Skin Lesion Computational Diagnosis of Dermoscopic Images: Ensemble Models based on Input Feature Manipulation

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

Background and Objectives: The number of deaths worldwide due to melanoma has risen in recent times, in part because melanoma is the most aggressive type of skin cancer. Computational systems have been developed to assist dermatologists in early diagnosis of skin cancer, or even to monitor skin lesions. However, there still remains a challenge to improve classifiers for the diagnosis of such skin lesions. The main objective of this article is to evaluate different ensemble classification models based on input feature manipulation to diagnose skin lesions. *Methods:* Input feature manipulation processes are based on feature subset selections from shape properties, colour variation and texture analysis to generate diversity for the ensemble models. Three subset selection models are presented here: 1) a subset selection model based on specific feature groups, 2) a correlation-based subset selection model, and 3) a subset selection model based on feature selection algorithms. Each ensemble classification model is generated using an optimum-path forest classifier and integrated with a majority voting strategy. The proposed models were applied on a set of 1104 dermoscopic images using a cross-validation procedure. *Results:* The best results were obtained by the first ensemble classification model that generates a feature subset ensemble based on specific feature groups. The skin lesion diagnosis computational system achieved 94.3% accuracy, 91.8% sensitivity and 96.7% specificity. *Conclusions:* The input feature manipulation process based on specific feature subsets generated the greatest diversity for the ensemble classification model with very promising results.

Keywords: Image Classification; Feature Extraction; Feature Selection; Ensemble of Classifiers; Computational Diagnosis.

[*] Corresponding author. 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. Tel.: +351 220413472; fax: +351 225081445 (João Manuel R. S. Tavares).

Email addresses: roberta.oliveira@fe.up.pt (Roberta B. Oliveira), aledir@sjrp.unesp.br (Aledir S. Pereira), tavares@fe.up.pt (João Manuel R. S. Tavares).

1. Introduction

Skin cancer is one of the most common cancers worldwide, and its incidence has increased in recent years [1]. Computational diagnosis systems have been developed to assist dermatologists in early diagnosis of skin cancer from dermoscopic images. The search for more efficient classifiers for these computational systems is a challenging task. Several studies have proposed an ensemble of classifiers, commonly known as a multiple classifier system or an ensemble classification model to improve skin lesion classifications from dermoscopic images [2-4]. An ensemble of classifiers consists of integrating several classification models in order to develop a more robust system that provides more accurate results than by using a single classifier [5]. There are different voting methods [6] for integration strategies based on the outputs of the input classifiers for ensemble classification models, e.g., majority voting that counts the votes for each class of all the input classifiers and then designates the class with the majority votes as the classification result. Statistical methods, such as average, sum, product and median can also be used for this same purpose [7], as well as for cases of numeric predictions.

One important requisite for constructing ensembles is to ensure diversity between the classification models, which can be performed by manipulating the modelling process or the input data. Manipulation of the modelling process consists of constructing the classification models by using either different learning algorithms or a single learning algorithm but with different parameters. The more popular approaches for input data manipulation are to manipulate the training samples and the input features. Algorithms used to manipulate the training samples can generate multiple hypotheses, in which a learning algorithm is applied to different subsets of the training samples. Bagging and boosting algorithms are the traditional ways to manipulate the training samples [5], and their hypotheses are integrated by a vote method. The bagging algorithm consists of randomly splitting the original dataset in several training subsets of the same size based on sampling with replacement, which can be applied to any learning algorithm. Likewise, the boosting algorithm combines the classification outputs using the same learning algorithm; however, this type of algorithm is iterative, where each new model is based on the result of the previously built one.

Algorithms for manipulating the input features generate ensembles based on different feature subsets available to the learning algorithm. This process can be, for example, the random splitting of a set of features into subsets [8], or by using a feature selection algorithm combined with manipulation of the training samples [4]. One challenge that affects the performance of classifiers is how to define which features are meaningful to describe the patterns of interest. Consequently, feature selection algorithms [9] can be used for the ensemble construction in order to achieve superior performance for skin lesion classifications.

This article presents ensemble classification models based on input feature manipulation to improve skin lesion computational diagnosis from dermoscopic images. Two examples of pigmented skin lesions in dermoscopic images are shown in Fig. 1. The main contributions of this study are the feature subset selection models based on specific feature groups and the feature selection algorithms for the input feature manipulation. To the best of our knowledge, few studies based on ensemble models and feature manipulation for skin lesion classification have been presented with successful results [10, 11].

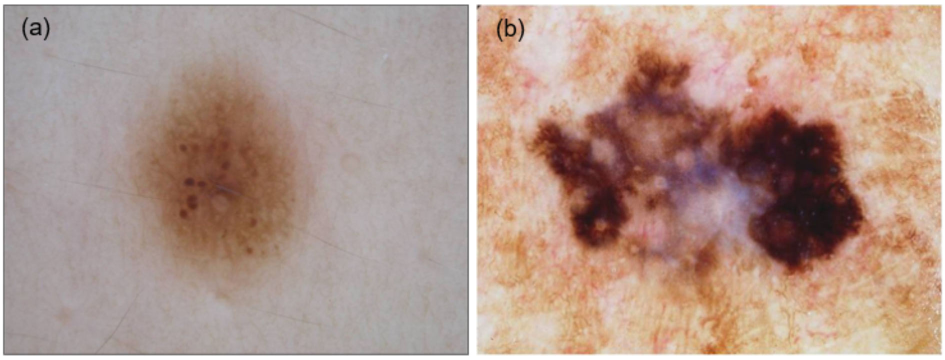

Fig. 1 - Two examples of pigmented skin lesions: (a) benign lesion and (b) malignant lesion.

This article is organized as follows: Studies relating to the ensemble methods for skin lesion classification are discussed in Section 2. The proposed ensemble classification models based on input feature manipulation are presented in Section 3. The experimental results and their discussion, which include the evaluation process, feature subset and feature selection evaluations, ensemble classification model evaluation and comparison between the classification algorithms used are given in Section 4. Finally, the conclusions drawn for the proposed ensemble classification models and future works about the skin lesion classification are pointed out in Section 5.

2. Related studies

An overview of computational methods for pigmented skin lesion classification in images, which addresses the feature extraction and selection, and classification steps, is presented in Oliveira, et al. [12]. The ensemble of classifiers based on input data manipulation has been recently adopted for skin lesion classification to achieve better results than single classifiers. Several algorithms can be used for constructing ensembles; e.g., the AdaBoost [13], which is a popular boosting algorithm that maintains a set of weighting systems for the training samples according to a computed error rate. In Barata, et al. [2], the proposed classification system using AdaBoost obtained the best results by using colour features and with combinations of two to five base classifiers for the detection of melanomas and nevi.

Random forest [14] is another ensemble algorithm used for skin lesion computational diagnosis. This algorithm is a variation of the bagging algorithm that is used to create an ensemble of decision trees that ensure the diversity by using a random selection of features to split each tree node. Its error rates are comparable to AdaBoost, but are more robust with respect to noise. Rastgoo, et al. [15] proposed an automatic system to differentiate melanoma from dysplastic nevi by using texture features and random forest. Barata, et al. [10] built a system for melanoma detection using the random forest algorithm based on the global and local feature fusion of colour and texture properties.

The random forest algorithm also obtained promising results in a system proposed by Garnavi, et al. [16]. The authors developed an optimized selection and integration of features derived from texture, border and geometrical properties. Rastgoo, et al. [11] proposed an automatic framework based on ensemble methods to differentiate melanoma from dysplastic and benign lesions. This framework used a random forest algorithm and a combination of colour and texture features based on global features. Maragoudakis and Maglogiannis [17] presented a novel ensemble classification algorithm for skin lesion diagnosis. The authors combined random forests with the Markov blanket notion to perform an inherent feature selection process in order to obtain more informative features. Using 32 features based on border, colour and texture properties, the classification result using a dataset of 1041 skin lesion images was increased from 4.5% to 6% in comparison with the traditional random forest, support vector machine (SVM) and k-nearest neighbour (KNN) algorithms, which were also combined with standard feature reduction techniques, namely, principal component analysis (PCA) and singular value decomposition (SVD).

Other ensemble classification models have also been proposed for skin lesion classification. In Abedini, et al. [18], an ensemble model, based on feature random subsets, a linear SVM classifier and forward model selection for the ensemble fusion, was proposed. The best results were obtained by concatenating the pattern prediction values, which are considered middle-level features. Schaefer, et al. [4] proposed a multiple classifier system to deal with imbalanced classes. Such a system consists of a random under-sampling method, an SVM using a polynomial kernel, and a neural network for the classifier fusion. In addition, a feature selection algorithm is applied to each classifier, and a diversity measure is used for pruning a pool of classifiers. The authors used features based on shape, colour and texture properties for the melanoma and benign lesion classification.

The ensemble of classifiers based on model manipulation process has also been adopted for skin lesion classification, which consists of constructing the multiple classification models by using different learning algorithms, or a single learning algorithm but with different parameters or structures. In Sboner, et al. [19], a novel multiple classifier system for the early diagnosis of melanoma was proposed based on the combination of different classification algorithms, which demonstrated a superior performance relatively to the use of each classifiers alone. The proposed system combines three different types of classifiers, namely, linear discriminant analysis (LDA), C4.5

decision tree and kNN classifiers, and uses 38 geometric and colourimetric features as input for the classifiers, and a voting scheme to combine the outputs of each classifiers. The performance of the proposed system was compared against the performances of each classifier when used alone and also relatively to the performances of eight dermatologists. The system achieved a performance that was significantly higher than the ones achieved by each classifier, and a performance comparable to the dermatologists.

A novel meta-ensemble model based on multiple neural network ensembles was proposed by Xie, et al. [20]. The authors used 57 features based on colour and texture properties from two lesion regions obtained by a combination of the self-generating neural network (SGNN) method and manual interaction followed by Otsu's threshold. In addition, the authors proposed novel lesion border features so that the model would be insensitive to the incompleteness of the lesion regions. The PCA technique was used to reduce the feature dimensions and to define the best feature subset. The meta-ensemble model is composed of three ensembles with different structures and network types. The model combines back-propagation (BP) neural networks with fuzzy-neural networks (FNNs) to increase individual net diversity. The standard boosting method was used to generate individual nets, and the voting and averaging methods were designed to combine the multiple outputs. The authors used two dermoscopy datasets to perform the experiments: a dataset that includes 240 images of the xanthous race and a dataset with 360 images of the caucasian race.

3. Description of the proposed ensemble classification models

In this section, the ensemble classification models based on input feature manipulation for skin lesion computational diagnoses, as well as the dermoscopic image dataset used are presented. Fig. 2 gives an overview of three different models developed for the input feature manipulation in order to generate diversity for the ensembles of classifiers. Given a dataset $T = \{\boldsymbol{x}_p, y_p\}$, with $p = 1,2 \dots, n$, according to the number of images n, where $\boldsymbol{x}_p$ is a sample, and y_p is the class to which it belongs. Each sample $\boldsymbol{x}_p$ is composed of a set of features F_{pq}, where $q = 1,2 \dots, m$, and m is the number of features. An ensemble $P = \{C_1, C_2, \dots, C_E\}$, with $i = 1,2, \dots, E$, and E is the ensemble size, where C_i $(i \in \{1,2, \dots, E\})$ is composed of the classification models obtained with the input feature manipulation, a base classifier using optimum-path forest (OPF) [21] and an integration strategy. One classification model is obtained in each iteration i by a subset of feature S_i $(i \in \{1,2, \dots, E\})$ that is sampled from F_{pq} based on specific feature groups or with a feature selection algorithm (Figs. 2a and 2b, respectively). The classification models are also obtained by applying several feature selection algorithms A_i $(i \in \{1,2, \dots, E\})$ from F_{pq} (Fig. 2c).

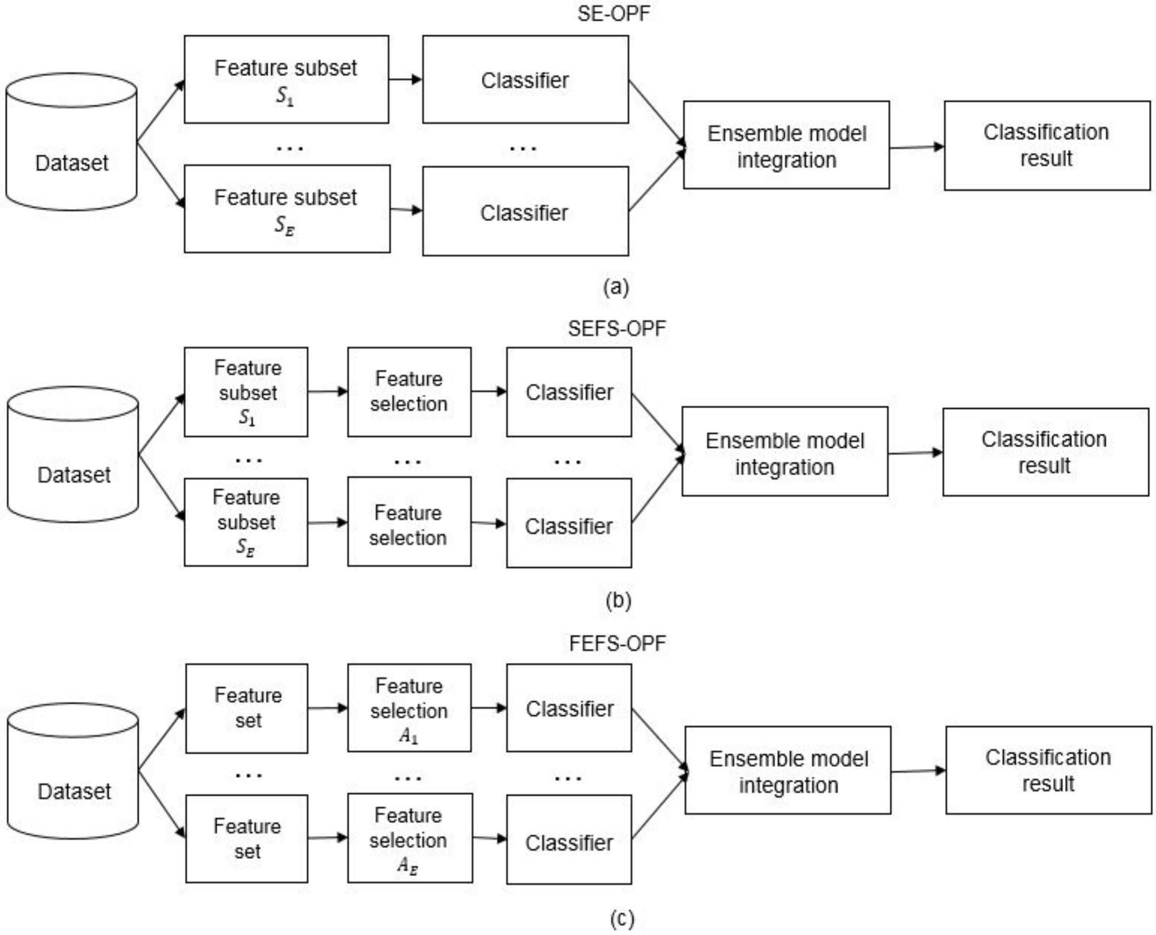

Fig. 2: Overview of the proposed ensemble classification models based on input feature manipulation for the skin lesion computational diagnosis: (a) feature subset ensemble (SE-OPF), (b) feature subset ensemble with a feature selection algorithm (SEFS-OPF), and (c) feature set ensemble with feature selection algorithms (FEFS-OPF).

3.1. Dermoscopic image dataset

The dermoscopic image dataset used in the experiments is composed of pigmented skin lesions, which were collected from the International Skin Imaging Collaboration (ISIC) dataset [22]. In addition, the 1279 images are paired with an 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, the extracted features from the images are based on shape properties, colour variation and texture analysis. The images in which the lesion did not fully fit within the image frame (174 images identified in the Appendix) were removed from the original dataset, since the shape properties are obtained from the lesion borders. Thus, in the end, a total of 1104 images were used from the original dataset. Of these, 916 images were benign lesions and 188 images were malignant lesions. The images of the dataset were proportionally resized to an average resolution of 400×299 pixels to simplify their processing.

3.2. Feature extraction and data pre-processing

The feature extraction process is based on the intensities of the pixels belonging to the binary masks defined by specialists, in which the non-zero pixels belong to the lesion, and the others to the background skin. A combination of features, based on shape properties, colour variation and texture analysis using different feature extraction methods, were used in this study. A total of 512 features were extracted for each skin lesion image. Of these, 18 features were related to the shape properties, 72 features to colour variation, and 420 features to the texture analysis.

a) Shape properties: shape measures are computed based on the geometrical properties, lesion asymmetry and border irregularity. To assess the geometrical properties of the lesion, the area, perimeter, equivalent diameter, compactness, circularity, solidity, rectangularity, aspect ratio and eccentricity [23-25] were computed. To assess the lesion asymmetry, three features were computed from the lesion, i.e., the average, variance and standard deviation. These features were obtained from the ratios between the shortest and longest distances of each pair of the semi-lines that represent the perpendicular lines by overlapping the two sub-regions of the lesion along an axis [26]. To assess the border irregularity, a number of peaks, valleys and straight lines of the border were computed using the vector product and inflexion point descriptors based on low and high irregularities of the border from a one-dimensional border [26].
b) Colour variation: the *RGB*, *HSV*, *CIE Lab* and *CIE Luv* colour spaces [27] were used to analyse the colour variation of the skin lesions. The *RGB* colour space is commonly used and the original *RGB* colour image can be converted to other colour spaces, and several studies have achieved good results from this colour space [23, 28]. The *HSV*, *CIE Lab* and *CIE Luv* colour spaces represent colours based on human perception. Furthermore, *CIE Lab* and *CIE Luv* are approximately perceptually uniform colour spaces and can simplify the identification of colour properties, as it is easy to maintain colour-difference ratios [29]. Six statistical measures, i.e., average, variance, standard deviation, minimum and maximum colours, and colour skewness, are computed for each colour channel in the region of the lesion using the aforementioned four-colour spaces that correspond to 12 channels.
c) Texture analysis: three different texture analysis methods were adopted to obtain the best features to represent the skin lesion texture based on colour images; namely, fractal dimension analysis [30], discrete wavelet transform (DWT) [31] and co-occurrence matrix [32]. The *RGB*, *HSV*, *CIE Lab* and *CIE Luv* colour spaces were also used for the texture analysis. The bi-dimensional fractal dimension using a box-counting method [30] is computed individually for each channel of the colour spaces. The energy and entropy measures from the coefficients obtained by DWT are computed for each of the 10 Haar wavelet sub-bands obtained by a-three-level decomposition, as well as for each channel of the colour spaces. Co-occurrence matrices were obtained for each channel of the colour spaces, and the intensities of each channel were quantized with 16 intensity levels. The distance between each reference pixel and its neighbours was one pixel, and four orientations $\theta = (0^{\circ}, 45^{\circ}, 90^{\circ}, 135^{\circ})$ were used. A normalized matrix was obtained from the matrices corresponding to the four orientations. From the normalized co-occurrence matrix, 14

statistical measures based on Haralick's texture features [32] were extracted from the image. These measures are the angular second moment, contrast, correlation, variance, inverse difference moment, sum average, sum variance, sum entropy, entropy, variance difference, entropy difference, information measure of correlation 1, information measure of correlation 2, and the maximal correlation coefficient. Therefore, 12 features were extracted from the fractal dimension analysis, 240 features were extracted from the discrete wavelet transform, and 168 features were extracted from the co-occurrence matrix.

As the values of the dataset obtained by feature extraction contain different ranges they were normalized into the same interval [0,1] for the skin lesion classification process. The normalization procedure scales all numeric values in the dataset by computing:

$$xn_{pq} = \frac{x_{pq} - \min(x_{pq})}{\max(x_{pq}) - \min(x_{pq})}, \qquad (1)$$

where $p = 1,2 \ldots, n$, $q = 1,2 \ldots, m$, n is the number of samples and m is the number of features. Thus, x_{pq} is the actual value of feature q in the sample p, with the minimum and maximum values of features of all the sets of samples, and xn_{pq} is the normalized value of same feature q in the same sample p. In addition, the unbalanced dataset problem is considered in this study, since the dataset is composed of 916 samples of benign lesions and 188 samples of malignant lesions. These unbalanced datasets concerning the number of samples in each class can decrease the accuracy of the evaluation results, since the classification tends to be based on the classes with the largest number of occurrences. Different sampling methods [33] have been used to solve such classification problems [4, 34]. Here, the resampling procedure was applied to the dataset [5]. This procedure produces a random subsample of the dataset using sampling with replacement and the class distribution is made into a uniform distribution.

3.3. Feature selection

The feature selection process aims to find the best feature subsets to generate the ensembles of classifiers. Feature selection algorithms are usually a combination of both search and evaluation methods [9]. Search methods can be applied to select a candidate subset from extracted features of skin lesions, which is evaluated and compared to the previous best subset until a given stopping criterion is reached. In this study, six feature selection algorithms were applied to generate different feature subsets for the ensemble of classifiers; namely, Pearson's correlation coefficient [35], gain ratio-based feature selection (GRFS) [5], information gain-based feature selection [35], relief-F [36], principal-component analysis (PCA) [37] and correlation-based feature selection (CFS) [38]. These algorithms have been commonly used for skin lesion feature selections [12] since they have several advantages, such as computationally efficiency, are simple and fast algorithms, independent evaluation criteria, and have the ability to overcome over-fitting.

All feature selection algorithms mentioned earlier are single-feature evaluators, with the exception of CFS that is a feature subset evaluator. The single-feature evaluators are used with a ranking method, where the features are ranked individually, according to their evaluation, i.e., the most relevant. The number of features to be maintained is previously defined. The feature subset evaluators measure a subset of features and they return a value that is used in the search [5]. In this study, both the greedy stepwise and best first search methods were adopted.

The greedy stepwise method searches feature subsets in either the forward or backward directions in a greedy way. The selection process must stop when the addition or removal of any feature occurs that worsens the outcome of the best-found subset until that moment. The best first method searches the feature subsets by greedy hill-climbing, and the search direction can be forward, backward or bi-direction. 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.

3.4. Base classifier and integration strategy

In this study, the focus is on homogeneous ensemble methods that are built with only one base classifier through input feature manipulation, and the classification model results are combined by an integration strategy. The number of base classifiers used defines the ensemble size. An OPF classifier [21] based on input feature manipulation for a set of training data was used to generate the ensemble classification models in this work.

The OPF classifier has been applied to address 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 samples are nodes of a complete graph; whose arcs are the links of all pairs of nodes. The arcs are weighted by the distances between the feature vectors of their corresponding nodes. The Euclidean $D_E(i,j)$, Chebyshev $D_C(i,j)$ and Manhattan $D_M(i,j)$ distance functions [5] were used to measure the distances between the feature vectors:

$$D_E(i,j) = \sqrt{\sum_{q=1}^{m} \left|x_{iq} - x_{jq}\right|^2}, \tag{2}$$

$$D_C(i,j) = \sum_{q=1}^{m} \left|x_{iq} - x_{jq}\right|, \tag{3}$$

$$D_M(i,j) = \max_{q=\{1,2,..,m\}} \left|x_{iq} - x_{jq}\right|, \tag{4}$$

where x_{iq} is the feature value of a sample i, x_{jq} is the feature value of a sample j, $q = 1,2 \dots, m$, and m is the number of features.

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 multiclass classifications and overlapping between classes, parameter independence and no assumption is based on the shape of the classes. Ensembles of OPF classifiers for reducing the size of the training set using under-sampling were proposed by Ponti Jr and Rossi [39]. The Weka library based on LibOPF was used to set up the OPF classifier [21] as proposed by Amorim, et al. [40].

Applying a good integration method is also important for the performance of the ensemble model. The challenge is how to integrate the results produced by the base classifiers. Here, the majority voting method [6] combines the classification results to generate an ensemble model. This method analyses which class receives the majority votes based on the results of all base classifiers and therefore the ensemble model must have an odd number of classifiers.

3.5. Input feature manipulation for the ensemble classification models

The input feature manipulation process aims to generate diversity for an ensemble classification model with the combination of the best feature subsets for the base classifier. In this section, three different models for the feature manipulation of skin lesions are presented. These models are based on specific feature groups and feature selection algorithms in order to create different feature subsets.

3.5.1. Feature subset selection model based on specific feature groups

The feature type and feature extraction algorithm were taken into account in order to establish the feature subset groups to be analysed. Hence, the extracted features were divided into: shape (18 features), colour (72 features) and texture (420 features) subsets. Also, the texture feature extraction algorithms based on all colour spaces, i.e., fractal texture (12 features), wavelet texture (240 features) and Haralick's texture (168 features), were studied independently. Moreover, the combination of the shape and colour subsets (90 features), the shape and texture subsets (438 features), and the colour and texture subsets (492 features) were evaluated. In addition, the colour feature extraction algorithms for each colour space alone, i.e., *RGB* colour (18 features), *HSV* colour (18 features), *LAB* colour (18 features), and *LUV* colour (18 features) subsets, and the texture feature extraction algorithms for each colour space individually, i.e., *RGB* texture (105 features), *HSV* texture (105 features), *LAB* texture (105 features), and *LUV* texture (105 features) subsets, were explored.

The combination of the colour and texture feature extraction algorithms were also taken into account for each colour space alone, i.e., *RGB* features (123 features), *HSV* features (123 features), *LAB* features (123 features), and *LUV* features (123 features) subsets. The colour and texture feature subsets were also combined with the shape subset for each colour space individually, i.e., shape + *RGB* features (141 features), shape + *HSV* features (141 features), shape + *LAB* features (141 features), and shape + LUV features (141 features) were analysed. Therefore, the specific feature subset groups built for feature manipulation were:

- Group 1: shape, colour, and texture (3 subsets);
- Group 2: fractal texture, wavelet texture, and Haralick's texture (3 subsets);
- Group 3: shape + colour, shape + texture, and colour + texture (3 subsets);
- Group 4: RGB colour, HSV colour, LAB colour, and LUV colour (4 subsets);
- Group 5: RGB texture, HSV texture, LAB texture, and LUV texture (4 subsets);
- Group 6: shape + RGB features, shape + HSV features, shape + LAB features, and shape + LUV features (4 subsets); and
- Group 7: RGB algorithms, HSV algorithms, LAB algorithms, and LUV algorithms (4 subsets).

The effectiveness of the feature groups is also evaluated individually in the experimental results section. The feature subset selection model generates a feature subset ensemble (SE-OPF). Algorithm 1 describes the procedure to set up this ensemble classification model, which was used for the input feature manipulation based on the feature subset groups and was also used by the OPF classifier [21] and majority voting [6].

```
Algorithm 1 SE-OPF
Require:
    Ensemble size $E$, training sample set $T$, feature set $F$, group-based feature subsets $S_i'$ from the
    feature set $F$

Procedure:
for $i = 1$ to $E$ do
    Select one feature subset $S_i$ from $S_i'$
    Train the OPF classifier $C_i$ using $T$ with the selected feature subset $S_i$
end for
for each new sample do
    Compute the majority voting $V$ of all classification models of the ensemble $C_i$
end for
```

3.5.2. Correlation-based feature subset selection model

The correlation-based feature subsets were set up using the feature subset groups discussed in the previous section and the CFS algorithm for the feature selection. The CFS algorithm [38] tries to find a set of features that are highly correlated with the class and have 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. Such an algorithm is adopted for this subset selection model, since experimental results using the OPF classifier [21] showed that this algorithm improved the classification performance more than the other feature selection algorithms.

The correlation-based subset selection model generates a feature subset ensemble with a feature selection algorithm (SEFS-OPF). Algorithm 2 describes the procedure to set up this ensemble model, which was used for feature input manipulation based on feature subset groups and the CFS algorithm, as well as the OPF classifier [21] and majority voting [6].

```
Algorithm 2 SEFS-OPF
Require:
    Ensemble size $E$,  training sample set $T$, feature set $F$, group-based feature subsets $S_i'$ from the
    feature set $F$

Procedure:
for $i = 1$ to $E$ do
    Select one feature subset $S_i$ from $S_i'$
    $FS$ ← Selected features from $S_i$ using the CFS algorithm
    Train the OPF classifier $C_i$ by using $T$ with the selected features $FS$
end for
for each new sample do
    Compute the majority voting $V$ of all classification models of the ensemble $C_i$
end for
```

3.5.3. Subset selection model based on feature selection algorithms

All features discussed in the previous sections were used to generate the feature subsets. The diversity for an ensemble classification model is obtained by using different feature selection algorithms; namely, correlation coefficient [35], GRFS [5], information gain [35], relief-F [36], PCA [37] and CFS [38]. This subset selection model generates a feature set ensemble with feature selection algorithms (FEFS-OPF). Algorithm 3 describes the procedure to set up this ensemble model, which was used for the input feature manipulation based on the feature selection algorithms A_i ($i \in \{1,2, \dots, E\}$), and with the OPF classifier [21] and majority voting [6].

```
Algorithm 3 FEFS-OPF
Require:
    Ensemble size $E$, training sample set $T$, feature set $F$

Procedure:
for $i = 1$ to $E$ do
    $FS$ ← Selected features from $F$ by using a feature selection algorithm $A_i$
    Train the OPF classifier $C_i$ using $T$ with the selected features $FS$
end for
for each new sample do
    Compute the majority voting $V$ of all classification models of the ensemble $C_i$
end for
```

4. Experimental results

In this section, the classification results are described. In order to evaluate the effectiveness of the ensemble models for the classification of benign and malignant skin lesions, three experiments were performed. First, the experiments for the feature subset and feature selection evaluations; second, the experiments for the ensemble classification model evaluation; and finally, the experiments to compare the results with the classification methods reported in the literature. In addition, the evaluation process used to evaluate the results is introduced.

4.1. Evaluation process

The performance of the ensemble classification models based on the input feature manipulation as described in the previous section was evaluated by using a stratified k-fold cross-validation procedure [5]. This kind of procedure consists of splitting the training set in k subsets of equal size; the procedure being repeated k times. In each procedure, one subset is used as a test set while the others are used as the training set. The best model based on its performance is chosen. Performance is the average accuracy obtained from each trial. The k-fold cross-validation procedure can be applied to avoid over-fitting while testing the capacity of the classifier to generalize. In addition, it has shown good results compared with other procedures [41].

The measures used to evaluate the performance of the classification are accuracy (ACC), sensitivity (SE) and specificity (SP), which are based on outcomes of the ensemble of classifiers, according to the majority voting. These outcomes represent the number of correct and incorrect classifications for each class, positive (benign) and negative (malignant). These measures are commonly used [12] and they are defined as: SE is the percentage of correctly classified positive samples with respect to all positive samples, SP is the percentage of correctly classified negative samples with respect to all negative samples, and ACC is the percentage of correctly classified positive and negative samples based on all samples.

A cost function C adopted from Barata, et al. [2] is used to deal with the trade-off between SE and SP, which is defined as:

$$C = \frac{c_{10}(1-SE)+c_{01}(1-SP)}{c_{10}+c_{01}}, \tag{5}$$

where c_{10} is the cost of an incorrectly classified benign lesion (FN), and c_{01} is the cost of an incorrectly classified malignant lesion (FP). The costs used to evaluate the classification were $c_{10} = 1$ and $c_{01} = 1.5$, since an incorrect classification of a malignant lesion is more critical. The lower the value of cost C, the better the classification is.

4.2. Evaluation of the feature subset and feature selection

In order to define the best feature subsets for the ensemble classification models, several subsets based on specific feature groups discussed in the previous section were evaluated. Table 1 shows the results for each feature subset using the OPF classifier. Three distance functions, i.e., Euclidean, Chebyshev and Manhattan were compared using this classifier, in order to find the distances between the feature vectors. The Euclidean distance was the best distance function for this classifier, according to the experiments using all features, which achieved an ACC of 92.3%. Consequently, this distance function was used for all other experiments in this study.

Table 1 - Performance results for the feature subsets compared to different feature groups (best result for each group is in bold).

Group	Feature subset	ACC
1	Shape	89.1%
	Colour	91.0%
	Texture	**91.6%**
2	Fractal texture	89.9%
	Wavelet texture	**90.7%**
	Haralick's texture	88.3%
3	Shape and colour	90.5%
	Shape and texture	91.3%
	Colour and texture	**91.7%**
4	RGB colour	90.6%
	HSV colour	**92.0%**
	LAB colour	90.3%
	LUV colour	90.3%
5	RGB texture	**91.8%**
	HSV texture	91.1%
	LAB texture	91.2%
	LUV texture	90.8%
6	Shape and RGB features	91.6%
	Shape and HSV features	**93.0%**
	Shape and LAB features	92.7%
	Shape and LUV features	91.7%
7	RGB features	90.8%
	HSV features	91.2%
	LAB features	**92.5%**
	LUV features	91.4%

The results in Table 1 indicate that there is diversity between the feature subsets. The three best feature subsets were the shape combined with the HSV features, the LAB features, and the HSV colour. The shape, colour and texture features provided an improvement to the classification when they were combined. The texture features, i.e., the fractal, wavelet and Haralick's features, achieved better results when the features were combined than when they were used individually. The feature extraction algorithms for each colour space provided better results when combined with the shape features.

The diversity for an ensemble classification model is also obtained by using different feature selection algorithms. Such algorithms were used to find the best features for the classification process. The single-feature evaluators use a ranking method, i.e., the correlation coefficient, GRFS, information gain, relief-F and PCA, and a set of retained number of features is empirically defined by

$N = \{25, 50, 75\}$, with the exception of PCA that chooses a sufficient number of eigenvalues to rank the new transformed features. The maximum number of features $F = 5$ was used to include the transformed features, 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 feature estimation defined the number of nearest neighbours $k = 10$ for the relief-F.

In the case of the feature subset evaluator, i.e., CFS, the greedy stepwise search method, in either forward or backward directions, is applied until the addition or removal of any feature produces a decrease in evaluation. Consequently, 37 features were selected in the forward direction and 50 in the backward direction. The best first search method was also carried out until five consecutive non-improving features, in the directions: forward (37 features), backward (50 features) or bi-direction (37 features) were found. However, experimental results, using the OPF classifier as discussed in the previous section, showed that this method did not improve the classification when applied with the stepwise search method. Therefore, only the stepwise search method was used with CFS and compared with the other feature selection algorithms.

Table 2 shows the best classification results using the feature selection algorithms. Although all the feature selection algorithms obtained good results, the OPF classifier using the features selected by the CFS algorithm achieved the best results. These algorithms were applied to generate the feature subsets for the ensemble classification models.

Table 2 - Comparing several feature selection algorithms (best result is in bold).

Feature selection	Features	ACC
Correlation coefficient	75	89.6%
GRFS	25	91.1%
Information gain	75	90.8%
Relief-F	75	91.0%
PCA	31	91.0%
CFS	50	**91.6%**

4.3. Evaluation of the ensemble classification models

The performance of the three ensemble classification models based on the input feature manipulation, OPF classifier and majority voting; namely, the SE-OPF, SEFS-OPF, FEFS-OPF algorithms, were evaluated using ten-fold cross-validation. Four ensembles of classifiers were generated for each algorithm, where $E = \{3,5,7,9\}$ describes the ensemble size, i.e., the number of base classifiers. Several subsets based on the combination of the specific feature groups were performed for the feature manipulation using the SE-OPF algorithm. The best subsets of the specific feature groups were performed for the feature manipulation based on the SEFS-OPF algorithm using the CFS algorithm for each ensemble. In addition, all extracted features were performed using different feature selection algorithms for the feature manipulation based on the FEFS-OPF algorithm. Table 3 shows the combination of the subsets and feature selection algorithms for each ensemble that achieved the best classification results.

Table 3 - Combination of the subsets and feature selection algorithms for each ensemble.

Ensemble classification model	Number of classifier	Feature subsets
SE-OPF	3	Shape, colour and texture
	5	Shape, *RGB* features, *HSV* features, *LAB* features and *LUV* features
	7	Shape, colour, texture, shape + *RGB* features, shape + *HSV* features, shape + *LAB* features and shape + *LUV* features
	9	Shape, *RGB* colour, *HSV* colour, *LAB* colour, *LUV* colour, *RGB* texture, *HSV* texture, *LAB* texture and *LUV* texture
SEFS-OPF	3	Shape + CFS, colour + CFS and texture + CFS
	5	Shape + CFS, *RGB* features + CFS, *HSV* features + CFS, *LAB* features + CFS and *LUV* features + CFS
	7	Shape + CFS, colour + CFS, texture + CFS, shape + RGB features + CFS, shape + *HSV* features + CFS, shape + *LAB* features + CFS and shape + *LUV* features + CFS
	9	Shape + CFS, *RGB* colour + CFS, *HSV* colour + CFS, *LAB* colour + CFS, *LUV* colour + CFS, *RGB* texture + CFS, *HSV* texture + CFS, *LAB* texture + CFS and *LUV* texture + CFS
FEFS-OPF	3	All features + PCA, all features + CFS and all features + GRFS
	5	All features + PCA, all features + correlation coefficient, all features + GRFS, all features + information gain and all features + relief-F
	7	All features + PCA, all features + correlation coefficient, all features + GRFS, all features + information gain, all features + relief-F, all features + CFS (best first) and all features + CFS (stepwise)
	9	All features + PCA + OPF (ED), all features + CFS + OPF (ED), all features + GRFS + OPF (ED), all features + PCA + OPF (CD), all features + CFS + OPF (CD), all features + GRFS + OPF (CD), all features + PCA + OPF (MD), all features + CFS + OPF (MD) and all features + GRFS + OPF (MD)

ED: Euclidean distance, CD: Chebyshev distance and MD: Manhattan distance

Table 4 shows the best classification results for each ensemble model. The SE-OPF algorithm achieved its best classification results using $E = 9$. Likewise, 9 classifiers for the ensemble yielded the best results for the SEFS-OPF algorithm, whereas the FEFS-OPF algorithm obtained its best results using $E = 3$. Although the SE-OPF algorithm did not have all the best classification measures, it resulted in a more balanced classification between the benign and malignant classes, i.e., with a lower classification cost. The classification results are presented in more detail in Fig. 3, which shows the variation of the accuracy, sensitivity and specificity measures, according to the ensemble size defined for each ensemble classification model.

Table 4 - Classification results for the ensemble classification models (best results are in bold).

Ensemble classification model	ACC	SE	SP	C
SE-OPF (feature subsets + OPF)	**94.3%**	**91.8%**	96.7%	**0.053**
SEFS-OPF (feature subsets + CFS + OPF)	93.9%	**91.8%**	96.0%	0.057
FEFS-OPF (all features + FS + OPF)	93.7%	90.4%	**96.9%**	0.057

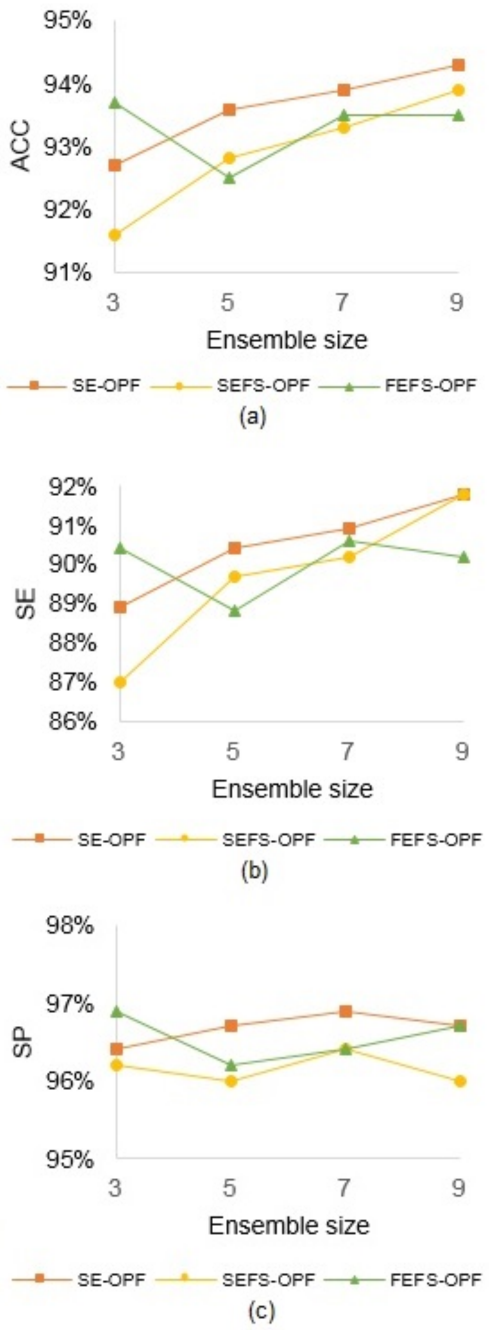

Fig. 3: Variation of the classification measures, according to the ensemble size established for each ensemble classification model: (a) accuracy, (b) sensitivity and (c) specificity.

4.4. Comparison between classification algorithms

The classification results achieved by the best ensemble model proposed here, based on the input feature manipulation as previously discussed, were compared against the ones obtained using three different ensemble algorithms. These algorithms are commonly used in the literature; namely, bagging [6], AdaBoost [13] and random forest [14]. The proposed ensemble model was also compared to the individual OPF classifier [21] to analyse the effectiveness of the ensemble algorithms. In addition, this classifier was adopted as a base classifier for the bagging and AdaBoost algorithms, since these algorithms can be used with any learning algorithm. The classification algorithms were applied with and without feature selection based on all the extracted features. The CFS algorithm was used in these experiments, since it improved the classification more than the other feature selection algorithms, as mentioned previously. Table 5 shows the results using different classification methods, as well as the results of the proposed model; the best results for each measure are shown in bold.

Table 5 - Comparative results between classification algorithms (best results are in bold).

Classification algorithms	ACC	SE	SP	C
OPF	92.3%	87.5%	**97.1%**	0.067
OPF (CFS)	91.6%	87.0%	96.2%	0.075
Bagging (OPF)	89.7%	85.9%	93.5%	0.095
Bagging (CFS + OPF)	91.8%	88.4%	95.3%	0.075
AdaBoostM1 (OPF)	92.3%	**92.3%**	92.3%	0.077
AdaBoostM1 (CFS + OPF)	91.6%	87.0%	96.2%	0.075
Random forest	93.9%	91.3%	96.6%	0.055
Random forest (CFS)	93.7%	90.4%	96.9%	0.057
Proposed model (SE + OPF)	**94.3%**	91.8%	96.7%	**0.053**

The results in Table 5 indicate that only the bagging and AdaBoostM1 algorithms achieved better results by using the features selected by the CFS algorithm rather than without the feature selection. Although the AdaBoostM1 algorithm without the feature selection yielded a better accuracy and achieved an average distinction between the benign and malignant classes, the cost was higher because the specificity was not very expressive. On the other hand, the random forest algorithm was more effective without the feature selection, since it obtained a better accuracy and a lower cost between the sensitivity and specificity. In addition, this algorithm obtained better classification results than the bagging and AdaBoostM1 algorithms. Likewise, the individual OPF classifier without the feature selection achieved better results than the bagging and AdaBoostM1 algorithms. Nevertheless, the accuracy obtained by the OPF classifier was not better than the random forest algorithm and the proposed model. Moreover, the classification cost was higher between the sensitivity and specificity. The proposed model showed good generalization between the benign and malignant classes. Furthermore, this model achieved a better accuracy and lower cost compared to other classification algorithms used in the literature.

5. Discussion

The main objective of this study was to develop and evaluate different classification models based on ensemble methods using input feature manipulation in order to improve the classification results for early image based diagnosis of skin cancer. The best proposed ensemble classification model achieved ACC = 94.3%, SE = 91.8% and SP = 96.7% in a dataset of 1104 dermoscopic images. This model was built by using the feature subset selection based on the combination of the specific feature subset groups for input feature manipulation, the OPF as base classifier, and the majority voting strategy to integrate several classification models. The classification results from the proposed model were more accurate than when the OPF classifier was used alone, and also more accurate than the standard ensemble algorithms, which are commonly found in the literature. Since this study did not use all the images from the original dataset, as mentioned previously, the results cannot be directly compared with the results obtained in the works that used the same dataset and the ground-truth lesion segmentation masks presented in Gutman, et al. [22]. These works used the full set of images from the data set which consisted of 1279 images and they divided them into test and training sets. The best results were achieved by Lequan, et al. [42] using the whole dataset, obtaining ACC = 0.855, SE = 0.547 and SP = 0.931. These latter authors proposed a novel method for melanoma recognition by leveraging very deep convolutional neural networks.

Several automatic diagnosis systems based on ensemble methods have been proposed in the literature for the skin lesion classification as described in the previous section about related studies. The results obtained from the proposed model are in line with those of other studies found in the literature, which also proposed ensemble methods to improve the classification of skin lesions and achieved high values of accuracy, sensitivity and specificity. In contrast, the studies presented in the literature usually computed the SE measure to represent the number, or percentage, of correctly classified malignant lesions, and the SP measure to represent the number, or percentage, of correctly classified benign lesions. For example, Barata, et al. [2] proposed a classification system using the AdaBoost algorithm that achieved SE = 96% and SP = 80% in a dataset of 176 dermoscopic images. Rastgoo, et al. [15] developed an automatic system based on texture features and random forest algorithm, which achieved SE = 98% and SP = 70% in a dataset of 180 dermoscopic images. Barata, et al. [10] built a classification system based on the fusion of global and local features using the random forest algorithm, which obtained SE = 98% and SP = 90% in a dataset of 200 dermoscopic images, and SE = 83% and SP = 76% in a dataset of 482 images. Rastgoo, et al. [11] proposed an automatic framework that used a random forest algorithm and a combination of colour and texture features based on global features, which obtained SE = 94% and SP = 92% in a dataset of 193 dermoscopic images. Abedini, et al. [18] developed an ensemble model based on feature random subsets, a linear SVM classifier and forward model selection for the ensemble fusion, which achieved ACC = 91%, SE = 97% and SP = 65% in a dataset of 200 dermoscopic images. Schaefer, et al. [4] proposed a multiple classifier system that consists of a random under-sampling method, an SVM with a polynomial kernel, and a neural network for the ensemble fusion, which obtained ACC = 93.83%, SE = 93.76% and SP = 93.84% in a dataset of 564 dermoscopic images. Xie, et al. [20] developed a novel meta-ensemble model based on multiple neural network ensembles, which achieved ACC = 94.17%, SE = 95% and SP = 93.75% in a dataset of 240 dermoscopic images of the xanthous race, and ACC = 91.11%, SE = 83.33% and SP = 95% in a dataset of 360 dermoscopic images of the caucasian race.

Usually, the automatic computational systems, like the ones proposed in the above-mentioned studies, include the segmentation of the skin lesions. Segmentation is an important step that allows the extraction of the regions of interest (ROIs) from an input image. Previous studies have shown that computational methods for image segmentation can provide suitable results for the identification of skin lesions in images [43, 44]. The classification results obtained can depend on the segmentation method used, since the features are extracted from the segmented ROIs. Thus, segmentation methods that obtain suitable ROIs may facilitate the classification process and lead to better classification results. The lack of a lesion segmentation process can be seen as a limitation of the present study; however, one should note that ground-truth lesion segmentation masks were used in order to obtain trustworthy classification results and conclusions. A possible skin lesion segmentation approach to be combined with the proposed ensemble classification model can be the one presented in Ma and Tavares [45], which is based on a level-set model and colour models and it has obtained very promising results.

Although the ensemble algorithms improve accuracy by combining the different classification models, these algorithms can present a high computational complexity and are rather hard to analyse [5]. Comprehensible models [46], which can be used to solve such problems, aim to produce a single classification model from an ensemble model without losing too much accuracy compared to using the integrated hypothesis model.

The proposed ensemble classification model based on input feature manipulation 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, JavaSE-1.8, and Weka 3.8 library for the classification algorithms. The feature extraction times for all the images from the binary masks were: shape - 10.26 min; colour - 10.12 min; fractal texture - 26.79 min; wavelet texture - 34.37 min; and Haralick's texture - 29.48 min. Finally, the best ensemble classification model required a total of 60.09 s to process all the samples. These values show 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. All algorithms were performed on an Intel(R) Core(TM) i5 CPU 650 @ 3.20 GHz with 8 GB of RAM, running Microsoft Windows 7 Professional 64-bits.

6. Conclusion and future works

In this article, three ensemble classification models based on input feature manipulation from the shape properties, colour variation and texture analysis, were presented; namely, the SE-OPF, SEFS-OPF and FEFS-OPF algorithms. The first model manipulates the features by using different subsets based on specific feature groups. The second model manipulates the features by using the CFS algorithm for the feature selection from the subsets defined in the first model. Finally, the third model manipulates the features by using different feature selection algorithms, i.e., correlation coefficient, GRFS, information gain, relief-F, PCA and CFS, from all extracted features. Each ensemble model was generated by using the OPF base classifier and integrated with the majority voting strategy. The effectiveness of the feature groups and feature selection algorithms used were individually evaluated to find the best features for the classification process, as well as to generate diversity for the ensemble classification models.

Promising results were achieved with the proposed ensemble classification models. The best classification results were obtained by the feature subset selection model based on feature groups (SE-OPF algorithm). Nine base classifiers were used for this model based on shape, *RGB* colour, *HSV* colour, *LAB* colour, *LUV* colour, *RGB* texture, *HSV* texture, *LAB* texture and *LUV* texture subsets, which yielded the following results: ACC = 94.3%, SE = 91.8% and SP = 96.7%. The feature manipulation process based on these specific feature subsets also provided an excellent generation of diversity for the ensemble classification model.

Future studies for pigmented skin lesion classification from dermoscopic images should search for new methods to develop more efficient and effective systems. In order to approach other challenges of dermoscopy image diagnoses, the proposed ensemble classification models should be taken into account in future works to identify the presence of global and local patterns. Discriminating between benign and malignant skin lesions is a challenging task for pattern analysis [47]. Essentially, the classification results can be improved by using deep learning architectures [48], since these architectures have revealed their capacity to learn from large amounts of data. Therefore, deep learning architectures should be taken into account in future works concerning skin lesion classification in dermoscopic images.

ACKNOWLEDGMENTS

The first author would like to thank CNPq ("Conselho Nacional de Desenvolvimento Científico e Tecnológico"), in Brazil, for her PhD grant. Authors gratefully acknowledge the funding of Project NORTE-01-0145-FEDER-000022 - SciTech - Science and Technology for Competitive and Sustainable Industries, co-financed by "Programa Operacional Regional do Norte" (NORTE2020), through "Fundo Europeu de Desenvolvimento Regional" (FEDER).

APPENDIX

The International Skin Imaging Collaboration (ISIC) image dataset [22] was used in the experiments presented in this article. This dataset is composed of 1279 images, however, 175 lesions did not fully fit into the image and were therefore excluded from this study; the excluded images were the XXXXX with XXXXX equal to: 00004, 00160, 00183, 00187, 00188, 00189, 00195, 00202, 00207, 00204, 00205, 00207, 00208, 00209, 00212, 00215, 00217, 00224, 00230, 00233, 00235, 00258, 00259, 00282, 00285, 00288, 00289, 00290, 00292, 00293, 00294, 00295, 00299, 00300, 00301, 00302, 00303, 00307, 00310, 00311, 00321, 00359, 00368, 00371, 00390, 00408, 00415, 00426, 00433, 00481, 00517, 00519, 01142, 09896, 09897, 09904, 09121, 09928, 09934, 09962, 09970, 09980, 09983, 09988, 09990, 09995, 10000, 10005, 10022, 10029, 10035, 10037, 10038, 10041, 10051, 10056, 10057, 10062, 10063, 10065, 10068, 10071, 10073, 10075, 10078, 10093, 10169, 10189, 10190, 10194, 10202, 10204, 10213, 10222, 10228, 10231, 10232, 10235, 10238, 10239, 10242, 10252, 10257, 10267, 10320, 10322, 10323, 10324, 10329, 10331, 10334, 10335, 10342, 10348, 10349, 10369, 10382, 10448, 10454, 10455, 10457, 10459, 10477, 10492, 10495, 10497, 10572, 10588, 10589, 10596, 10603, 10604, 10605, 10847, 10857, 11079, 11088, 11104, 11105, 11109, 11112, 11120, 11121, 11122, 11126, 11128, 11139, 11149, 11151, 11156, 11158, 11159, 11167, 11175, 11203, 11212, 11229, 11229, 11300, 11322, 11327, 11329, 11334, 11347, 11348, 11349, 11350, 11356, 11360, 11361, 11373, 11374, 11387, 11390, 11402.

REFERENCES

[1] American Cancer Society, Global cancer facts & figures, 3 ed., American Cancer Society, 2015.
[2] C. Barata, M. Ruela, M. Francisco, T. Mendonça, J.S. Marques, Two systems for the detection of melanomas in dermoscopy images using texture and color features, IEEE Systems Journal 8 (2013) 965-979. doi: 10.1109/JSYST.2013.2271540.
[3] M. Abedini, N.C.F. Codella, J.H. Connell, R. Garnavi, M. Merler, S. Pankanti, J.R. Smith, T. Syeda-Mahmood, A generalized framework for medical image classification and recognition, IBM Journal of Research and Development, 59 (2015) 1-18. doi: 10.1147/JRD.2015.2390017.
[4] G. Schaefer, B. Krawczyk, M.E. Celebi, H. Iyatomi, An ensemble classification approach for melanoma diagnosis, Memetic Computing, 6 (2014) 233-240. doi: 10.1007/s12293-014-0144-8.
[5] I.H. Witten, E. Frank, M.A. Hall, Data mining: Practical machine learning tools and techniques, Morgan Kaufmann, 2011.
[6] T.G. Dietterich, Ensemble methods in machine learning, Multiple classifier systems, Springer, Berlin, Heidelberg, 2000, pp. 1-15.
[7] M.M. Rahman, P. Bhattacharya, B.C. Desai, "A multiple expert-based melanoma recognition system for dermoscopic images of pigmented skin lesions," in proceedings of the 8th IEEE International Conference on International Conference on BioInformatics and BioEngineering, pp. 1-6, 2008. doi: 10.1109/BIBE.2008.4696799.
[8] M. Blachnik, Ensembles of instance selection methods based on feature subset, Procedia Computer Science, 35 (2014) 388-396. doi: 10.1016/j.procs.2014.08.119.
[9] M. Dash, H. Liu, Feature selection for classification, Intelligent data analysis, 1 (1997) 131-156. doi: 10.1016/S1088-467X(97)00008-5.
[10] C. Barata, M. Emre Celebi, J.S. Marques, "Melanoma detection algorithm based on feature fusion," in proceedings of the 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society pp. 2653-2656, 2015.
[11] M. Rastgoo, O. Morel, F. Marzani, R. Garcia, "Ensemble approach for differentiation of malignant melanoma," in proceedings of the The International Conference on Quality Control by Artificial Vision 2015, pp. 953415-953419, 2015. doi: 10.1117/12.2182799.

[12] R.B. Oliveira, J.P. Papa, A.S. Pereira, J.M.R.S. Tavares, Computational methods for pigmented skin lesion classification in images: Review and future trends, Neural Computing and Applications, 27 (2016) 1-24. doi: 10.1007/s00521-016-2482-6.

[13] Y. Freund, R.E. Schapire, A decision-theoretic generalization of on-line learning and an application to boosting, Journal of Computer and System Sciences, 55 (1997) 119-139. doi: 10.1006/jcss.1997.1504.

[14] L. Breiman, Random forests, Machine learning, 45 (2001) 5-32. doi: 10.1023/A:1010933404324.

[15] M. Rastgoo, R. Garcia, O. Morel, F. Marzani, Automatic differentiation of melanoma from dysplastic nevi, Computerized Medical Imaging and Graphics, 43 (2015) 44-52. doi: 10.1016/j.compmedimag.2015.02.011.

[16] R. Garnavi, M. Aldeen, J. Bailey, Computer-aided diagnosis of melanoma using border- and wavelet-based texture analysis, IEEE Transactions on Information Technology in Biomedicine 16 (2012) 1239-1252. doi: 10.1109/titb.2012.2212282.

[17] M. Maragoudakis, I. Maglogiannis, "Skin lesion diagnosis from images using novel ensemble classification techniques," in proceedings of the 10th IEEE International Conference on Information Technology and Applications in Biomedicine, pp. 1-5, 2010. doi: 10.1109/ITAB.2010.5687620.

[18] M. Abedini, Q. Chen, N.C.F. Codella, R. Garnavi, X. Sun, Accurate and scalable system for automatic detection of malignant melanoma, in: M.E. Celebi, T. Mendonca, J.S. Marques (Eds.) Dermoscopy image analysis, CRC Press, 2015, pp. 293-343.

[19] A. Sboner, C. Eccher, E. Blanzieri, P. Bauer, M. Cristofolini, G. Zumiani, S. Forti, A multiple classifier system for early melanoma diagnosis, Artificial Intelligence in Medicine, 27 (2003) 29-44. doi: 10.1016/S0933-3657(02)00087-8.

[20] F. Xie, H. Fan, Y. Li, Z. Jiang, R. Meng, A. Bovik, Melanoma classification on dermoscopy images using a neural network ensemble model, IEEE Transactions on Medical Imaging, 36 (2017) 849-858. doi: 10.1109/TMI.2016.2633551.

[21] J.P. Papa, A.X. Falcao, C.T. Suzuki, Supervised pattern classification based on optimum-path forest, International Journal of Imaging Systems and Technology, 19 (2009) 120-131. doi: 10.1002/ima.20188.

[22] D. Gutman, N.C.F. Codella, E. Celebi, B. Helba, M. Marchetti, N. Mishra, A.C. Halpern, "Skin lesion analysis toward melanoma detection: A challenge," in the International Symposium on Biomedical Imaging (ISBI) 2016, hosted by the International Skin Imaging Collaboration (ISIC), arxiv preprint arxiv:1605.01397.

[23] H. Iyatomi, K. Norton, M.E. Celebi, G. Schaefer, M. Tanaka, K. Ogawa, "Classification of melanocytic skin lesions from non-melanocytic lesions," in proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society pp. 5407-5410, 2010. doi: 10.1109/iembs.2010.5626500.

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

[25] M.E. Celebi, H. Iyatomi, W.V. Stoecker, R.H. Moss, H.S. Rabinovitz, G. Argenziano, H.P. Soyer, Automatic detection of blue-white veil and related structures in dermoscopy images, Computerized Medical Imaging and Graphics, 32 (2008) 670-677. doi: 10.1016/j.compmedimag.2008.08.003.

[26] R.B. Oliveira, N. Marranghello, A.S. Pereira, J.M.R.S. Tavares, A computational approach for detecting pigmented skin lesions in macroscopic images, Expert Systems with Applications, 61 (2016) 53-63. doi: 10.1016/j.eswa.2016.05.017.

[27] M. Tkalcic, J.F. Tasic, "Colour spaces: Perceptual, historical and applicational background," in proceedings of the IEEE Region 8 EUROCON 2003: Computer as a Tool pp. 304-308, 2003. doi: 10.1109/EURCON.2003.1248032.

[28] M.E. Celebi, H.A. Kingravi, B. Uddin, H. Iyatomi, Y.A. Aslandogan, W.V. Stoecker, R.H. Moss, A methodological approach to the classification of dermoscopy images, Computerized Medical Imaging and Graphics, 31 (2007) 362-373. doi: 10.1016/j.compmedimag.2007.01.003.

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

[30] M. Al-Akaidi, Fractal speech processing, Cambridge university press, 2004.

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

[32] R.M. Haralick, K. Shanmugam, I.H. Dinstein, Textural features for image classification, IEEE Transactions on Systems, Man and Cybernetics, SMC-3 (1973) 610-621. doi: 10.1109/TSMC.1973.4309314.

[33] N.V. Chawla, Data mining for imbalanced datasets: An overview, in: O. Maimon, L. Rokach (Eds.) Data mining and knowledge discovery handbook, Springer, 2005, pp. 853-867.

[34] M. Rastgoo, G. Lemaitre, J. Massich, O. Morel, F. Marzani, R. Garcia, F. Meriaudeau, "Tackling the problem of data imbalancing for melanoma classification," in proceedings of the 3rd International Conference on Bioimaging, pp., 2016.

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

[36] I. Kononenko, Estimating attributes: Analysis and extensions of relief, in: F. Bergadano, L. De Raedt (Eds.) Machine learning: Ecml-94, Springer, Berlin, Heidelberg, 1994, pp. 171-182.

[37] D. Hand, H. Mannila, P. Smyth, Principles of data mining, The MIT Press, 2001.

[38] M.A. Hall, "Correlation-based feature selection for discrete and numeric class machine learning," in proceedings of the 17th International Conference on Machine Learning, pp. 359-366, 2000.

[39] M.P. Ponti Jr, I. Rossi, Ensembles of optimum-path forest classifiers using input data manipulation and undersampling, in: Z.-H. Zhou, F. Roli, J. Kittler (Eds.) Multiple classifier systems, Springer, Berlin, Heidelberg, 2013, pp. 236-246.

[40] W.P. Amorim, A.X. Falcão, M.H. de Carvalho, "Semi-supervised pattern classification using optimum-path forest," in proceedings of the 27th SIBGRAPI Conference on Graphics, Patterns and Images, pp. 111-118, 2014. doi: 10.1109/SIBGRAPI.2014.45.

[41] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in proceedings of the 14th International Joint Conference on Artificial Intelligence, pp. 1137-1145, 1995.

[42] Y. Lequan, H. Chen, Q. Dou, J. Qin, P.A. Heng, Automated melanoma recognition in dermoscopy images via very deep residual networks, IEEE Transactions on Medical Imaging (2016) 1-11. doi: 10.1109/TMI.2016.2642839.

[43] M.E. Celebi, Q. Wen, H. Iyatomi, K. Shimizu, H. Zhou, G. Schaefer, A state-of-the-art survey on lesion border detection in dermoscopy images, in: M.E. Celebi, T. Mendonca, J.S. Marques (Eds.) Dermoscopy image analysis, CRC Press, 2015, pp. 97-129.

[44] M.E. Celebi, H. Iyatomi, G. Schaefer, W.V. Stoecker, Lesion border detection in dermoscopy images, Computerized medical imaging and graphics, 33 (2009) 148-153. doi: 10.1016/j.compmedimag.2008.11.002.

[45] Z. Ma, J.M.R.S. Tavares, A novel approach to segment skin lesions in dermoscopic images based on a deformable model, IEEE Journal of Biomedical and Health Informatics, 20 (2016) 615-623. doi: 10.1109/JBHI.2015.2390032.

[46] C. Ferri, J. Hernández-Orallo, M.J. Ramírez-Quintana, From ensemble methods to comprehensible models, in: S. Lange, K. Satoh, C.H. Smith (Eds.) Discovery science, Springer, Berlin, Heidelberg, 2002, pp. 165-177.

[47] G. Argenziano, H.P. Soyer, S. Chimenti, R. Talamini, R. Corona, F. Sera, M. Binder, L. Cerroni, G. De Rosa, G. Ferrara, R. Hofmann-Wellenhof, M. Landthaler, S.W. Menzies, H. Pehamberger, D. Piccolo, H.S. Rabinovitz, R. Schiffner, S. Staibano, W. Stolz, I. Bartenjev, A. Blum, R. Braun, H. Cabo, P. Carli, V. De Giorgi, M.G. Fleming, J.M. Grichnik, C.M. Grin, A.C. Halpern, R. Johr, B. Katz, R.O. Kenet, H. Kittler, J. Kreusch, J. Malvehy, G. Mazzocchetti, M. Oliviero, F. Özdemir, K. Peris, R. Perotti, A. Perusquia, M.A. Pizzichetta, S. Puig, B. Rao, P. Rubegni, T. Saida, M. Scalvenzi, S. Seidenari, I. Stanganelli, M. Tanaka, K. Westerhoff, I.H. Wolf, O. Braun-Falco, H. Kerl, T. Nishikawa, K. Wolff, A.W. Kopf, Dermoscopy of pigmented skin lesions: Results of a consensus meeting via the internet, Journal of the American Academy of Dermatology, 48 (2003) 679-693. doi: 10.1067/mjd.2003.281.

[48] Y. Bengio, Learning deep architectures for ai, Foundations and trends® in Machine Learning, 2 (2009) 1-127. doi: 10.1561/2200000006.