Biomedical Signal Processing and Control 57 (2020) 101765

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

Skin lesion classification enhancement using border-line features – The melanoma vs nevus problem

Pedro M.M. Pereira [a,b,*], Rui Fonseca-Pinto [b,c], Rui Pedro Paiva [a], Pedro A.A. Assuncao [b,c], Luis M.N. Tavora [c], Lucas A. Thomaz [b], Sergio M.M. Faria [b,c]

[a] CISUC, Department of Informatics Engineering, University of Coimbra
[b] Instituto de Telecomunicações, Portugal
[c] Polytechnic of Leiria, Portugal

ARTICLE INFO

Article history:
Received 21 August 2019
Received in revised form 16 October 2019
Accepted 10 November 2019
Available online 24 November 2019

Keywords:
Medical imaging
Skin lesion
Image segmentation
Feature extraction
Classification

ABSTRACT

Machine learning algorithms are progressively assuming an important role as a computational tool to support clinical diagnosis, namely in the classification of pigmented skin lesions. The current classification methods commonly rely on features derived from shape, colour, or texture, obtained after image segmentation, but these do not always guarantee the best results. To improve the classification accuracy, this work proposes to further exploit the border-line characteristics of the lesion segmentation mask, by combining gradients with local binary patterns (LBP). In the proposed method, these border-line features are used together with the conventional ones to enhance the performance of skin lesion classification algorithms.

When the new features are combined with the classical ones, the experimental results show higher accuracy, which impacts positively the overall performance of the classification algorithms. As the medical image datasets usually present large class imbalance, which results in low sensitivity for the classifiers, the border-line features have a positive impact on this classification metric, as evidenced by the experimental results. Both the features' usefulness and their impact are assessed in regard to the classification results, which in turn are statistically tested for completeness, using three different classifiers and two medical image datasets.

1. Introduction

Digital image processing techniques are gaining increasing relevance in many computational tools that assist clinicians in their diagnostic decisions, as noted in the dermatology field for skin lesion classification, namely for the problem of differentiating melanoma and nevus [1]. The algorithms employed in such tasks range from those using Deep Learning, where the algorithm automatically select which types of features will be employed for classification, to other classic Machine Learning (ML) algorithms which require hand-crafted features. The use of Deep Learning algorithms has achieved significant performances (e.g. [2]), however those algorithms require rich precisely annotated datasets that are not usually available.

Still, many different types of features have been specially tailored for this problem, ranging from colour, texture, and shape. In this work we employ a recent sub-group of hand-crafted features, which are extracted from the border-line of the skin lesions [3]. The pipeline approach proposed in this work encompasses detailed segmentation methods, so that this new type of features may be explored.

The main contribution of this paper is to demonstrate the relevance of the newly proposed features (extracted from the segmentation border-line information) for improving a classification algorithm, in addition to other commonly used features [4]. Hence, this work does not aim to validate the segmentation or classification algorithms, but the information gain achieved with the proposed features. The proposed method uses two existing segmentation techniques: Gradient-based Histogram Thresholding (GHT) [5] and a variant of the recent Local Binary Patterns Clustering (LBPC) [6]. The border-line features are then extracted and used as input for automatic classification of melanocytic lesions using ML algorithms. To this end only images of melanoma and nevus are used. Melanoma is the most aggressive form of skin cancer [7] and one-third of all melanomas arise from pre-existing nevus, thus detection and removal of such nevus is of utmost importance in the prevention of melanoma. In summary, this paper exploits the additional insight that, based on statistical features, border-line information

* Corresponding author at: DEI – FCTUC, University of Coimbra, Portugal.
E-mail address: pedrommpereira@co.it.pt (P.M.M. Pereira).

https://doi.org/10.1016/j.bspc.2019.101765
1746-8094/

provides relevant information to enhance discrimination of skin lesions.

The reminder of the paper is organised as follows: Section 2 presents the literature involved in this work. Section 3 introduces the proposed approach, detailing the segmentation techniques, feature extraction and the pigmented skin lesion classification. Section 4 details about the datasets utilised for this study. Section 5 presents and discusses the results with statistical validation information, and Section 6 highlights the conclusions and future work.

2. Background

Segmentation provided by dermatologists (i.e., clinical segmentation), which is broadly used as ground-truth in the context of ML algorithms, presents some associated methodological constraints. In fact, handmade lesion segmentation aims at establishing a guide to surgical excisions or to be a follow-up reference for future medical examinations, defining a region of interest that comprises not only the pigmented lesion but also some of the surrounding area in the transition to healthy skin [8]. The transition region between the lesion and the clinical segmentation presents perceptual challenges to the human visual system. In fact, ambiguity in contrast variations, blurred edges or changes in lightning conditions contribute to the lack of accuracy in eye-based segmentation approaches [9]. Given these limitations, computational methods have been developed as alternative to be used in the segmentation step of ML algorithms [10,11]. Hence, the latter widely established option for more accurate computer vision delineation methods is considered in this work, rather than coarser versions available for clinical segmentation.

Segmentation methods have been widely used as a preprocessing step in skin imaging systems [12,13,4,14], and in some applications, prior to the segmentation step a few stages must precede the main algorithm, including artefact/hairs removal techniques such as [15,16]. A comprehensive review of such methods is presented in [17].

Only recently, delineation information of skin lesions borderline has emerged as an input feature in ML algorithms. Based on a detailed segmentation border, several feature characteristics of its perimeter may be determined. For such purpose, some research work has been carried out in this topic. In [1], using a clustering algorithm, the authors extracted features related to the perimeter border from 90 images and classified them as Benign vs Malignant and Unknown. Similarly, in [18] the authors also extracted the perimeter information, as well as colour and texture features. These emerging features provide new dimensions to the solution and have potential to be added directly to most existing approaches. One of such works is presented in [19], where several features are benchmarked with Matlab classifiers (present in the classifier app), attaining an average accuracy of 87%. Other implementations, like those based on Deep Learning, namely [2,20], already obtain state-of-the-art results but it is more difficult to include such features in their approaches. However, most neural network applications, like [21], are easily adaptable since they already behave like a feature classifier by construct. Currently, it was also introduced in [3] a feature-based descriptor for skin lesions that mainly includes some types of border-related features.

In spite of the fact that such approaches provide state-of-the-art results, they do not utilise the same datasets or classifiers, making the comparison of their results a difficult task, which will not be performed in the scope of this work.

3. Proposed skin lesion classification approach

This section presents the method's pipeline, which comprises three steps: Segmentation, Feature Extraction, and Classification. Firstly, the lesion image is segmented with the GHT and the LBPC algorithms (introduced in previous publications from the authors, i.e. [5,6]). Then common border-line features of the binary segmentation mask (presented for the first time in this paper) are extracted and used by the lesion classifier in the third and final step.

3.1. Segmentation

In this work two methods are used for segmentation of pigmented skin lesion images: GHT [5] and LBPC [6]. While GHT exploits luminance intensity variations, LBPC is more sensitive to texture patterns, therefore in both methods the luminance (grayscale) image is used, as described in the following sections. Given an RGB image as input, the corresponding luminance image (Y) is obtained by means of a weighted sum of R, G, and B channels, as defined in Rec.ITU-R BT.601-7 [22], Eq. (1).

$$Y = 0.2989 * R + 0.5870 * G + 0.1140 * B. \quad (1)$$

3.1.1. Gradient-based Histogram Thresholding

In general, skin lesion images produce bi-modal colour histograms [23], as can be seen in the workflow of the GHT method presented in Fig. 1. Therefore, two dominant peaks are identified in the histogram of the grayscale image, which are associated to the two major regions of interest: normal skin and lesion. In order to differentiate these two regions, in [5] the authors proposed a novel GHT approach specially tailored for this conditions, which determines a threshold value between the two peaks. This value can be calculated based on an optimum criterion that targets the maximum gradient along the segmentation mask perimeter [5]. This average gradient estimation involves only the component perpendicular to the segmentation curve and is calculated as the mean RGB gradient.

Nevertheless, sometimes the gradient-based thresholding algorithm may result in segmentation contours that only include inner parts of the lesion, as shown in Fig. 1 (green segmentation contour). Although the line does indeed exhibit the highest possible gradient, it can be observed that it excludes relevant portions of the lesion-to-skin transition region. To overcome this limitation, a clustering-based mask of the original image is determined using the *k-means++* algorithm [24]. Despite the fact that this segmentation strategy does not result in border-lines with mean gradient values as high as the ones previously obtained, it is more effective in including the regions that are to be associated with lesion area. The final GHT segmentation combines the strengths of both techniques by determining the optimum border-line (blue curve in Fig. 1) as a trade-off between high gradient values, from the gradient-based histogram, and high accurate lesion regions (often larger), from the clustering approach (a detailed discussion can be found in [5]).

3.1.2. Local Binary Pattern Clustering (LBPC)

We also tackled the problem of image segmentation following a LBPC approach, where the original algorithm was adapted as follows. The LBPC algorithm combines the LBP characteristics [25] with *k-means++* clustering, as depicted in the schematic representation of Fig. 2. The RGB image is converted to grayscale and a set of pixel-wise LBPs are determined (named "LBP" in Fig. 2) – although other colour transformations have been studied, the best performance was achieved using grayscale. Then, a modified LBP (binary) image is generated by setting to zeros all the elements whose LBP values are powers of two or zero (LBP $\in \{0, 2^n\}$, $n \geq 0$), and the remainder of this image is set as ones. These elements have been chosen due to their ability to provide rich information about smooth healthy skin regions despite the various intensity levels, according to [6]. Then, a new image is created by expanding the regions around the selected pixels with a homogeneous Gaussian kernel, empirically defined with a size of 13 pixels and

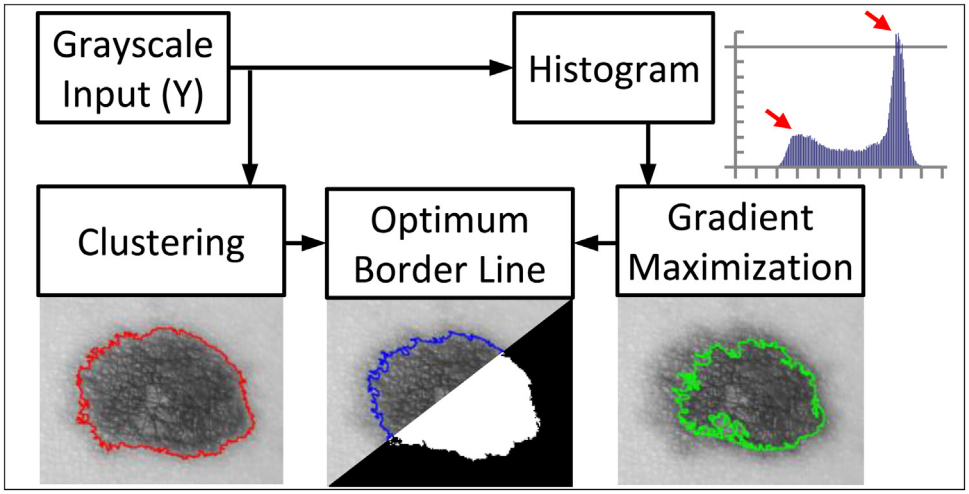

Fig. 1. Gradient-based Histogram Thresholding method workflow: given a grayscale input image, an histogram is produced to find the two dominant colour peaks (arrows) of the image, which provide boundaries for the RGB gradient maximisation step (green segmentation line); the same input image goes through a clustering step (red segmentation line); finally, from the two previous segmentation lines, an optimum border-line optimum is obtained (blue curve) from which image binarisation produces the final mask.

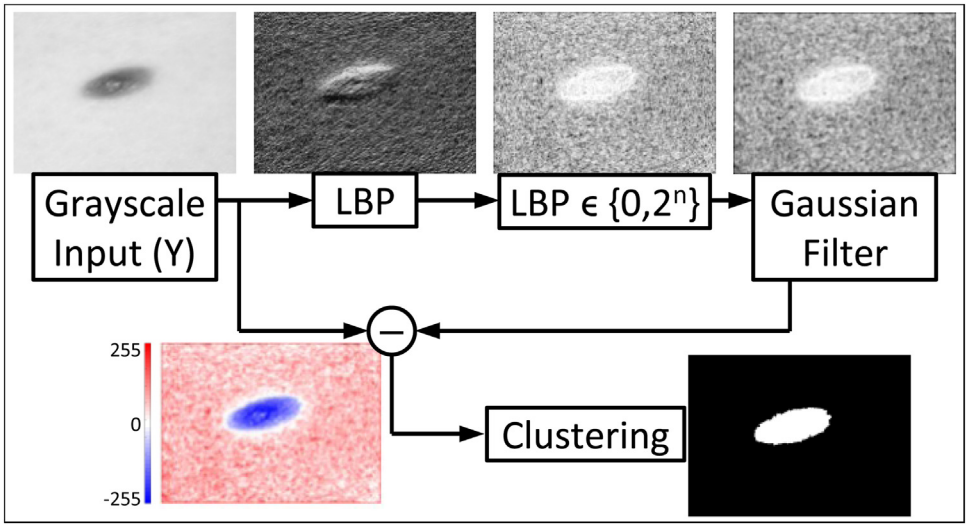

Fig. 2. Local Binary Pattern Clustering method workflow: given a grayscale input image, the pixel-wise LBP information (with values ranging from 0 to 255) is determined, based on a specific set of LBPs (0 and possible powers of 2), and smoothed with a Gaussian filter; this information is subtracted from the input grayscale image (which now ranges from −255 to 255, mapped from red to blue respectively) and fed to a *k-means* algorithm that clusters the information into two regions.

standard deviation 3. Finally, the transformed LBP image is subtracted from the original grayscale image from the first step and the resulting pixel values are input to the *k-means++* clustering algorithm (resorting to Euclidean distance and a maximum of 100 iterations) that separates the information into two cluster regions: lesion and normal skin. At this point each pixel is assigned to only one of the clusters. The pixel labels are then used to form a binary segmentation mask.

When compared to the original version of the LPBC algorithm, the novelty herein introduced is this image subtraction step, rather than the original representation inspired on the CIE $L^*a^*b^*$ colour space, as it was found to lead to better performances at this study' classification step.

The resulting segmentation output mask obtained with these two techniques can be observed in Fig. 3 for image *B*355*b* of the Dermofit dataset [26]. As it can be seen, the segmentation from both the GHT and LBPC algorithms provides much higher detail on the lesion borders than that of the dataset Ground-Truth (Fig. 3b).

3.2. Feature extraction from the segmentation border

In order to extract features from the proposed segmentation masks, the detailed border-lines are reshaped from their rounded lesion-shape to an unfolded line, resulting in the lines shown in Fig. 4a and b. The line unfolding is carried out by firstly calculating the centre of mass of the segmented region. Then, the Euclidean distance d (in pixels) from each pixel to the centre of mass is represented by $d(i)$ and, from this representation, the new line-segment is obtained. This unfolded line maintains all the original information, except the lesion shape. As can be observed from Fig. 4a and b, the lines segments originated from both algorithms have different sizes. This is due to the segmentation boundary generated by each corresponding algorithm. Although the algorithms provide similar shaped-segmentations, GHT displays a smoother curve than LBPC. Hence the LBPC segmentation is intrinsically larger (in terms of perimeter pixels) than GHT. This is further evidenced observing Fig. 3.

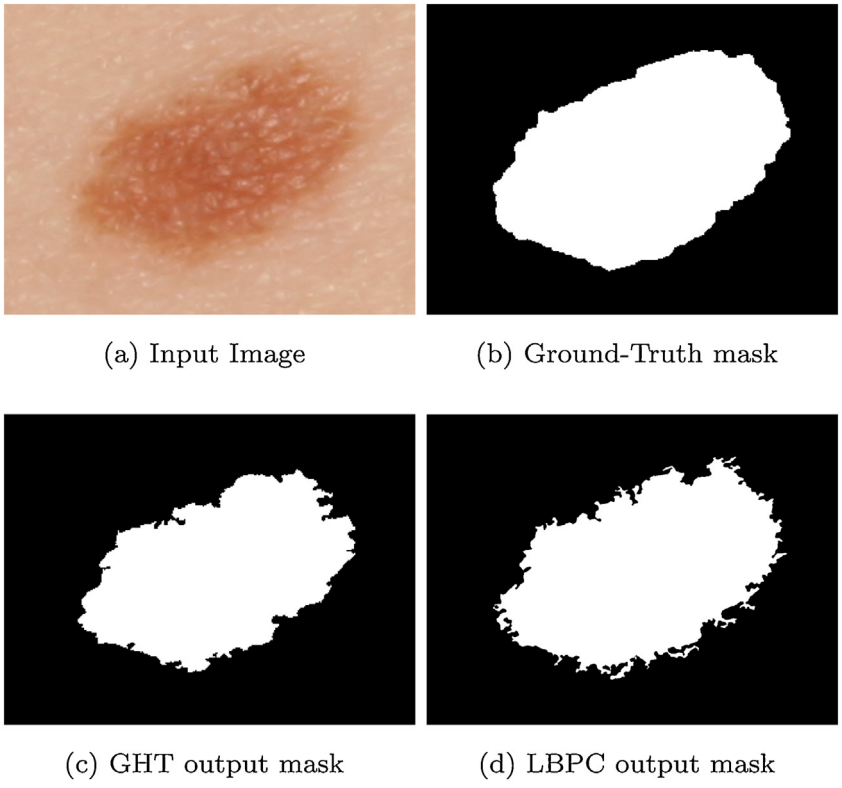

Fig. 3. Segmentation results for *B355b* image of the Dermofit dataset [26].

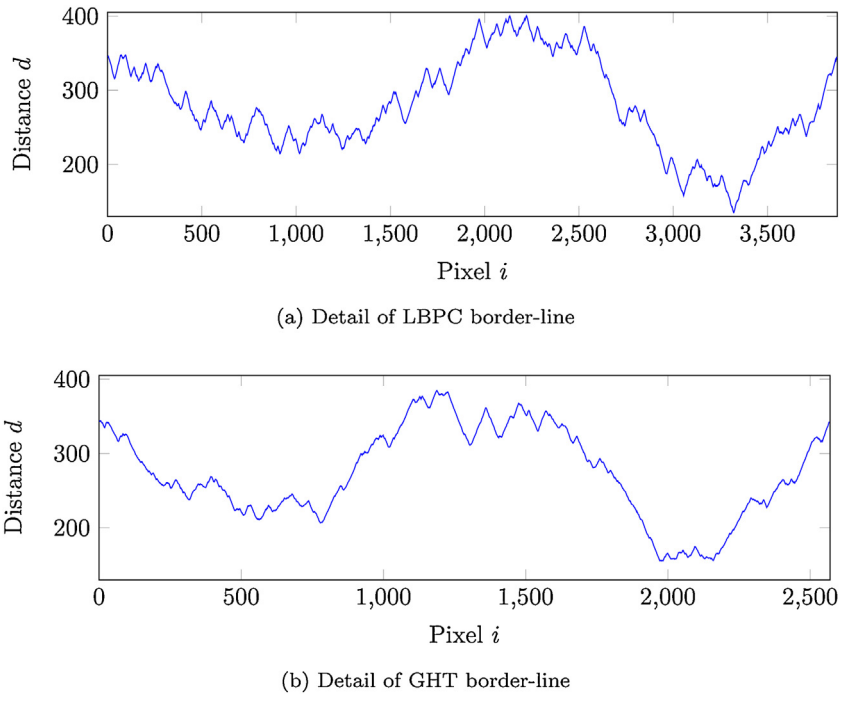

Fig. 4. Border-lines extracted from *B355b* image of Dermofit dataset [26].

Based on this representation, a new set of features were extracted from the unfolded border-line (as can be seen in Fig. 4a), namely: root-mean-square level (*F1*); average d value (*F2*); height of main peak (*F3*) and height and position of second peak of a autocorrelation sequence calculation (*F4–5*); magnitude of the highest peak of each of the first six bins of a Discrete Fourier Transform spectrum (DCT) using 4096 points, where the sampling resolution is $2\pi/4096$ rad/sample (dividing it in 32 equal-size bins) (*F6–11*); frequency component corresponding to the six points of the previous features (*F12–17*); sum of values of five equal-length segments produced by splitting its periodogram power spectral density (PSD) [27] (*F18–22*). The number of peaks/segments (*F6-22*) was optimised using correlation analysis.

3.3. Classification

As previously mentioned, the main research question addressed in this work is to verify whether segmentation border details and the type of lesion might be somehow correlated. This is done in a nevu versus melanoma setting. It is known that the melanoma is the most aggressive form of skin cancer [7] and one-third of all melanomas arise from pre-existing nevus. Thus, detection and removal of such nevus is of utmost importance in the prevention of melanoma. If such hypothesis is true, the use of border-line features might prove to be useful in providing additional discriminatory information that will help to improve the classification accuracy of skin lesions. To test and validate the raised hypothesis, three classifiers were used: two of them are based on a linear Support Vector Machine (SVM) of similar parameters, while the third implements a Feedforward Neural Network (FNN) for classification. Deep Learning classifiers were not selected for this study due to the selected segmentation algorithms variable length outputs and the datasets' size constraints. The experiments were made in a MSI GT683DXR-423US laptop, which provides an Intel® Core™i7-2670QM CPU @ 2.20GHz with 8GB of RAM.

The first classifier, namely the SMO, employs an SVM classifier using Sequential Minimal Optimization. This classifier was proposed in [4] for skin lesions classification and implements a robust supervised learning method with a linear kernel function that is solved iteratively through the sequential minimal optimization. The classifier, imported from Weka 3.8.2, is employed to enable comparison with [4]. In this algorithm, the SVM problem is broken into a series of smaller sub-problems, which are solved analytically [28]. For this method default literature parameters were used: complexity constant of 0.5 and epsilon of 1×10^{-7}.

The second classifier, namely the ISDA, also employs an SVM classifier, however, instead of solving the problem with the sequential minimal optimization, as in the SMO, this version uses the Iterative Single Data Algorithm proposed in [29]. For this classifier, an existing implementation present in Matlab™R2018b was used. Unlike SMO, ISDA solves a series of one-point minimisation that does not respect the linear constraint and does not explicitly include the bias term in the model. The ISDA implementation uses the same parameters, as in the SMO.

The third classifier, referred to as FFN, is a Feed Forward Network (present in Matlab™R2018b *patternnet* function) that was implemented based on a common *rule of thumb*, which states that the number of neurons n in a network should be determined taking into consideration the number of samples, features (inputs) and possible classifications (outputs), expressed by $n = (\#\text{sample} * (\#\text{inputs} + \#\text{output}))/w$ where, the weight w was set to 2, halving the result, in order to force lower overfitting probability, as it limits the networks' number of degrees of freedom. In this network, the traditional sigmoid activation function [30] was employed. This network was trained using Scaled Conjugate Gradient Backpropagation [31] with cross-entropy as the network performance measurement and no normalisation or regularisation for simplicity. The FFN classifier was included in this experiment because it was shown to be a universal approximator and could thus provide better results [32]. The following default literature parameters were used in this method: 5×10^{-5} for derivative approximation (sigma), 5×10^{-7} for the indefiniteness of the Hessian (lambda), a minimum performance gradient of 1×10^{-6}, and maximum six validation fails.

For the tests, the SVM classifiers were trained using 90% of the available data and tested on the remaining, unseen, 10% of the data. For the FNN, the network was trained on 70% of the data, validated on untrained 20% (to prevent overfitting), and later tested using the remaining 10% of the data. The results obtained in the different classifiers are presented in terms of the average of all tests' that have been repeated 10 times using 10-fold cross-validation (100 executions). Training and test proportions of 70–30% and 50–50% were also considered, presenting similar results.

4. Datasets

This section describes the two utilised datasets, namely the MED-NODE dataset [33] and the Dermofit dataset [26]. Both are used in this experiment since they provide different acquisition methods and constraints.

4.1. MED-NODE

The MED-NODE dataset consists of 70 melanoma and 100 nevus images from the digital image archive of the Department of Dermatology of the University Medical Centre Groningen. The images were acquired with a Nikon D3 or Nikon D1x body and a Nikkor 2.8/105 mm micro lens, with an average distance of 33 cm between the lens and the targeted lesion. Images of pigmented skin lesions originate only from patients of Caucasian origin (majority of the population in the Netherlands). Prior to the dataset release, scale and size (along with other operations like hair removal) were performed manually. These operations are lesion-region dependent since some manual pre-processing was performed to set the images' range from 349×321 to 1880×1867 pixels. Each image shows a single region of interest that contains both healthy skin and lesion, and associated classification label.

4.2. Dermofit

The Dermofit dataset is comprised of images with high quality focus, collected under similar conditions by the Department of Dermatology of the University of Edinburgh. Although the lesions' diagnostics in this dataset span across 10 different classes, only melanoma and nevus are of interest for this research. Using this criterion, 407 images were selected, obtaining an unbalanced setting of 331 nevi and 76 melanomas. These images were acquired using a Canon EOS 350D SLR camera at an average distance of 50 cm. Their size range from 209×169 to 2176×2549 pixels. Along with each image of the lesion surrounded by some healthy skin, the dataset provides a ground-truth binary segmentation mask of the lesion area (GT) and a classification label. Images from the dataset were used after a pre-processing stage that aimed at removing hair from the images, using the algorithm described in [15].

5. Experimental results and discussion

The skin lesion images used in the experiments were taken from two publicly available datasets, namely the MED-NODE and the Dermofit as described in Section 4. For each dataset, the image data was evaluated by using the three classifiers described in Section 3.3. In each case, two feature sets were employed. Firstly, 10 features (*F23–32*) proposed in [4] are used as input to the classifiers (five of which assess the lesions' asymmetry aspects, one assesses border condition and four consider lesions' colour attributes). Later, in order to assess the contribution of the detailed border-line information to the classifiers performance, the remaining 22 features (*F1–22*) described in Section 3.2 are used as input to the classifiers, thus resulting in a total of 32 features. The results obtained in these assessments are expressed in terms of percentage of classification accuracy (*Acc.*), Specificity (*SP*), and Sensitivity (*SE*). This experiment, as described, is estimated to take two and half hours to execute – including segmentation and feature extraction of all images in the dataset, and later classification of this data using the four different sets of features for the three classifiers.

Table 1
Results for the MED-NODE dataset (×100).

Seg.	Ft. (#)	SVM-SMO			SVM-ISDA			FFN		
		Acc.	SE	SP	Acc.	SE	SP	Acc.	SE	SP
GHT	*F23–F32* (10)	73 ± 1.5	45 ± 3.7	92 ± 3.0	76 ± 1.2	66 ± 2.4	83 ± 1.4	**76** ± 1.9	63 ± 4.6	84 ± 2.1
	F1–F32 (32)	**74** ± 1.2	56 ± 6.4	86 ± 6.2	**78** ± 2.0	66 ± 2.4	86 ± 1.3	**76** ± 2.4	63 ± 4.7	84 ± 2.7
LBPC	*F23-F32* (10)	75 ± 1.2	49 ± 3.0	93 ± 2.0	77 ± 1.3	69 ± 2.2	83 ± 0.9	75 ± 1.7	64 ± 4.1	83 ± 1.7
	F1–F32 (32)	**78** ± 1.3	58 ± 5.6	91 ± 3.4	**79** ± 1.5	65 ± 2.7	88 ± 1.1	**77** ± 1.9	66 ± 5.8	86 ± 2.2

Table 2
Results for the Dermofit Dataset (×100).

Seg.	Ft. (#)	SVM-SMO			SVM-ISDA			FFN		
		Acc.	SE	SP	Acc.	SE	SP	Acc.	SE	SP
GT	*F23–F32* (10)	82 ± 0.1	3 ± 1.4	100 ± 0.4	83 ± 0.6	17 ± 2.5	98 ± 0.3	84 ± 1.9	38 ± 5.6	95 ± 1.5
	F1–F32 (32)	**88** ± 0.9	49 ± 4.2	96 ± 1.1	**89** ± 0.5	59 ± 1.2	96 ± 0.5	**88** ± 0.8	51 ± 3.9	96 ± 0.8
GHT	*F23–F32* (10)	83 ± 0.2	9 ± 1.3	100 ± 0.3	88 ± 0.3	52 ± 1.6	97 ± 0.3	86 ± 0.8	49 ± 2.8	96 ± 0.9
	F1–F32 (32)	**88** ± 0.4	40 ± 3.9	99 ± 1.3	**90** ± 0.4	56 ± 2.1	98 ± 0.2	**88** ± 0.9	54 ± 2.9	96 ± 0.8
LBPC	*F23–F32* (10)	81 ± 0.7	5 ± 5.4	99 ± 0.8	84 ± 0.8	27 ± 1.9	97 ± 0.6	86 ± 1.1	50 ± 5.2	95 ± 1.2
	F1–F32 (32)	**87** ± 0.6	43 ± 3.8	97 ± 1.0	**89** ± 0.5	64 ± 1.7	95 ± 0.5	**91** ± 0.8	68 ± 3.5	96 ± 0.6

5.1. MED-NODE dataset

The experimental results for the MED-NODE dataset can be seen in Table 1.

When using the GHT segmentation, adding the proposed border-line features led to limited improvements of 1% and 2% on SMO and ISDA, respectively; and no improvements whatsoever to the FFN classifier. With these results the best classification results correspond to ISDA, which achieves 78% accuracy, followed by the FFN with 76% accuracy. In this scenario, the main problem faced by classification algorithms is to wrongly classify the melanoma samples as nevus, which leads to poor sensitivity results. It is, however, worth noting that with SMO the inclusion of the proposed border-line features have substantially increased the sensitivity by 11%.

The obtained results show that using border-line features extracted from LBPC-based segmentations consistently leads to better results in all classification methods. This is likely to be associated with the dense local texture information provided by LBPs localised detail which, in this case, led to improvements of 3% for the SMO and 2% for the remaining classifiers. Moreover, the sensitivity also increased 9% with the SMO, showing that for these classifiers the addition of the border-line features to the commonly used features helps solving the main issues in the classification, as discussed in the previous paragraph.

5.2. Dermofit dataset

The Dermofit dataset poses a different challenge to the classifiers due to the unbalanced dataset, despite providing more data for training. As previously mentioned, the dataset was classified using the proposed approach by testing each of the 2 proposed segmentation methods (GHT, LBPC) plus the provided segmentation Ground-Truth (GT). The features were extracted from each segmented image using both methods. Then, they were tested separately with three classifiers (SMO, ISDA, and FFN) to perform the lesion classification, achieving the results depicted in Table 2.

When using the provided GT segmentation, the additional border-line-based features led to accuracy enhancements of 6%, 6%, and 4% on SMO, ISDA, and FFN, respectively. The best performance was obtained with the ISDA classifier (89% accuracy). It is also relevant to notice the significant gains in sensitivity (46%, 42%, and 13%), which indicates that adding the new features helps the classifiers to better cope with the class imbalance.

Concerning GHT-based segmentations, the first observation is that it yields better results than GT even with only the initial features, which can be seen as an indication of a higher quality lesion segmentation. With the additional border-line-based features, gains of 5%, 2%, and 2% were now observed in accuracy, with ISDA reaching a top score level of 90%. As with GT, more significant increases were observed in the sensitivity (31%, 4%, and 5%), showing that, again, the new border-line features make the classifiers better prepared to handle the dataset class imbalance.

Finally, the LBPC segmentation method presents results similar to those previously discussed. In this case, gains of 6%, 5%, and 5% were obtained in the accuracy, and of 38%, 36%, and 18% in sensitivity, with FFN outperforming the other two classifiers.

On a global analysis, GHT features (with the ISDA classifier) and LBPC-based features (with the FFN classifier) led to best results in terms of accuracy (90% and 91%), but with the latter exhibiting a better sensitivity (68% against 56%), indicating that it better handles the dataset class imbalance problem previously described.

5.3. Feature statistics

To validate the previous results, this subsection presents statistical information about the discussed material. While Section 5.3.1 provides insight about each feature's usefulness, in Section 5.3.2 the maintenance of all features is justified, instead of using only the most useful.

To improve the presentation of the results, only the Dermofit dataset and the SMO classifier are considered. The dataset was chosen due to its large class unbalance. The choice of the classifier is due to its low computational constraints.

5.3.1. Feature information

Three separate feature selection algorithms were used due to their different capabilities to evaluate the worthiness of an attribute: one to measure the correlation (Pearson's) between it and the class, dubbed Correlation; one to measure the information gain with respect to the class, dubbed InfoGain; and another dubbed ReliefF [34–36] that evaluates by repeatedly sampling an instance and considering the value of the given attribute for the nearest instance of the same and different class. All the above were executed using their default literature parameters. Table 3 shows the results resorted by the algorithms' metric. All experiments were done with 10-fold cross-validation.

Table 3
Feature Evaluation using 3 algorithm metrics. Numbers in Attr column refer to the features proposed in Section 3.2

Correlation		InfoGain		Relief	
Merit	Attr	Merit	Attr	Merit	Attr
26.6 ± 3.0	1	34.0 ± 1.3	1	10.2 ± 0.4	1
25.2 ± 3.0	2	31.6 ± 2.0	2	9.3 ± 0.5	2
22.1 ± 1.0	6	17.7 ± 2.0	3	3.9 ± 0.3	12
12.4 ± 1.7	15	7.0 ± 1.6	19	3.6 ± 0.4	4
12.3 ± 2.0	16	7.5 ± 3.8	5	2.7 ± 0.7	6
12.2 ± 2.4	3	6.8 ± 0.7	6	2.5 ± 0.4	3
12.0 ± 2.0	17	8.7 ± 4.9	12	2.1 ± 0.1	20
11.9 ± 2.0	14	4.6 ± 2.7	4	2.1 ± 0.1	22
11.0 ± 1.7	13	5.2 ± 0.7	21	2.1 ± 0.1	21
8.7 ± 1.7	12	5.2 ± 0.6	18	2.1 ± 0.2	19
5.9 ± 2.4	18	5.2 ± 0.6	22	1.7 ± 0.2	18
7.5 ± 1.5	4	4.9 ± 0.6	20	1.2 ± 0.1	10
6.1 ± 2.6	19	0.0 ± 0.0	17	1.1 ± 0.3	11
6.2 ± 1.7	11	0.0 ± 0.0	13	0.9 ± 0.3	7
5.2 ± 2.6	20	0.0 ± 0.0	16	0.8 ± 0.1	8
4.9 ± 2.5	22	0.0 ± 0.0	14	0.9 ± 0.3	5
3.5 ± 1.0	7	0.0 ± 0.0	15	0.7 ± 0.2	9
4.4 ± 2.5	21	0.0 ± 0.0	7	0.6 ± 0.1	15
3.1 ± 1.3	10	0.0 ± 0.0	8	0.6 ± 0.1	17
3.1 ± 1.4	8	0.0 ± 0.0	10	0.6 ± 0.1	16
2.7 ± 1.2	9	0.0 ± 0.0	9	0.6 ± 0.1	14
1.9 ± 1.3	5	0.0 ± 0.0	11	0.5 ± 0.1	13

In the presented results, it is clear that features *F1* and *F2* are the most relevant. This may be due to their indirect ability to provide a proportionality of the lesions' average size. Then *F6* and *F3* are the next dominant features across the three feature selection algorithms, again providing information about lesion dimension. Apart from these, the following most significant features belong to the magnitude of the six peaks of the DCT and the five periodogram PSD.

By the magnitudes provided by the algorithm metrics, many features seem to provide very little information. Specifically with InfoGain, there are 10 features that seem to be completely useless. But this is not case. If removed, they have a strong negative combined impact in the classifier.

5.3.2. Feature selection

As mentioned in the previous subsection, there are many features that provide very little information or correlation apart from the first two. Nevertheless, including all the other features moves the sensitivity up by 18.4% and specificity down by only 0.9%. This, overall, improves the accuracy by 2.7%, which is why all features where used in this work. Fig. 5 provides more detail on the feature contribution for the SMO classifier. The initial x-label '*i*' denotes the use of the previously mentioned 10 base-features from [4]. Afterwards, at each step along the *x*-axis, each feature (*F*) denoted in the *x*-label is incrementally included in the dataset to plot the corresponding Accuracy, Specificity and Sensitivity *y*-axis values.

Table 4
Classification Significance Results

Seg.	Metric	*H0*		*H1*		T
		≥0	*p*-value	≤0	*p*-value	
GT	Acc.	Rejected	0.0000	Not rejected	1.0000	v
	SE	Rejected	0.0000	Not rejected	1.0000	v
	SP	Not rejected	1.0000	Rejected	0.0000	*
GHT	Acc.	Rejected	0.0000	Not rejected	1.0000	v
	SE	Rejected	0.0000	Not rejected	1.0000	v
	SP	Not rejected	0.9873	Rejected	0.0127	*
LBP	Acc.	Rejected	0.0000	Not rejected	1.0000	v
	SE	Rejected	0.0000	Not rejected	1.0000	v
	SP	Not rejected	0.9935	Rejected	0.0065	*

5.4. Classification significance

This section focuses on the classifier's statistical information, using the features obtained in the previous section. As previously mentioned, only the Dermofit dataset and the SMO classifier are considered for the same reasons.

The results provided by the classifiers were validated with a corrected paired *t*-test [37] in order to assess if the obtained results with 32 features (32F) are significantly better than the previous 10 features (10F), using a significance $\alpha = 0.05$. Two hypothesis are tested: *H0)* verifies if the 32F are significantly worse than 10F; and *H1)* verifies if the 32F are significantly better than 10F. If both null-hypothesis are confirmed then it means 32F and 10F are equal. Table 4 shows the overall results for the SMO classifier over the three used metrics. The conclusion for the corrected paired *t*-test is given in column T. The annotation indicates whether a specific result is statistically better (v) or worse (*) than the baseline scheme (10F). Note that they are never statistically equals.

With the presented results it is possible to state that the results obtained with the 32F are significantly better in terms of accuracy and sensitivity but worse (even if slightly) in terms of specificity, as expected from the previous experiments.

6. Conclusions

This work offers contributions to improve the automatic classification of melanocytic skin lesions (namely Nevus vs Melanoma), showing the importance of the lesion border information. Two image segmentation methods are exploited to provide the lesions contours, namely the GHT and the proposed LBPC, from which the border-line features are extracted.

The achieved results confirm that segmentation accuracy contributes to enhance the classification performance, namely in

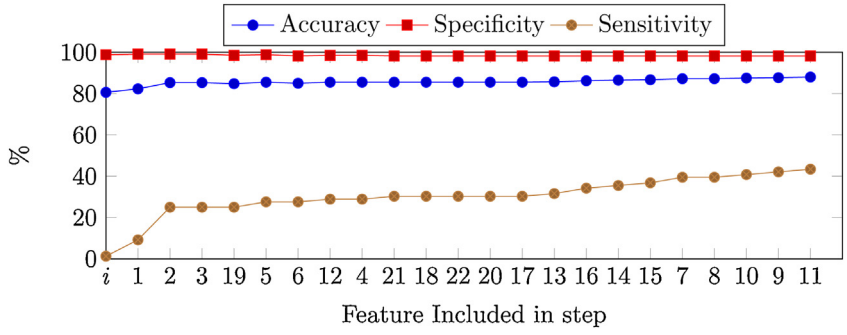

Fig. 5. Feature inclusion plot with Accuracy, Specificity and Sensitivity metrics for the Dermofit dataset and the SMO classifier.

methods based on GHT and LBPs, which clearly outperform the GT segmentation provided for the Dermofit dataset. The results obtained with the three considered classifiers confirm that adding border-line lesion features does indeed contribute to improve the performance of automatic classification algorithms.

Moreover, the use of finer segmentation algorithms such as GHT and LBPC was found to be particularly suited for this approach. In fact, the features extracted from their spatially detailed border-lines segmentation improved the classification performance by figures above the gains obtained with the coarser GT segmentation line provided for the Dermofit dataset.

It is shown that using border-line based features together with other commonly used sets can lead to classification results with accuracy above 90% in the tested datasets. Additionally, it was shown that these features improve the sensitivity, which is important when dealing with class imbalanced datasets, as commonly occurs with medical image datasets. Hence, future endeavours might include these types of features to compensate for class imbalance and improve the classification results in general.

Acknowledgements

This work was supported by the Fundação para a Ciência e Tecnologia, Portugal, under PhD Grant SFRH/BD/128669/2017 and project PlenoISLA PTDC/EEI-TEL/28325/2017, and Instituto de Telecomunicacoes project UID/EEA/50008/2019, through national funds and where applicable co-funded by FEDER – PT2020.

Declaration of Competing Interest

The authors declare no conflicts of interest.

References

[1] N.B. Linsangan, J.J. Adtoon, J.L. Torres, Geometric analysis of skin lesion for skin cancer using image processing, in: 2018 IEEE 10th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM), Philippines, 2018, pp. 1–5.
[2] A. Namozov, Y.I. Cho, Convolutional neural network algorithm with parameterized activation function for melanoma classification, in: 2018 International Conference on Information and Communication Technology Convergence (ICTC), Jeju Island, Korea, 2018, pp. 417–419.
[3] S.A. Mahdiraji, Y. Baleghi, S.M. Sakhaei, Bibs, a new descriptor for melanoma/non-melanoma discrimination, in: Iranian Conference on Electrical Engineering (ICEE), Iran, Mashhad, 2018, pp. 1397–1402.
[4] M.H. Jafari, S. Samavi, N. Karimi, S.M.R. Soroushmehr, K. Ward, K. Najarian, Automatic detection of melanoma using broad extraction of features from digital images, in: International Conference of the IEEE Engineering in Medicine and Biology Society, Orlando, USA, 2016, pp. 1357–1360.
[5] P.M.M. Pereira, L.M.N. Tavora, R. Fonseca-Pinto, R.P. Paiva, P.A.A. Assuncao, S.M.M. Faria, Image segmentation using gradient-based histogram thresholding for skin lesion delineation, International Conference on BioImaging (2019).
[6] P.M.M. Pereira, R. Fonseca-Pinto, R.P. Paiva, L.M.N. Tavora, P.A.A. Assuncao, S.M.M. de Faria, Accurate segmentation of dermoscopic images based on local binary pattern clustering, 2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO) (2019) 314–319, http://dx.doi.org/10.23919/MIPRO.2019.8757023.
[7] L.A. Goldsmith, F.B. Askin, A.E. Chang, et al., Diagnosis and treatment of early melanoma: NIH consensus development panel on early melanoma, JAMA 268 (10) (1992) 1314–1319.
[8] G. Day, R. Barbour, Automated skin lesion screening – a new approach, Melanoma Res. 11 (1) (2001) 31–35.
[9] E. Claridge, A. Orun, Modelling of edge profiles in pigmented skin lesions, in: Proceedings of Medical Image Understanding and Analysis, Portsmouth, United Kingdom, 2002, pp. 53–56.
[10] I. Cheng, X. Sun, N. Alsufyani, Z. Xiong, P. Major, A. Basu, Ground truth delineation for medical image segmentation based on local consistency and distribution map analysis, in: International Conference of the IEEE Engineering in Medicine and Biology Society, Milan, Italy, 2015, pp. 3073–3076.
[11] R. Kéchichian, H. Gong, M. Revenu, O. Lezoray, M. Desvignes, New data model for graph-cut segmentation: Application to automatic melanoma delineation, IEEE International Conference on Image Processing (2014) 892–896.
[12] R. Fonseca-Pinto, M. Machado, A textured scale-based approach to melanocytic skin lesions in dermoscopy, in: International Convention on Information and Communication Technology, Electronics and Microelectronics, Opatija, Croatia, 2017, pp. 279–282.
[13] I. Pirnog, R.O. Preda, C. Oprea, C. Paleologu, Automatic lesion segmentation for melanoma diagnostics in macroscopic images, in: European Signal Processing Conference, Nice, France, 2015, pp. 659–663.
[14] V. Zheludev, I. Pölönen, N. Neittaanmäki-Perttu, A. Averbuch, P. Neittaanmäki, M. Grönroos, H. Saari, Delineation of malignant skin tumors by hyperspectral imaging using diffusion maps dimensionality reduction, Biomed. Signal Process. Control 16 (2015) 48–60, http://dx.doi.org/10.1016/j.bspc.2014.10.010, URL http://www.sciencedirect.com/science/article/pii/S1746809414001608.
[15] J. Koehoorn, A.C. Sobiecki, D. Boda, A. Diaconeasa, S. Doshi, S. Paisey, A. Jalba, A. Telea, Automated digital hair removal by threshold decomposition and morphological analysis, in: International Symposium on Mathematical Morphology and Its Applications to Signal and Image Processing, Reykjavik, Iceland, 2015, pp. 15–26.
[16] Q. Abbas, M. Celebi, I.F. García, Hair removal methods: A comparative study for dermoscopy images, Biomed. Signal Process. Control 6 (4) (2011) 395–404, http://dx.doi.org/10.1016/j.bspc.2011.01.003, URL http://www.sciencedirect.com/science/article/pii/S1746809411000048.
[17] S. Pathan, K.G. Prabhu, P. Siddalingaswamy, Techniques and algorithms for computer aided diagnosis of pigmented skin lesions – a review, Biomed. Signal Process. Control 39 (2018) 237–262, http://dx.doi.org/10.1016/j.bspc.2017.07.010, URL http://www.sciencedirect.com/science/article/pii/S1746809417301428.
[18] S. Mane, S. Shinde, A method for melanoma skin cancer detection using dermoscopy images, in: 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), Maharashtra, India, 2018, pp. 1–6.
[19] N. Hameed, A. Shabut, M.A. Hossain, A computer-aided diagnosis system for classifying prominent skin lesions using machine learning, in: 2018 10th Computer Science and Electronic Engineering (CEEC), Colchester, UK, 2018, pp. 186–191.
[20] E.Z. Chen, X. Dong, X. Li, H. Jiang, R. Rong, J. Wu, Lesion attributes segmentation for melanoma detection with multi-task u-net, in: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), Venice, Italy, 2019, pp. 485–488.
[21] S. Majumder, M.A. Ullah, Feature extraction from dermoscopy images for an effective diagnosis of melanoma skin cancer, in: 2018 10th International Conference on Electrical and Computer Engineering (ICECE), Dhaka, Bangladesh, 2018, pp. 185–188.
[22] B. Series, Studio Encoding Parameters of Digital Television for Standard 4:3 and Wide-screen 16:9 Aspect Ratios, Standard, International Telecommunication Union, Geneva CH, 2011, March.
[23] S. Khalid, U. Jamil, K. Saleem, M.U. Akram, W. Manzoor, W. Ahmed, A. Sohail, Segmentation of skin lesion using cohen-daubechies-feauveau biorthogonal wavelet, Springerplus 5 (1) (2016) 1603.
[24] D. Arthur, S. Vassilvitskii, k-means++: the advantages of careful seeding, in: Proceedings of the ACM-SIAM symposium on Discrete algorithms, New Orleans, USA, 2007, pp. 1027–1035.
[25] T. Ojala, M. Pietikä inen, D. Harwood, A comparative study of texture measures with classification based on featured distributions, Pattern Recognit. 29 (1) (1996) 51–59.
[26] L. Ballerini, R.B. Fisher, B. Aldridge, J. Rees, A color and texture based hierarchical k-nn approach to the classification of non-melanoma skin lesions, in: Color Medical Image Analysis, Springer, 2013, pp. 63–86.
[27] F. Auger, P. Flandrin, Improving the readability of time-frequency and time-scale representations by the reassignment method, IEEE Trans. Signal Process. 43 (5) (1995) 1068–1089.
[28] J.C. Platt, Sequential minimal optimization: a fast algorithm for training support vector machines, tech. rep, in: Advances in Kernel Methods – Support Vector Learning, 1998.
[29] V. Kecman, T.-M. Huang, M. Vogt, Iterative Single Data Algorithm for Training Kernel Machines from Huge Data Sets: Theory and Performance, Springer, Berlin, Heidelberg, 2005, pp. 255–274.
[30] G. Cybenko, Approximation by superpositions of a sigmoidal function, Math. Control Signals Syst. 2 (4) (1989) 303–314.
[31] M.F. Møller, A scaled conjugate gradient algorithm for fast supervised learning, Neural Netw. 6 (4) (1993) 525–533.
[32] B.C. Csáji, Approximation with artificial neural networks, MSc thesis, 2001, 24, 48.
[33] I. Giotis, N. Molders, S. Land, M. Biehl, M.F. Jonkman, N. Petkov, Med-node: a computer-assisted melanoma diagnosis system using non-dermoscopic images, Expert Syst. Appl. 42 (19) (2015) 6578–6585.
[34] K. Kira, L.A. Rendell, A practical approach to feature selection, Machine Learning Proceedings 1992 (1992) 249–256.
[35] I. Kononenko, Estimating attributes: analysis and extensions of relief, European Conference on Machine Learning (1994) 171–182.
[36] M. Robnik-Šikonja, I. Kononenko, An adaptation of relief for attribute estimation in regression, Machine Learning: Proceedings of the Fourteenth International Conference (ICML'97) 5 (1997) 296–304.
[37] C. Nadeau, Y. Bengio, Inference for the generalization error, Advances in Neural Information Processing Systems (2000) 307–313.