelligent data analysis 1 (3):131-156. doi:10.1016/S1088-467X(97)00008-5

43. Liu H, Yu L (2005) Toward integrating feature selection algorithms for classification and clustering. IEEE Transactions on Knowledge and Data Engineering 17 (4):491-502. doi:10.1109/TKDE.2005.66

44. Kononenko I (1994) Estimating attributes: Analysis and extensions of RELIEF. In: Bergadano F, De Raedt L (eds) Machine Learning: ECML-94, vol 784. Lecture Notes in Computer Science. Springer, Berlin, Heidelberg, pp 171-182. doi:10.1007/3-540-57868-4_57

45. Cover T, Hart P (1967) Nearest neighbor pattern classification. IEEE Transactions on Information Theory 13 (1):21-27. doi:10.1109/tit.1967.1053964

46. Shannon CE (1948) A mathematical theory of communication. The Bell System Technical Journal 27 (1):379-423. doi:10.1002/j.1538-7305.1948.tb01338.x

47. Hall MA (Year) Correlation-based Feature Selection for Discrete and Numeric Class Machine Learning. In: Proceedings of the 17th International Conference on Machine Learning, San Francisco, USA, June 29 - July 02 2000. Morgan Kaufmann Publishers Inc., 657793, pp 359-366

48. Hand D, Mannila H, Smyth P (2001) Principles of Data Mining. The MIT Press, London

49. Kohavi R (Year) A study of cross-validation and bootstrap for accuracy estimation and model selection. In: Proceedings of the 14th International Joint Conference on Artificial Intelligence, Quebec, Canada, August 20-25 1995. Morgan Kaufmann Publishers Inc., pp 1137-1145

50. Congdon P (2007) Bayesian statistical modelling, vol 704. 2 edn. John Wiley & Sons, Chichester

51. Quinlan JR (1993) C4.5: programs for machine learning. Morgan Kaufmann Publishers Inc., USA,

52. Haykin SS (1999) Neural networks: a comprehensive foundation. Prentice Hall, Englewood Cliffs, USA

53. Burges CJC (1998) A tutorial on support vector machines for pattern recognition. Data mining and knowledge discovery 2 (2):121-167. doi:10.1023/A:1009715923555

54. Han J, Kamber M (2006) Data Mining: concepts and techniques. Elsevier, San Francisco

55. Platt JC (1999) Fast training of support vector machines using sequential minimal optimization. In: Advances in kernel methods. MIT Press Cambridge, USA, pp 185-208

56. Amorim WP, Falcão AX, de Carvalho MH (Year) Semi-supervised pattern classification using optimum-path forest. In: 27th SIBGRAPI Conference on Graphics, Patterns and Images, Rio de Janeiro, August 26-30 2014. IEEE, pp 111-118. doi:10.1109/SIBGRAPI.2014.45

57. Gutman D, Codella NCF, Celebi E, Helba B, Marchetti M, Mishra N, Halpern AC (Year) Skin Lesion Analysis toward Melanoma Detection: A Challenge at the International Symposium on Biomedical Imaging (ISBI) 2016, hosted by the International Skin Imaging Collaboration (ISIC), arXiv preprint arXiv:1605.01397

58. Arroyo JLG, Zapirain BG (2014) Detection of pigment network in dermoscopy images using supervised machine learning and structural analysis. Computers in biology and medicine 44:144-157. doi:10.1016/j.compbiomed.2013.11.002

59. Maglogiannis I, Delibasis KK (2015) Enhancing classification accuracy utilizing globules and dots features in digital dermoscopy. Computer methods and programs in biomedicine 118 (2):124-133. doi:10.1016/j.cmpb.2014.12.001

60. Zortea M, Schopf TR, Thon K, Geilhufe M, Hindberg K, Kirchesch H, Møllersen K, Schulz J, Skrøvseth SO, Godtliebsen F (2014) Performance of a dermoscopy-based computer vision system for the diagnosis of pigmented skin lesions compared with visual evaluation by experienced dermatologists. Artificial intelligence in medicine 60 (1):13-26. doi:10.1016/j.artmed.2013.11.006

61. Ma Z, Tavares JMRS (2016) A Novel Approach to Segment Skin Lesions in Dermoscopic Images Based on a Deformable Model. IEEE Journal of Biomedical and Health Informatics 20 (2):615-623. doi:10.1109/JBHI.2015.2390032

62. Lequan Y, Chen H, Dou Q, Qin J, Heng PA (2016) Automated Melanoma Recognition in Dermoscopy Images via Very Deep Residual Networks. IEEE Transactions on Medical Imaging 1-11. doi:10.1109/TMI.2016.2642839

63. Toussaint GT (Year) Solving geometric problems with the rotating calipers. In: Proceedings of IEEE Melecon, Athens, 1983. pp 1-8

64. Yu-Len H, Ruey-Feng C (Year) Texture features for DCT-coded image retrieval and classification. In: IEEE International Conference on Acoustics, Speech, and Signal Processing, Phoenix, 15-19 Mar 1999 1999. IEEE, pp 3013-3016. doi:10.1109/ICASSP.1999.757475

65. Chang T, Kuo CCJ (1993) Texture analysis and classification with tree-structured wavelet transform. IEEE Transactions on Image Processing 2 (4):429-441. doi:10.1109/83.242353

66. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP (2002) SMOTE: synthetic minority over-sampling technique. Journal of Artificial Intelligence Research 16:321-357. doi:10.1613/jair.953

67. Schaefer G, Krawczyk B, Celebi ME, Iyatomi H (2014) An ensemble classification approach for melanoma diagnosis. Memetic Computing 6 (4):233-240. doi:10.1007/s12293-014-0144-8

68. Sharma K, Virmani J (2017) A Decision Support System for Classification of Normal and Medical Renal Disease Using Ultrasound Images: A Decision Support System for Medical Renal Diseases. International Journal of Ambient Computing and Intelligence 8 (2):52-69. doi:10.4018/IJACI.2017040104

69. Wang D, He T, Li Z, Cao L, Dey N, Ashour AS, Balas VE, McCauley P, Lin Y, Xu J (2016) Image feature-based affective retrieval employing improved parameter and structure identification of adaptive neuro-fuzzy inference system. Neural Computing and Applications:1-16. doi: 10.1007/s0052

70. Azzabi O, Njima CB, Messaoud H (2017) New Approach of Diagnosis by Timed Automata. International Journal of Ambient Computing and Intelligence 8 (3):76-93. doi:10.4018/IJACI.2017070105

71. Li Z, Shi K, Dey N, Ashour AS, Wang D, Balas VE, McCauley P, Shi F (2017) Rule-based back propagation neural networks for various precision rough set presented KANSEI knowledge prediction: a case study on shoe product form features extraction. Neural Computing and Applications 28 (3):613-630. doi: doi.org/10.1007/s0052

72. Ghosh A, Sarkar A, Ashour AS, Balas-Timar D, Dey N, Balas VE (2015) Grid color moment features in glaucoma classification. International Journal of Advanced Computer Science Applications 6 (9):1-14. doi: 10.14569/IJACSA.2015.060913

73. Li Z, Dey N, Ashour AS, Cao L, Wang Y, Wang D, McCauley P, Balas VE, Shi K, Shi F (2017) Convolutional Neural Network Based Clustering and Manifold Learning Method for Diabetic Plantar Pressure Imaging Dataset. Journal of Medical Imaging and Health Informatics 7 (3):639-652. doi:10.1166/jmihi.2017.2082

74. Kuncheva LI (2014) Combining pattern classifiers: methods and algorithms. 2 edn. John Wiley & Sons, New Jersey

75. Bengio Y (2009) Learning deep architectures for AI. Foundations and trends® in Machine Learning 2 (1):1-127. doi:10.1561/2200000006