Expert Systems With Applications 145 (2020) 113127

Contents lists available at ScienceDirect

Expert Systems With Applications

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

A hierarchical structure based on Stacking approach for skin lesion classification

Ghasem Shakourian Ghalejoogh, Hussain Montazery Kordy*, Farideh Ebrahimi

Faculty of Electrical and Computer Engineering, Babol Noshirvani University of Technology, Shariati Ave., Babol, Iran

ARTICLE INFO

Article history:
Received 2 June 2019
Revised 4 November 2019
Accepted 8 December 2019
Available online 10 December 2019

Keywords:
Skin cancer
Melanoma
Ensemble classifiers
Meta learning
Stacking

ABSTRACT

Malignant melanoma is the most dangerous type of skin cancer. The diagnosis of melanoma in the early stages can greatly increase the possibility of its successful treatment. In the recent years, automated systems have played an pivotal role in increasing the skin cancer diagnosis rate. The main objective of this paper was to improve the performance of a skin cancer automated diagnostic system by introducing a new approach to combining classifiers in the classification stage. Therefore, the Stacking Ensemble Method based on the Meta Learning algorithm was proposed for the skin lesion classification. To classify skin lesions as melanoma, dysplastic and benign, two new hybrid approaches of Structure Based on Stacking (SBS) and Hierarchical Structure Based on Stacking (HSBS) were introduced to combine the heterogeneous classifiers. The proposed methods for skin lesions classification were implemented and evaluated based on the dermoscopic images of two PH^2 and Ganster datasets using the Five Fold Cross Validation procedure and different numbers of the selected features. The results showed that the SBS approach had a good performance in diagnosing melanoma lesions from non-melanoma lesions for both datasets. Moreover, the results indicated that the HSBS method compared to the SBS approach and other works on the same dataset offers a far better performance in classifying skin lesions as benign, dysplastic and melanoma.

1. Introduction

Skin cancer is one of the most common cancers in the world. The main cause of the cancer is exposure to the ultraviolet radiation emitted from the sun. Squamous cell carcinoma, basal cell carcinoma, and melanoma are the three most common types of cancer. Malignant Melanoma is the most dangerous and deadly type of skin cancer and hence, is more important than the two other types. According to the American Cancer Association (2018), the number of 91,270 new cases of melanoma were estimated in the United States, of which 9320 people died because of this disease (Cancer Fact & Figures, 2018). Diagnosis of melanoma at the early stages can greatly increase the possibility of successful treatment. Dermatologists follow the "ABCDE" rule to identify melanoma. This rule examines the lesion based on the five characteristics of asymmetry, boundary disorder, color variation, diameter greater than 6 mm, and evolution over time (Queen, 2017). Therefore, if a lesion has the above-mentioned symptoms, it is known to be a malignant tumor or melanoma and is treated by surgery or other methods. Due to similarities between benign, dysplastic and melanoma lesions, it is difficult to diagnose melanoma correctly with the naked eye and a mistake in the diagnosis results. As a result, the use of dermoscopy method has been proposed to examine the lesion. Dermoscopy is a non-invasive method for imaging from surface of the skin. This method provides a more accurate examination of the lesion by illumination and magnification (Gandhi & Kampp, 2015). Fig. 1 shows the comparison of the clinical photography and dermoscopy images of melanoma lesion.

Employing dermoscopy, due to its complexity, requires a lot of experience and expertise. Studies have shown that if a dermatologist does not have enough experience, the accuracy of the melanoma diagnosis reduces (Schaefer, Krawczyk, Celebi & Iyatomi, 2014). Therefore, the use of computer diagnosis systems is proposed to reduce diagnosis errors due to lack of expertise and experience of a dermatologist. According to studies, in cases where the dermatologist uses the computerized system to detect melanoma, the diagnosis accuracy increases from 75% to over 92% (Xie et al., 2016). A skin cancer detection system generally has four stages of preprocessing, segmentation, feature extraction, and classification. The main objective of this paper is to propose a new approach for combining heterogeneous base classifiers in order to classify skin lesions and achieve favorable diagnosis rate based on stacking method. Moreover, in this paper, in addition

* Corresponding author.
E-mail addresses: ghasemsh@stu.nit.ac.ir (G.S. Ghalejoogh), hmontazery@nit.ac.ir (H.M. Kordy), f.ebrahimi96@nit.ac.ir (F. Ebrahimi).

https://doi.org/10.1016/j.eswa.2019.113127
0957-4174/

Fig. 1. Image of Melanoma skin lesion by clinical photography and dermoscopy: a) clinical photography of melanoma; b) dermoscopy image of melanoma.

to distinguishing melanoma from non-melanoma based on the stacking method, the classification of lesions into three classes of benign, dysplastic and melanoma is presented by introducing a new hierarchical structure based on stacking.

Accordingly, a binary image of hairs is created using the median filter and morphological operators in the preprocessing stage and then, by the image inpainting approach, the hairy areas of the image are removed. In order to segment the lesion area, after improving the contrast between the area of the lesion and the normal skin, the image from the RGB color space is transferred to the HSV color space. Then, the image is thresholded by the Otsu method and finally, in the post-processing stage, the unwanted areas of the image are removed from the image to obtain the final segmented image. At the feature extraction stage, according to the "ABCD" rule, three types of features based on the shape, color and texture are extracted from the lesion area of each image. After this, to reduce the dimensions of the feature vector and to eliminate unnecessary features, the feature selection process is applied based on two filter and wrapper methods. Finally, in the data classification phase, instead of using an independent classifier for the classification of skin lesions, a new hybrid approach based on the Meta learning method is used to improve the performance of the classification algorithm. Fig. 2 shows the different stages of the implementation of a skin cancer diagnosis system.

This paper is organized in five sections. In the first part, some of the research on skin cancer is reviewed. The datasets used in this research are presented in the second section. The proposed methodology and its implementation are summarized in the third section. The fourth section examines the results of the implementation of the proposed methods and the final section is allocated to the conclusion.

2. Related works

Research on Computer-aided diagnosis (CAD) systems for skin lesions is conducted in different fields according to different stages of a diagnostic system. The first studies for the automatic classification of pigmented lesions were published in 1987 (Cascinelli, Ferrario, Tonelli, & Leo, 1987). After that, various approaches for Preprocessing, Segmentation, Feature Extraction and Classification were provided. Below, the articles related to CAD systems for skin cancer are reviewed to remove hairs from the image, Abbas, Garcia, Emre Celebi and Ahmad (2013), used a Gaussian filter which detected the hairs. Then, by thresholding method, a binary mask was produced from hairs. After removal of additional areas by morphological operators, Fast Marching Inpainting method was used to remove the hair from image.

Pathan, Prabhu and Siddalingaswamy (2018a,b) used a bank of Gabor filters and CIEl*a*b color space for hairs detection and used Mumford-Shah-based inpainting method to remove hairs in dermoscopic images.

The introducted algorithm for preprocessing and hairs removal by Jaworek and Tadeusiewicz (2013) is based on Gaussian filter and Top-Hat transform. They replaced neighborhood pixels to remove hairs from the image.

Applying a color normalisation technique, namely Automatic Color Equalization (ACE), as a preprocessing step for improving the image contrast was proposed by Schaefer, Rajab, Celebi and

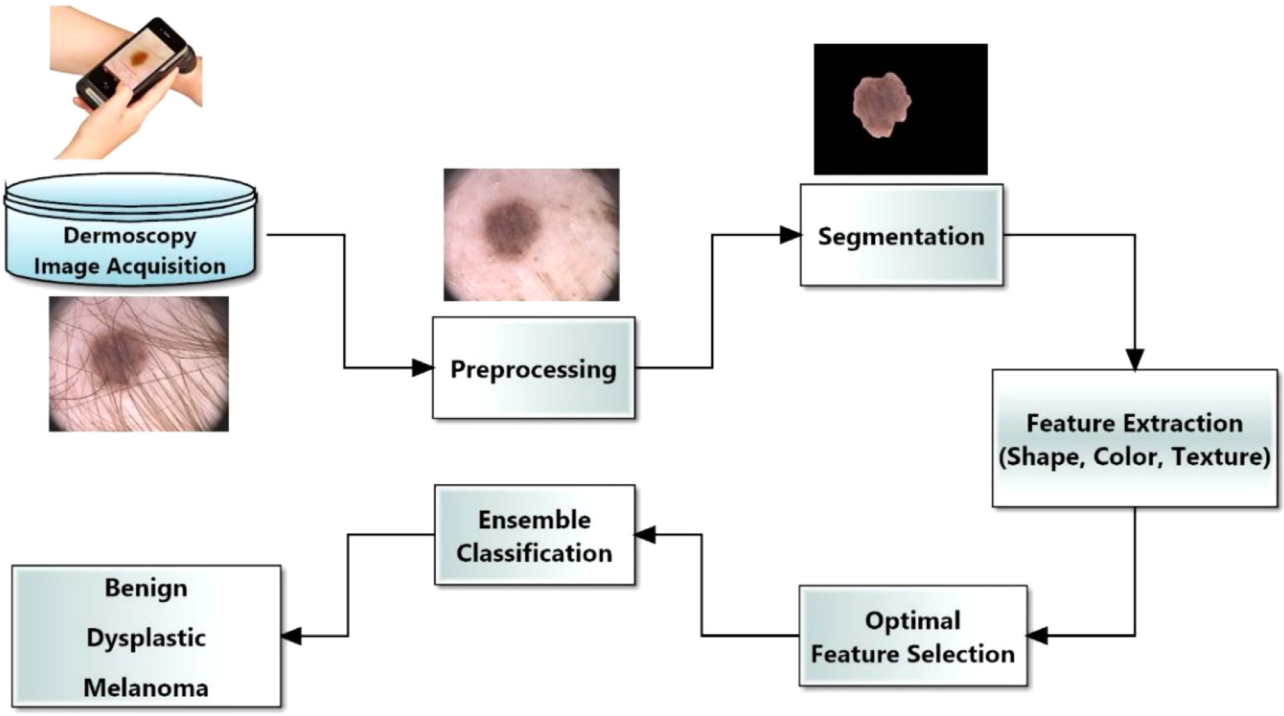

Fig. 2. Steps in Computer aided diagnosis system.

Iyatomi (2011). In this research, they employed two techniques for lesion segmentation: One employing an iterative segmentation scheme, and the other based on co-operative neural network edge detection.

In another study by Celebi et al. (2007), the JSEG algorithm was used to segment the skin lesion. This method consists of three stages of preprocessing, segmentation and post processing. So, in the preprocessing stage, the image was first filtered by using the median filter. Then, an estimate of the lesion area was obtained based on the Otsu thresholding method. In the following, border areas were determined based on local similarity and then Region Merging operations were applied. Finally, to achieve the final boundaries, additional areas were removed and the remaining pixels were merged.

The segmentation of the lesion based on the wavelet transform was presented in a study by Khalid et al. (2016). In this study, the Blue channel of the RGB image was decomposed into two levels, and then it was thresholded. The final binary image was obtained by smoothing the edges and filling the holes of the lesion area.

Garnavi, Aldeen and Bailey (2012) introduced a border detection method which encompasses two stages; the first stage applies global thresholding to detect an initial boundary of the lesion, and the second stage applies adaptive histogram thresholding on optimized color channels of X (from the CIE XYZ color space) to refine the border. In the feature extraction stage, texture features based wavelet tree and boundary series features were calculated. Then, feature vector dimensions were reduced by Gain Ratio Feature Selection (GRFS) method. Finally, in the classification stage, four classifiers- Random forest, SVM, LMT and HNB-were used and evaluated by 289 dermoscopic images. Reported results showed that Random forest with 23 selected features provides a good performance.

According to the presented method in a study by Rastgoo, Garcia, Morel and Marzani (2015), the boundary areas were first determined from the XYZ color space through the Global thresholding and adaptive histogram on the X channel. Next, global and local features were extracted from the lesion area. For shape analysis, features such as Thinness Ratio, Border asymmetry were extracted. Color variance in RGB, HIS and LAB color spaces and color histogram in RGB were calculated as color features. The LBP, GLCM, HOG and SIFT features were used to analyze the texture of the lesion area. In the classification stage, SVM, Gradient boosting and Random Forest classifiers were used to classify Melanoma and Dysplastic lesions. The obtained results for this research showed that the global features and RF classifier provide a good result with 98.46% sensitivity.

Pigmented Networks Detecting skin lesion for the purpose of melanoma diagnosis is presented in a study by Pathan et al. (2018b). In this research, after removing the hairs from dermoscopic images, typical and atypical pigment networks were extracted from the lesion area. Then, Color, Texture and Geometric features were extracted from pigmented networks. Finally, the SVM and ANN classifiers were used to classify skin lesions as melanoma and non-melanoma. The reported result for PH^2 dataset showed that using pigmented networks features and ANN classifier provide a good performance with an accuracy of 96.7%.

The use of Ensemble models for the diagnosis of skin cancer was proposed by Oliveira, Pereira and Tavares (2017). According to this research, 510 features were first extracted from the lesion area of each image based on the shape, color and texture. Then, by using the homogeneous classifiers and features manipulation, various models of the ensemble classifiers were created which ultimately reached to an accuracy of 94.3%.

The research by Schaefer et al. (2014) addressed the classification of skin lesions for unbalanced datasets based on the combination of the classifiers. According to this method, for balancing data, different subsets were created based on the Under sampling method. Then, for each of these subsets, the Support Vector Machine (SVM) was used as a base classifier and was considered as the feature selection operation. Finally, after pruning the weak classifiers and integrating their predictions by the Neural Network, the final structure was created and evaluated. The evaluation results indicated that their proposed method has good performance with the sensitivity of 93.76% and specificity of 93.84%.

Table 1
Summary of related works for dermoscopic images preprocessing.

Author (s)	Dataset	Method
Abbas et al. (2013)	100 Dermoscopy Images of EDRA CDROM	Guassian Filter + Threshold-ing + Fast Marching Inpainting
Pathan et al. (2018)	200 Dermoscopy Images of PH^2 Dataset	Gabor filter bank + CIEl*a*b color space + Mumford-Shah-based Inpainting
Jaworek et al. (2013)	50 Dermoscopy Images of Naples and Graz university	Guassian Filter + Top-Hat transform
Schaefer et al. (2011)	100 Images of EDRA Interactive Atlas of Dermoscopy	Automatic Color Equalization (ACE)
Pathan et al. (2018)	200 Dermoscopy Images of PH^2 Dataset	Laplacian Filter + Median filter + Bottom-Hat Transfom + Mumford-Shah-based Inpainting
Oliveira et al. (2016)	408 Macroscopic images	Anisotropic Diffusion filter

The application of Pigmented Skin lesions detecting in macroscopic images was presented by Oliveira, Marranghello, Pereira and Tavares (2016). In this research, for macroscopic image preprocessing, the anisotropic diffusion filter was applied to remove artifacts. Then, Chan-Vese model was used in order to the lesion area segmentation. For classification of pigmented skin lesions, after creating feature vector that was based on shape Asymmetry, irregularity border, Statistical measure of Color and box-counting method (BCM) for Texture analysis, the SVM classifier with adapted histogram intersection kernel was proposed. The evaluation of the proposed method by a dataset including 408 macroscopic images showed that the histogram intersection provides a better performance compared to the RBF function for pigmented skin lesions classification.

In order for Melanoma classification on dermoscopy images, Xie et al. (2016) proposed a Neural Network Ensemble Model. This work involve three steps. First, the lesion area was segmented by a Self-Generating Neural Network (SGNN) model. Then boundary, color and texture features were extracted from tumor region and in the classification stage, input features feeding to different types of Neural Network ensemble model in first level and outputs were integrated in the second level in order to create final decision. This model evaluated by dermoscopic images of two datasets provided an accuracy of 94.1% and 91.11% for each dataset, respectively.

All related studies presented above are summarized in Tables 1–3 below. In the reviewed research, most of the studies are based on the diagnosis of Melanoma from Non-melanoma, while dysplastic moles are considered as Benign cases. This is while dysplastic has a higher risk of becoming Melanoma (Rastgoo et al., 2015). Therefore, in this paper, dysplastic lesion is considered as an independent class. In the feature selection stage, often, filter based feature selection methods are used. Hence, in this paper, in addition to the filter based method, a Wrapper-based approach was also

Table 2
Summary of related works for skin lesions segmentation.

Author (s)	Dataset	Method
Schaefer et al. (2011)	100 Images of EDRA Interactive Atlas of Dermoscopy	Iterative segmentation + ISO-DATA + Co-operative Neural Network
Celebi et al. (2007)	596 Dermoscopy Images of Naples, florence and Graz university	JSEG algorithm + Region Merging
Khalid et al. (2016)	200 Dermoscopy Images of PH^2 Dataset	Discrete Wavelet Transform
Garnavi et al. (2012)	289 Dermoscopy images	Global thresholding + Adaptive Histogram thresholding + XYZ color space
Oliveira et al. (2016)	408 Macroscopic images	Chan-Vese Model
Xie et al. (2016)	240 Dermoscopy images of xanthous race Dataset and 360 Dermoscopy images caucasian race Dataset	Self-Generating Neural Network (SGNN) model
Rastgoo et al. (2015)	180 Images of Vienna general Hospital	Global thresholding + Adaptive Histogram thresholding + XYZ color space

examined in the feature selection process. Moreover, for combining classifiers to create ensemble classification, the homogeneous classifiers were mostly used and the required variety to combine classifiers was created by input features manipulation. However, in this research the heterogeneous base classifiers to ensemble classification was employed to provide a robust performance. On the other hand, instead of using the voting-based methods to combining classifiers, a separate classifier was used to ensemble base classifiers. Also, Abuzaghleh, Barkana and Faezipour (2015) showed that combining two independent classifiers based on the hierarchical method to classify the skin lesions into three categories can improve the rate of skin cancer diagnosis. According to this method, first, melanoma and non-melanoma lesions were distinguished from each other on the first classifier and then, on the second classifier, two benign and dysplastic lesions were classified. Therefore, in the present study, instead of classifying the skin lesions in one level, a two-level hierarchical approach according to the method by Abuzaghleh et al. (2015), was introduced. The difference between the proposed method in the present study and the method by Abuzaghleh et al. (2015), lies in the fact that instead of using an independent classifier at each level, the ensemble of the heterogeneous classifiers based on the stacking method was used to provide a good and stable performance to classify skin lesions into the three categories of benign, dysplastic and melanoma.

3. Datasets

In the present study, the PH^2 dataset (Mendonca, Celebi, Mendonca & Marques, 2015), and the presented dataset in Ganster et al. (2001), were used to implement and evaluate the proposed method. The Dermoscopic images in PH^2 were obtained under the same conditions using a magnification of 20x. This dataset contains 200 dermoscopic images of lesions, including 80 Benign moles, 80 Dysplastic moles and 40 Melanomas. They are 8-bit RGB color images with a resolution of 768 × 560 pixels.

The Ganster Dataset (Ganster et al., 2001), also has over 5300 dermoscopic images of the three types of Benign, Dysplastic and Melanoma lesions provided by the department of dermatology at Vienna general hospital. In this department, the imaging is performed by a hand-held CCD camera (one chip color sensor) that is combined with an epiluminescence microscope in order to produce digitized ELM images of skin lesions. The images have a spatial resolution of 632 × 387 pixels. To balance the data, 270 images of this dataset including 100 benign samples, 100 dysplastic samples and 70 malignant samples were selected. The information of the two datasets is summarized in Table 4. Fig. 3 shows three examples of the images in these two datasets.

Table 3
Summary of related works for skin cancer classification.

Author (s)	Features	Feature Selection	Classification	Result
Garnavi et al. (2012)	Wavelet tree-Based Texture and Boundary Series analysis	Gain Ratio Feature Selection (GRFS)	Random forest (Melanoma vs Non-melanoma)	91.26% Accuracy
Rastgoo et al. (2015)	Local and Global features	–	RF,GB,SVM (Melanoma vs Dysplastic)	98.46% Sensitivity
Pathan et al. (2018)	Color, Texture and Geometric features from Pigmented Networks	–	SVM, ANN (Melanoma vs Non-melanoma)	96.7% Accuracy
Oliveira et al. (2017)	510 features based on the Shape, Color and Texture	PCA, Relief-f, GRFS, Correlation Coefficient, CFS and Information Gain	Ensemble OPF classifier and feature manipulation (Melanoma vs Non-melanoma)	94.3% Accuracy
Schaefer et al. (2014)	Geometric features, Statistical features of Color and Texture analysis	Correlation based Feature Selection (CFS)	Ensemble SVM Classifier and Combinig votes by Neural Network (Melanoma vs Non-melanoma)	93.76% Sensitivity
Oliveira et al. (2016)	Shape Asymmetry, irregularity Border, Color variety and Texture analysis	–	SVM Classifier with Adapted Histogram Intersection kernel (Nevus–Melanoma)	74.36% Accuracy
Xie et al. (2016)	Boundary, Color and Texture features	PCA	Neural Network Meta-Ensemble Model (Melanoma vs Non-melanoma)	94.1% (Dataset1) and 91.11% (Dataset2)

Table 4
Information of two datasets used in this research.

Dataset	Total Images	Selected Images	Images Details	Images Classes (#)
PH^2	200 Sample	200	RGB 768 × 560 pixels	Benign (80) Dysplastic (80) Melanoma (40)
Ganster	Over 5300 Sample	270	RGB 632 × 387 pixels	Benign (100) Dysplastic (100) Melanoma (70)

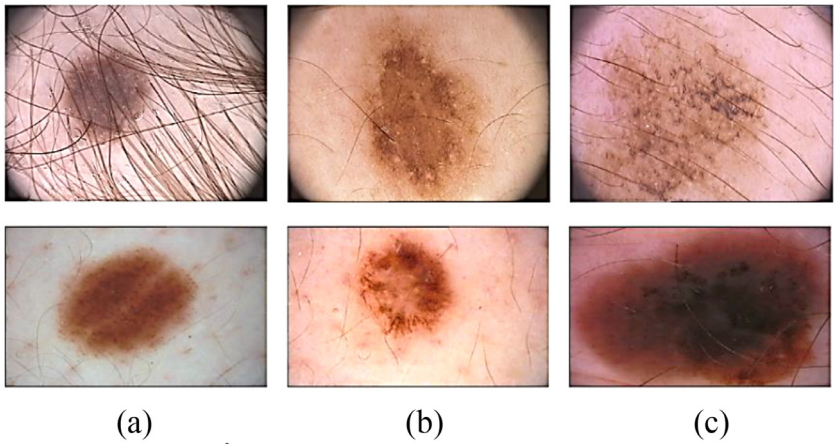

Fig. 3. Three Sample of images in the PH^2 Dataset (Mendonca et al., 2015), (up) and the Ganster Dataset (Ganster et al., 2001),(bottom): a) Benign; b) Dysplastic; c) Melanoma.

4. Proposed method

In order to implement the proposed method, the steps that indicated in Fig. 2 were followed. First, the algorithm for preprocessing of the dermoscopic image is presented. Second, the segmentation of the lesion is provided. Then, the different methods of feature extraction based on Shape, Color and Texture are described. Finally, the proposed method for classifying skin lesions based on the combination of the classifiers has been stated.

4.1. Preprocessing

The first step in this section is to improve the contrast in the image. To do so, first, the pixels of the image in each of the three red, green and blue channels are scaled to range between zero and one. As a result of this, dark pixels take an intensity close to zero and bright pixels close to one. Then, the resulting improved image is obtained by squaring the pixels value. This makes dark areas darker and bright areas brighter. In the next step, the image is filtered using a Median filter with 25 × 25 window dimensions and is subtracted from the original image to produce a high-pass filter image. The dimensions of the Median filter window are selected experimentally and according to visual quality as well as the processing time. If the dimensions of the median filter window are small, the hairs will not be removed from the image well, and with an increase of the window size greater than 25 × 25, processing time will increase, but will not affect the image quality of the output.

The high-pass filtered image includes details of the image that includes the hairs. Given that hair pixels have values less than zero in high-pass filtered image, so the high-pass filtered image is thresholded to keep pixels smaller than zero and remove other pixels. Then, for the detection of hairs in the image, considering the characteristics of hair such as Deviation, Length and Structure, several linear structural elements with the length of 30 every 15° (from 0 to 180) using the Top-Hat transform on the reverse of high-pass filtered image is applied. The resulting image, on which the hairs are distinct from other areas, becomes a binary image via the Otsu thresholding method. Finally, by removing the small regions with areas less than 300 pixels from the image, the final binary mask of the hair formed. The entire above-described process is repeated on two other channels of RGB color space, and to remove the hairs in the image, inpainting method presented by Agrawal, Sinha, Kumar and Bagai (2015), based on Partial Derivative Equations (PDE), is modeled. This method, based on Laplace Equation, determines the desired pixel value by its four neighborhoods in the above, below, left, and right based on Eq. (1).

$$u_{i,j} - 0.25 \times (u_{i-1,j} + u_{i+1,j} + u_{i,j-1} + u_{i,j+1}) = 0 \tag{1}$$

In the above equation, u is the intensity of the pixel and i and j represent the location of the pixels. The steps mentioned for the preprocessing are summarized in Eqs. (2) to (11). Fig. 4 illustrates the complete steps for implementing the preprocessing algorithm, and the obtained images from each step for the Red, Green and Blue channel of the sample image.

$$I(R,G,B)_{(0,1)} = \frac{I(R,G,B) - \min(I(R,G,B))}{\max(I(R,G,B)) - \min(I(R,G,B))} \tag{2}$$

$$R,G,B_{\text{Channel Conttast}} = I(R,G,B)_{(0,1)} \wedge 2 \tag{3}$$

$$\text{Median Filter}(R,G,B_{CC}) = \text{medfilt2}(R,G,B_{CC}, [25\ 25]) \tag{4}$$

$$\text{High pass Filter}\ (R,G,B_{CC}) = R,G,B_{CC} - \text{Median Filter}(R,G,B_{CC}) \tag{5}$$

$$\text{Pixels Smaller than Zero}_{R,G,B} = \text{High pass Filter}\ (R,G,B_{CC})_{(i,j)} < 0 \tag{6}$$

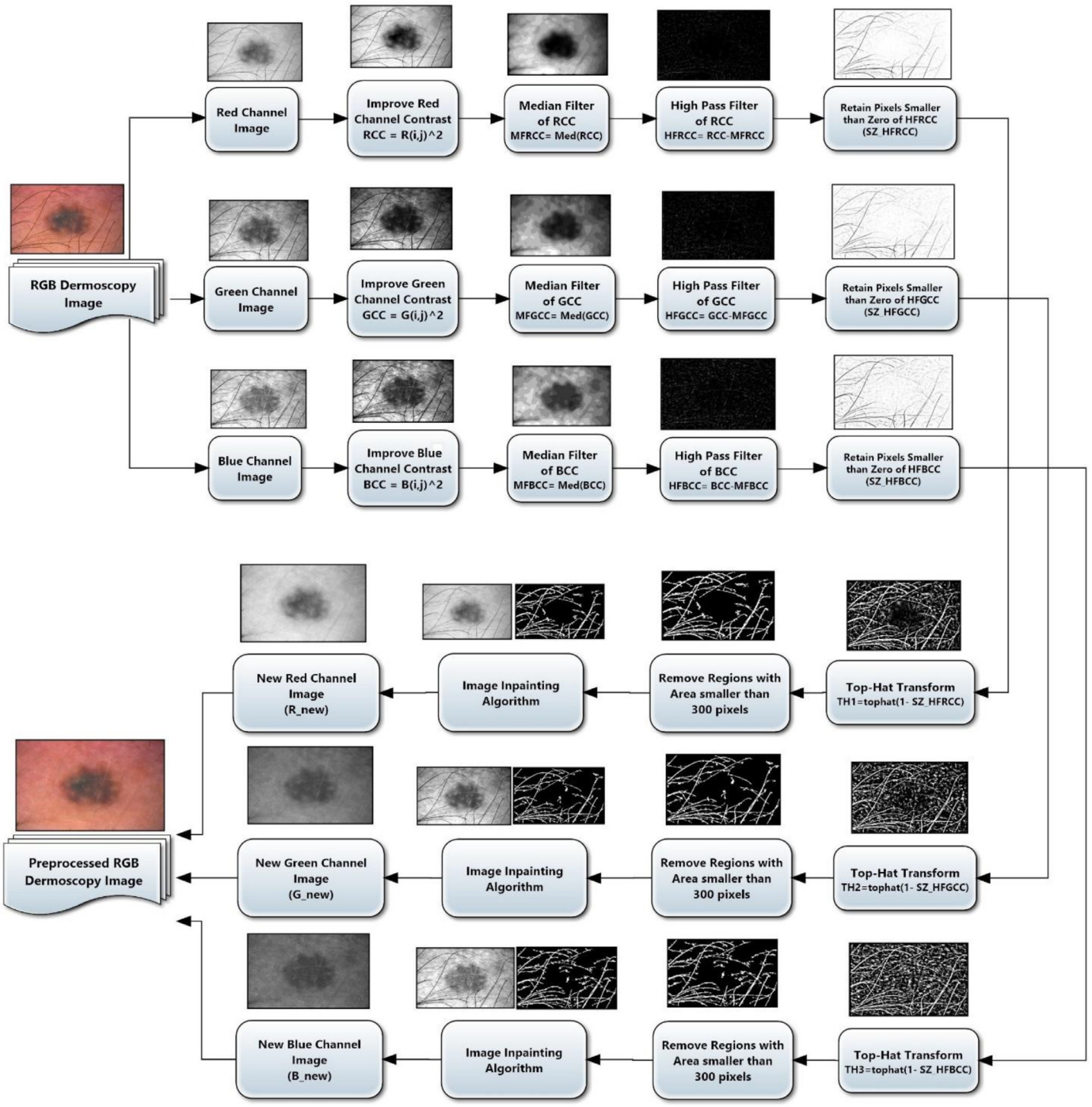

Fig. 4. Preprocessing Algorithm Steps for removing Hairs in Dermoscopy Images.

$$\text{Reverse Image}_{R,G,B} = 1 - \text{Pixels Smaller than Zero}_{R,G,B} \tag{7}$$

$$\text{Top} - \text{Hat transform Image}_{R,G,B} = \text{tophat}(\text{Reverse Image}_{R,G,B}, b) \quad (\text{tophat}(f, b) = f - f \circ b) \tag{8}$$

$$R, G, B_{(bw)} = \text{Otsu}(\text{Top} - \text{Hat transform Image}_{R,G,B}) \tag{9}$$

$$\text{bw\_Mask}_{R,G,B} = \text{Regions Area}(R, G, B_{(bw)} > 300) \tag{10}$$

$$R, G, B_{(New)} = \text{inpainting}\left(I(R, G, B)_{(0,1)}, \text{bw\_Mask}_{R,G,B}\right) \tag{11}$$

The above equations (R, G, B), express that operations apply individually to the Red, Green and Blue channels of input RGB image. In Eq. (8), f is a gray scale image, $\circ$ denotes the opening operation, and b is the structure element.

The required time to run the proposed algorithm by a system with 4 GB of RAM and a Dual-Core processor using MATLAB 2016 for an image of Ganster dataset was 7.25 s and for the PH^2 dataset image was recorded in 14.39 s.

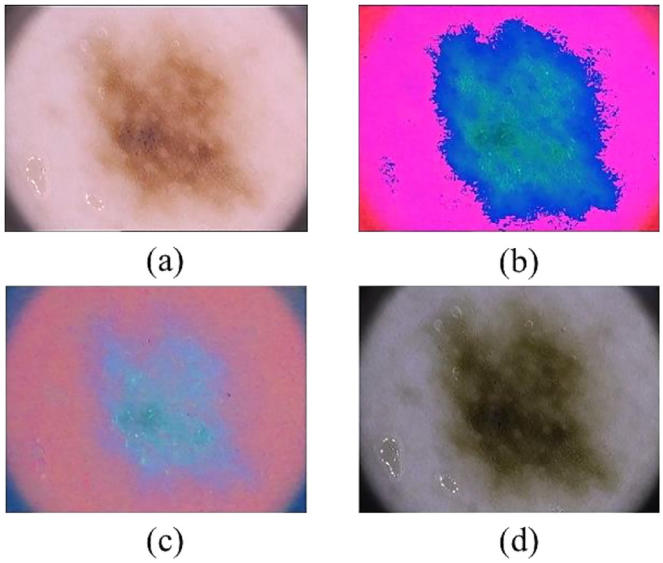

Fig. 5. Different Color Spaces for a dermoscopy image: a) RGB color space; b) HSV color space; c) CIEl*a*b color space; d) XYZ color space (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.).

4.2. Segmentation

After applying the Preprocessing, lesion segmentation was initiated with the aim of separating the area of the lesion from healthy skin areas. So, first, according to Eq. (12), the intensity of each pixel in the red channel was replaced by its average value in three Red, Green and Blue channels. Then, to improve the contrast in the image, by the imadjust function, the intensity range of the pixels was linearly mapped from a primary range to a secondary range.

$$I_{new}(i,j,r) = \frac{I(i,j,r) + I(i,j,g) + I(i,j,b)}{3} \tag{12}$$

In the equation above, I, is the original image and I_{new} is the improved image. Using the RGB color space for the segmentation of images does not seem appropriate due to disadvantages such as correlations between its channels. Therefore, the use of other color spaces such as HSV, CIEl*a*b and XYZ was considered for the segmentation of lesions (Kaur & Joshi, 2015; Rastgoo et al., 2015). Here, the HSV color space is used because it provides a better visual perception and a good distinction between the lesion area and the healthy skin. Fig. 5 shows the comparison of different color spaces for a sample image. The HSV color space consists of three components of Hue, Saturation and Value. After color space transforming, the saturation and reverse of the Value channel were selected from the HSV image and were thresholded by the Otsu method. In order to remove the additional areas, the morphological operators of Erosion and Dilation with the structural element of the Disk with dimensions of 4 and 25 were used, respectively, to obtain an estimate of the area of the lesion. After that, the holes were filled in the image and finally, a binary mask was applied to the image to eliminate the Dark Corners created during imaging. To determine the final boundaries of the lesion, the segmented images of the two channels were merged and the edges were smoothed with the help of a median filter with dimension of 8 × 8. This dimensions were determined empirically with respect to the boundaries of the lesions. Fig. 6 shows an example of the detected boundaries by the proposed algorithm. The required time to segment an image based on the proposed algorithm was recorded in about one second.

4.3. Feature extraction

To analyze the area of the lesion, three types of features based on Shape, Color and Texture were extracted from the dermoscopy images for the classification task. To extract the Shape-based features, the binary segmented image of the lesion was analyzed. Fig. 7 shows two samples of the binary segmented images for benign and malignant tumors. The features used to describe the shape features are: Lesion Area, Greater Diameter, Circularity Index, Asymmetry, Aspect Ratio, Eccentricity, Elongation, Solidity, Convexity, Rectangularity Index and Equivalent Diameter.

Complex Networks Discriptors (CND) were used to analyze the boundaries of the lesion. Complex Networks have been introduced in the past decade by combining concepts such as Graph theory and Statistics. Graph theory is used to model a shape in the form of Complex Networks. Consider S as the contour of an image, where $S = (s_1; s_2; \ldots; s_N)$ and s_i are typical vectors in the form of $s_i = (x_i; y_i)$, whose components are discrete numerical values representing the coordinates of point i of the contour. In order to apply the Complex Networks theory to the problem, a representation of contour S should be built like graph $G =)V; Ei(.$ Each pixel of the contour is represented as a vertex in the network (i.e., $S = V$). A set of non-directed edges E connecting each pair of the vertices makes up the network. This set E is calculated using the Euclidean distance (Scabini et al., 2017):

$$d(s_i, s_j) = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \tag{13}$$

Therefore, the network is represented by $N \times N$ weight matrix W and in order to scale the invariant of weights, the weight array W was scaled to values between 0 and 1 according to Eq. (14) (Backes, Casanova & Bruno, 2009).

$$w_{ij} = W\left(\left[w_i, w_j\right]\right) = d(s_i, s_j)$$
$$W = \frac{W}{\max_{w_{ij} \in W}} \tag{14}$$

All the nodes in the created network are interconnected, which is why they display a regular behavior that contradicts the concept of a complex network and does not provide specific information about the structure of the shape. Therefore, in order to achieve a complex network, it is a desirable method of applying threshold on the edges of a regular network. Accordingly, a group of edges with weights less than the threshold are selected according to Eq. (15).

$$E^t = \left\{\left(s_i, s_j\right) | W_{v_i,v_j} \le T_L, s_i \, and \, s_j \in S\right\} \tag{15}$$

Using the thresholded network, the shape can be described by a feature vector derived from different values of the T_L. Therefore, the threshold of T_L increases at regular intervals from T_{ini} to the T_Q. The above desirable values, according to the experiments performed in Backes et al. (2009), ranged from 0.025 to 0.95, with a distance of 0.075. From the created networks, a feature vector based on degree descriptors was extracted. The degree for a node is equal to the number of edges that are directly connected to that node. This feature was calculated from the adjacency matrix A according to Eq. (16). Here, matrix A is the thresholded matrix.

$$k_i = \sum_{j=1}^{N} A_{ij} \tag{16}$$

In the above equation, N is the number of vertices in the network. The degree features of the matrix A for each threshold were calculated using the mean and maximum criteria based on Eqs. (17) and (18).

$$k_k = \max(k_i) \tag{17}$$

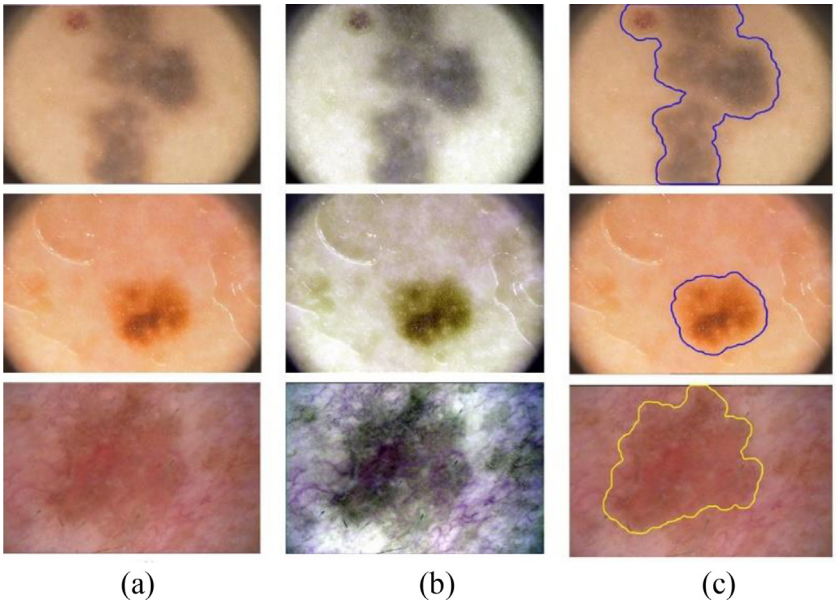

Fig. 6. Detected Boundaries from Segmentation Algorithm: a) Input Dermoscopy Image; b) Contrast Enhancement; c) Detected Boundary of skin Lesion.

Fig. 7. Binary Mask of Lesion: a) Melanoma; b) Benign.

$$k_\mu = \frac{1}{N}\sum_{i=1}^{N} k_i \tag{18}$$

The final feature vector is created by combining the mean and maximum features at different thresholds. This vector can be considered as a shape signature, which in addition to the higher differentiation power than other shape descriptors, such as Fourier descriptors and Zernike moments, has the desired properties such as resistance to scaling, rotation, and noise (Backes et al., 2009).

- Color analysis of the lesion is one of the important characteristics in the diagnosis of melanoma, which has been considered by dermatologists. A malignant lesion has often more color variations than benign type. This color variation may appear in the form of shades of black, brown, and tan. Also, the lesion may be white, gray, red, pink or blue (Satheesha, Satyanarayana, Prasad & Dhruve, 2017). For the reasons mentioned, use of color information in skin cancer detection systems is common. Here, the statistical measures of mean, variance, standard deviation and skewness from each channel of color spaces RGB, HSV, CIE l*a*b and OPP were considered as features. For the analysis of the uniform distribution of color in the area of the lesion as well as the amount of color changes in the lesion relative to the color of the healthy skin around it, the average ratios were used according to Eqs. (19) and (20) (Satheesha et al., 2017).

$$\frac{\mu R}{\mu G}, \frac{\mu R}{\mu B}, \frac{\mu G}{\mu B} \tag{19}$$

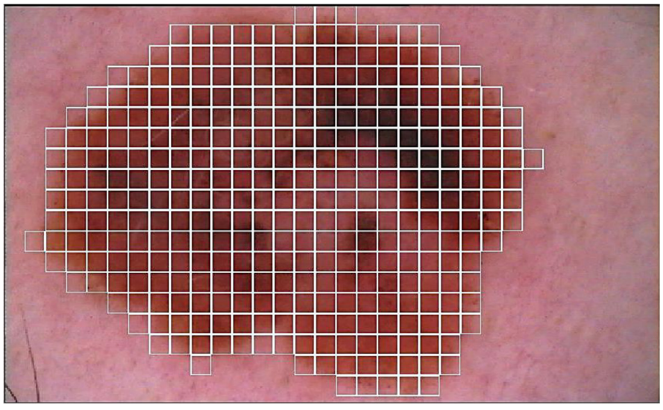

Fig. 8. Patching the area of the lesion into blocks of 20 × 20.

$$\frac{\mu R}{\overline{\mu R}}, \frac{\mu G}{\overline{\mu G}}, \frac{\mu B}{\overline{\mu B}} \tag{20}$$

In Eq. (20), $\overline{\mu}_{R,G,B}$ is the mean value of the color in the areas around the lesion.

- The use of Gray Level Co-occurrence Matrix (GLCM) is a common method for analyzing lesion texture. This matrix was first introduced by Haralick in 1973 (Haralick & Shanmugam, 1973). The GLCM elements represent the number of pairs of pixels that occur with a specific gray level at certain intervals and angles. Dimensions of this symmetric matrix are equal to the number of gray levels of the image. In this paper, the 14 features presented in Haralick and Shanmugam (1973) were extracted from GLCM matrix in four directions of 0, 45, 90 and 135° as texture features. These features include: Angular Second Moment, Contrast, Energy, Correlation, Variance, Local Homogeneity, Sum Average, Sum Entropy, Sum Variance, Entropy, Difference Variance, Difference Entropy and Information measures of Correlation1,2.

To extract the above features, the area of the lesion of each image was first blocked into 20 × 20 patches, as shown in Fig. 8.

This dimensions are determined based on the values stated by Kasmi and Mokrani (2016). Then, from all of these blocks, the GLCM features were extracted and the mean value of each of them was considered as the final feature.

In total, 162 features were extracted from the lesion area of each image, which are:

11 features of shape geometric analysis, 26 features from the lesion boundaries analysis by Complex Networks, 69 features of lesion color analysis, 56 features of lesion texture analysis by GLCM. Due to the large dimensions of the feature vector and the possibility of existence of redundant features, the Feature Selection process was used which is described in the next section.

4.4. Feature selection

Choosing the optimal features and reducing the dimensions of the feature vector play an important role in the classification of data. Dimension reduction to decrease the computational cost and reduction of noise to improve the classification accuracy are the advantages of the feature selection process (Ding & Peng, 2005). In general, feature selection approaches are divided into three categories based on Filter, Wrapper and Eembedded methods. Filter-based methods select the features independent of the learning algorithm. They rely on various measures of general characteristics of training data such as Distance, Information, Dependency, and Consistency (Novovičová, Somol, Haindl & Pudil, 2007). On the contrary, the Wrapper FS methods require one predetermined learning algorithm and use its classification accuracy as performance measure to evaluate the quality of selected features set. These methods tend to give superior performance as they find features better suited to the predetermined learning algorithm.They also tend to be more computationally expensive (Chandrashekar & Sahin, 2014). Embedded methods want to reduce the computation time taken up for classifying different subsets which is done in wrapper methods. The main approach is to incorporate the feature selection as a part of the training process (Chandrashekar & Sahin, 2014).

- **Feature Selection based on Mutual Information** is a filter based approach. In information theory, mutual information is one of the most well-known criteria for the expression of dependency. For discrete categorical variables, the mutual information of two variables x and y is defined based on their joint probabilistic distribution P(x, y) and the respective marginal probabilities p(x) and p(y):

$$I(x,y) = \sum_{i,j} p(x_i, y_j) \log \frac{p(x_i, y_j)}{p(x_i)\, p(y_j)} \tag{21}$$

This criterion is a powerful tool for measuring the relationship between random variables that can be considered as a mathematical solution to find and select optimal features. Accordingly, the method of Maximum Relevance and the Minimum Redundancy (mRMR) is proposed by Peng, Long and Ding (2005). The purpose of this method is to select the features with the most relevance to their target. On the other hand, because various features associated with the target may have additional information or have a great deal of dependency on each other, their removing does not change the classification work. Therefore, the redundancy criterion needs to be considered in the feature selection process. Therefore, in order to consider these two criteria in the feature selection process, the Eq. (22) was defined (Ding & Peng, 2005).

$$\Phi(X_S, Y) = \frac{1}{|S|} \sum_{i \in S} I(X_i; Y) - \frac{1}{|S|^2} \sum_{i,j \in S} I(X_i; X_j) \tag{22}$$

In the above relation, the first sentence calculates the average of the Relevance of the features to the target, and the second sentence measures the mean of the dual redundancy among the selected attributes. Therefore, maximizing $\Phi(X_S,Y)$, identifies a sub-set of desirable features.

- **Sequential Feature Selection** is the wrapper-based method that chooses a set of features that perform best performance based on the classification accuracy criterion. This method has a high computational complexity and has a very low speed. However, the advantage of this method is to provide a high classification accuracy. This approach chooses Best features based on two Forward and Backward algorithms. The Sequential Forward Feature Selection (SFFS) algorithm starts with an empty set and adds one feature for the first step which gives the highest value for the objective function. From the second step onwards, the remaining features are added individually to the current subset and the new subset is evaluated (Chandrashekar & Sahin, 2014). A Sequential Backward Feature Selection (SBFS) algorithm can also be constructed which is similar to SFFS but the algorithm starts from the complete set of variables and removes one feature at a time whose removal gives the lowest decrease in the predictor performance (Chandrashekar & Sahin, 2014). In this paper, the Backward method was used to select optimal features.

4.5. Classification

The purpose of data classification is to divide the feature space into labeled categories separated by decision boundaries. Often, an independent classifier is used to classify skin lesions. However, studies have shown that a better performance can be achieved if different classifier's votes are combined (Schaefer et al., 2014). Various methods are proposed for combining classifiers that are generally divided into two categories of Weighting and Meta-learning methods (Rokach, 2010). Weighting methods are useful if the Base-Classifiers perform the same task and have comparable success. Meta-learning methods are best suited for cases in which certain classifiers consistently correctly classify, or consistently misclassify, certain instances (Rokach, 2010). The general structure for combining the classifiers is shown in Fig. 9.

In this paper, the Stacking method was used to combine the base classifiers. Stacking is a Meta-learning based method. This approach, also known as Stack Generalization, was first introduced by Wolpert (1992). The classification based on this method consists of two steps, which in the first stage (level 0) base classifiers were trained using training data. From the first stage outputs, a new dataset was created that used to training the second stage classifiers (Level 1). Fig. 10 shows the structure of the Stacking method. A specific method for determining the number of classifiers in the first level is not defined. However, studies have shown that using the three classifiers as heterogeneous can provide a better performance than voting methods (Ahmed, Rasool, Afzal & Siddiqi, 2017). In the proposed structure for base classifiers training and also the production of the second dataset for training the Meta-Classifier, the following steps are followed (Aggarwal, 2014):

- Training the first level classifiers by original training data
- Creating new dataset based on Base Classifiers (BC) prediction
- Training the second level classifier (Meta-Classifier) based on the new dataset that created in the previous step

After training the Meta-Classifier, it can be used to combine prediction of the Base Classifiers. For an unknown instance of x, its predicted label is derived from the Stacking process through $h'(h_1(x),h_2(x), \ldots ,h_T(x))$, where $\{h_1(x),h_2(x), \ldots ,h_T(x)\}$ is the first stage classifiers and h' is classifier on second stage (Aggarwal, 2014).

In the above algorithm, the same dataset was used to train first-level classifiers and prepare the training data for the second-level classifiers, which may lead to over-fitting. To solve this

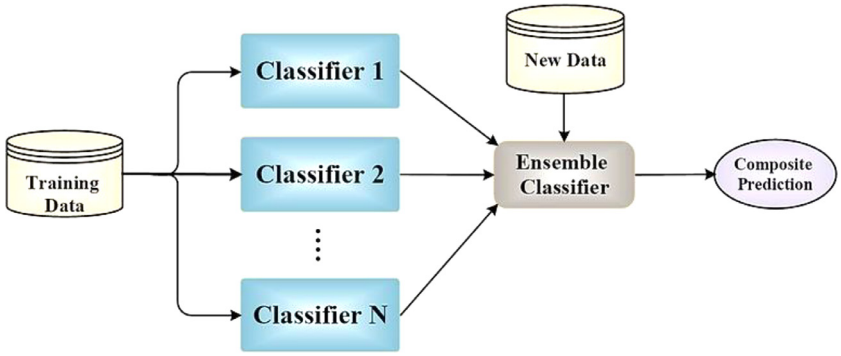

Fig. 9. Ensemble Classification.

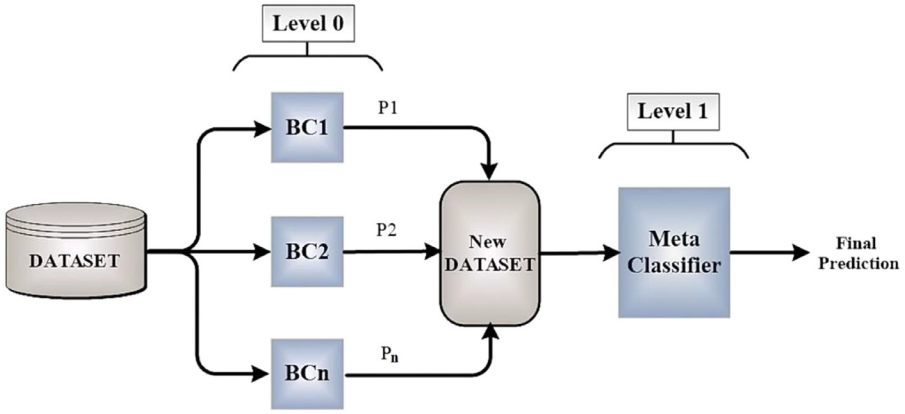

Fig. 10. Ensemble Classification Based on Stacking.

problem, we can incorporate the idea of cross validation in stacking. The idea of cross validation can be combined with Stacking to avoid using the same training set for building both first and second level classifiers (Aggarwal, 2014). Instead of using all the training examples to get first-level classifiers, we partitioned original data into K subsets and get K classifiers, each of which is trained only on $K - 1$ subsets. Each classifier is applied to the remaining one subset and the output of all the first-level classifiers constitutes the input feature space for the second-level classifier. Note that over-fitting is avoided because the data to train the first-level classifiers and the data to receive the predicted labels from the first-level classifiers are different. After we build the second-level classifier based on the first-level classifiers predicted labels, we can re-train the first-level classifiers on the whole training set so that all the training examples are used. Applying the second-level classifier on the updated first-level classifier output will give us the final ensemble output (Aggarwal, 2014).

In this paper, two approaches were proposed based on the Stacking algorithm for classifying skin lesions. In the first method, according to Fig. 10, the data were classified into three classes. The second method classified the lesions based on the hierarchical approach. Independent popular classifiers such as K-Nearest Neighborhood (KNN), Support Vector Machine (SVM), Elman Neural Network (ENN) and Multilayer Perceptron (MLP) Neural Network were used to create proposed hybrid approaches. Figs. 11 and 12 illustrate the using individual classifiers in the first and second stages for the proposed hybrid structures. The classification of data in Hierarchical Structure Based on Stacking (HSBS) is shown in Fig. 12. The figure shows that in the first stage Melanoma samples are distinguished from Non-melanoma and in the second stage, Benign and Dysplastic lesions are separated.

5. Experimental results and discussion

In this section, we used two PH^2 and Ganster datasets in order to evaluate the proposed methods. In our experiments, Five Fold Cross Validation was carried out for training and testing the proposed methods. Moreover, to determine the optimum dimensions of the feature vector, different numbers of selected features were examined. The derived results of the SBS method on the PH^2 dataset are presented in Table 5**.** In this table**,** classifying skin lesions into three classes of Melanoma, Dysplastic and Benign according to the two feature selection methods is presented. The obtained results show that the Stacking algorithm provides a better performance when 20 selected features of the mRMR method are used or 10 features of the SBFS mehtod are employed. The results from the base classifiers for each feature selection method are shown in Fig. 13. Separate results show that the SBS Ensemble method provides a better performance than the independent classifiers.

The details of the selected features for the SBS classification method are shown in Fig. 14. A review of the selected features suggests that color-based features have been considered in both feature selection methods, while texture features were only selected in the SBFS method. Additionally, Shape and Complex Networks features were selected more in the mRMR method. Therefore, it can be concluded that Color and Shape-based features have less redundancy than GLCM Texture features.

Next, the performance of the HSBS algorithm was evaluated in the same way as the previous method. In the HSBS method, the lesions classified into two classes of Melanoma and Non-melanoma, as well as Benign and Dysplastic moles classification, were separately evaluated according to various selected features.

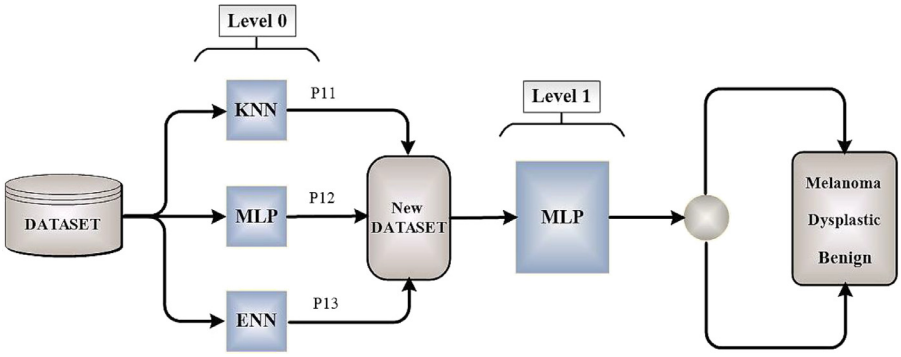

Fig. 11. Proposed SBS Ensemble method for Classification Skin Lesions to Three Classes.

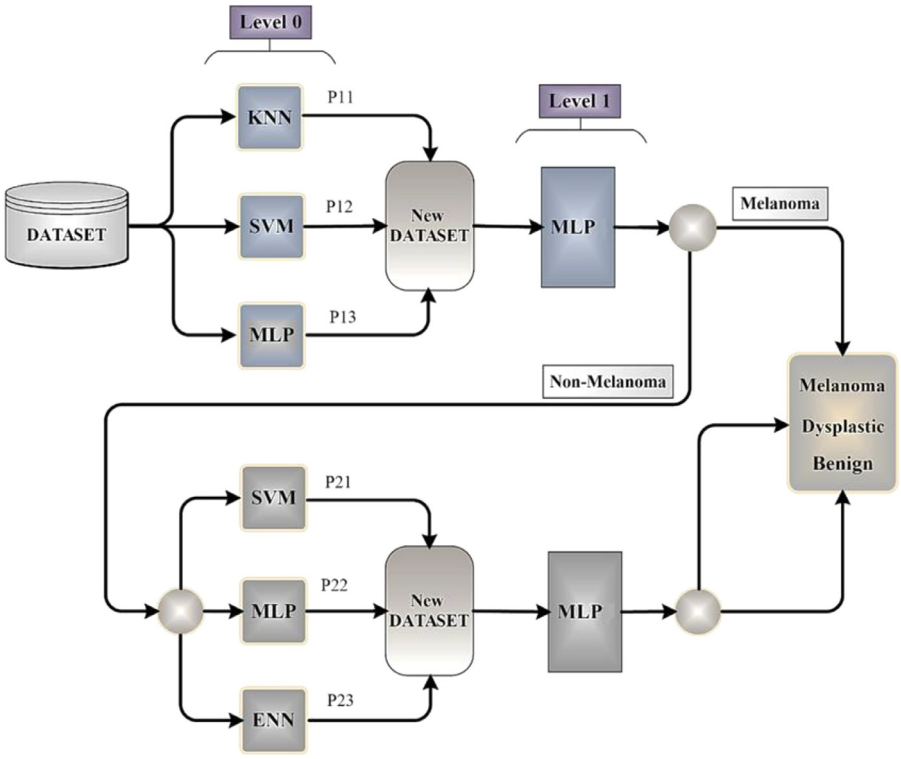

Fig. 12. Proposed HSBS Ensemble method for Classification Skin Lesions to Three Classes.

Table 5
Results of the SBS method to classify the Melanoma, Benign and Dysplastic lesions on PH^2 dataset.

No. Feautres	mRMR Method Evaluation Criteria			SBFS Method Evaluation Criteria		
	Accuracy%	Sensitivity%	Specificity%	Accuracy%	Sensitivity%	Specificity%
10	91.67	87.5	93.75	**92.3**	**88.5**	**94.35**
15	91.67	87.5	93.75	89.3	84	92
20	**92**	**88**	**94**	90.3	85.5	92.7
25	91.3	87	93.5	92	88	94

Tables 6 and 7 depict the obtained results of the classification of the Melanoma and Non-melanoma lesions as well as the Benign and Dysplastic lesions. The final results of the proposed method by integrating these two structures (Fig. 12) are presented in Table 8. The obtained results show that the Stacking structure has a good performance for the classification of melanoma and non-melanoma lesions, with an accuracy of 98.5%, sensitivity of 97.5%, and specificity of 98.75% based on the 20 selected features from both of the feature selection methods.

The results in Table 7 show that the using of the 20 selected features of the SBFS method provides a better performance for the classification of Benign and Dysplastic lesions compared to the mRMR method. The evaluation of the HSBS method shows that the proposed method has a much better performance compared to the presented approach in Fig. 11. Moreover, the comparison of the results of the two feature selection methods in Table 8 indicates that the 20 selected features of the SBFS method with 96% accuracy, 94% sensitivity, and 97% specificity have the

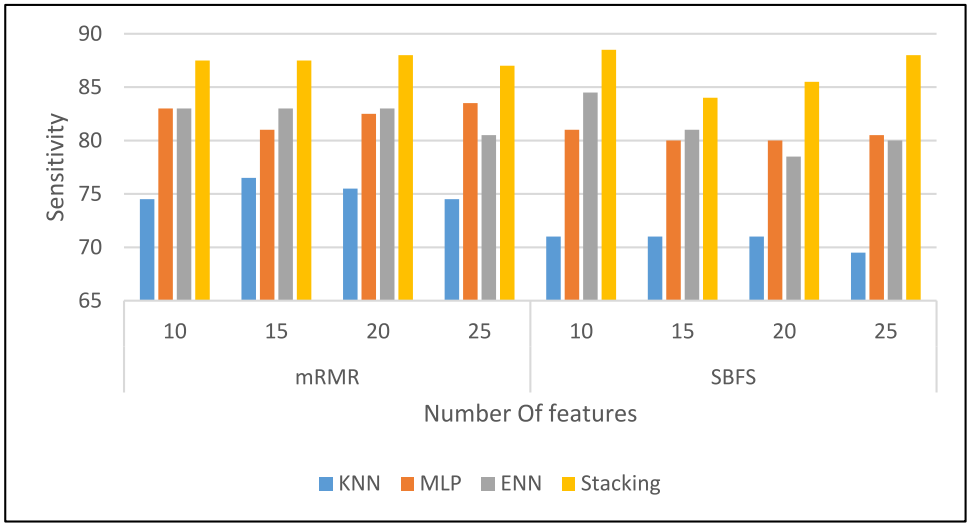

Fig. 13. Base Classifiers results in SBS Classification Method.

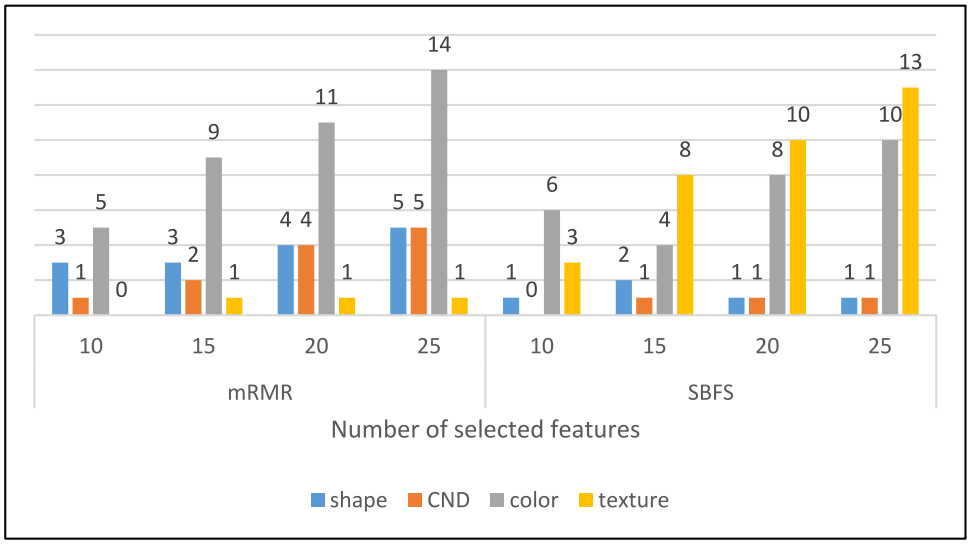

Fig. 14. Details of the selected features for the SBS Classification method.

Table 6
Classification results for classify the skin lesions to Melanoma and Non-melanoma on PH^2 dataset.

No. Feautres	mRMR Method Evaluation Criteria			SBFS Method Evaluation Criteria		
	Accuracy%	Sensitivity%	Specificity%	Accuracy%	Sensitivity%	Specificity%
10	97.5	95	98.13	97.5	92.5	98.7
15	98	92.5	**99.38**	97	85	**100**
20	**98.5**	**97.5**	98.7	**98.5**	**97.5**	98.75
25	98	95	98.7	96	82.5	99.4

Table 7
Classification results for classify the skin lesions to Benign and Dysplastic on PH^2 dataset.

No. Feautres	mRMR Method Evaluation Criteria			SBFS Method Evaluation Criteria		
	Accuracy%	Sensitivity%	Specificity%	Accuracy%	Sensitivity%	Specificity%
10	88.75	**93.75**	83.75	88.1	90	86.25
15	88.75	90	87.5	91.87	91.25	92.5
20	**91.25**	90	**92.5**	**94.4**	**93.75**	**95**
25	87.5	91.25	83.75	83.75	82.5	85

Table 8
Results of HSBS algorithm evaluation for classification of Melanoma, Benign and Dysplastic lesions on PH^2 dataset.

No. Feautres	mRMR Method Evaluation Criteria			SBFS Method Evaluation Criteria		
	Accuracy%	Sensitivity%	Specificity%	Accuracy%	Sensitivity%	Specificity%
10	92.33	88.5	94.25	92	88	94
15	92.67	89	94.5	93.67	90.5	95.25
20	**94**	**91**	**95.5**	**96**	**94**	**97**
25	92	88	94	88.67	83	91.5

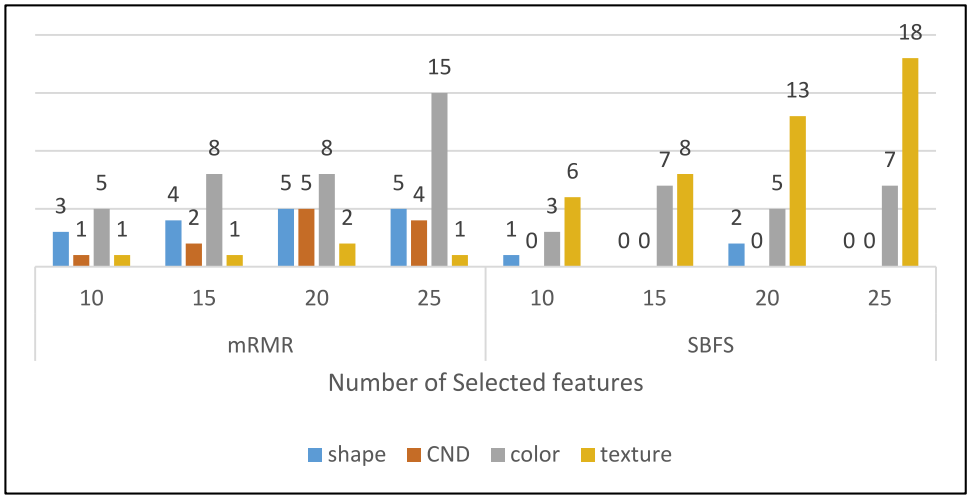

Fig. 15. Details of the selected features for the first level of HSBS Classification method (Melanoma and non-melanoma lesions).

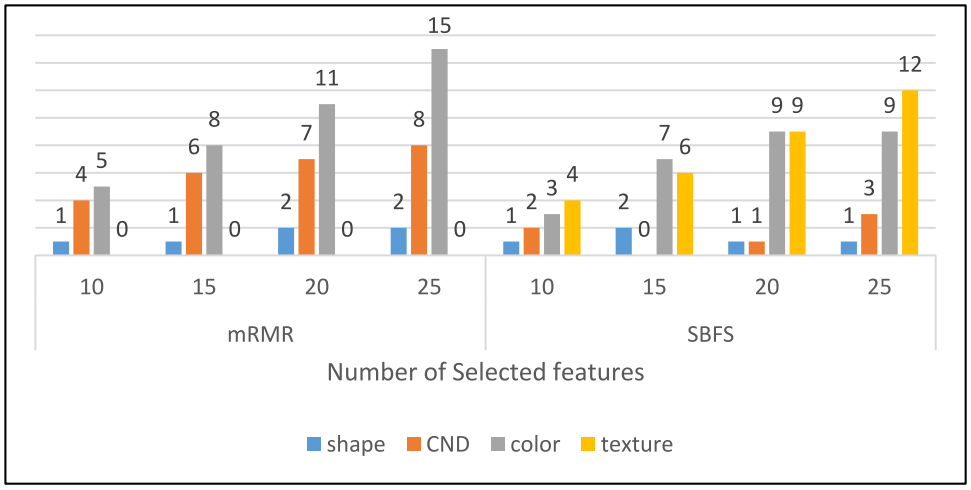

Fig. 16. Details of the selected features for Second level of HSBS Classification method (Benign and Dysplastic lesions).

most favorable effect on the improvement of the classification performance.

The details of the selected features for both levels of the HSBS classification method are shown in Figs. 15 and 16. According to the diagrams in Figs. 15 and 16, like the SBS method, the Shape and Color based features are more focused in the mRMR approach, while the texture features are more selected by the SBFS method. The results of first and second level base classifiers for the HSBS Ensemble method are shown in Figs. 17 and 18. The results show that the Stacking method has a better sensitivity than the independent classifiers in both stages of the HSBS algorithm. Figs. 19 and 20 show the effect of increasing the number of the selected features for the mRMR and SBFS methods for different approaches to the classification of skin lesions. Studying the charts from both feature selection methods indicates that by increasing the number of the selected features to a certain number, the classification accuracy improves in various methods of classification. Here, the desirable number of features for the classification of skin lesions in most cases is 20.

The degree of variation in individual base classifiers and Stacking Structure for Classifying melanoma and Non-melanoma based on the optimal number of features were investigated by using two feature selection methods during the Five Fold Cross Validation process by drawing the box plots that are presented in Figs. 21 and 22. The plotted graphs show that using the selected features of the mRMR method provides a more balanced performance than the

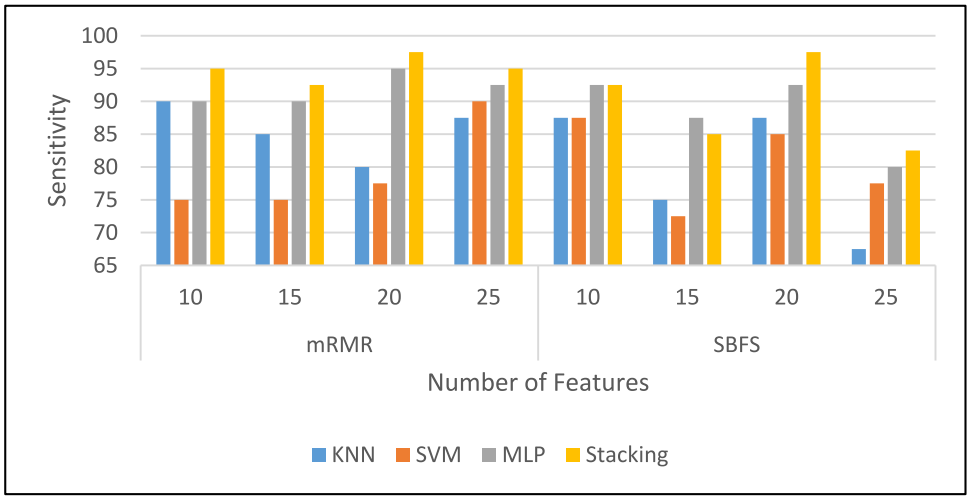

Fig. 17. Base Classifiers Sensitivity for First Level of HSBS Method for Classification the Melanoma and non-melanoma Lesions.

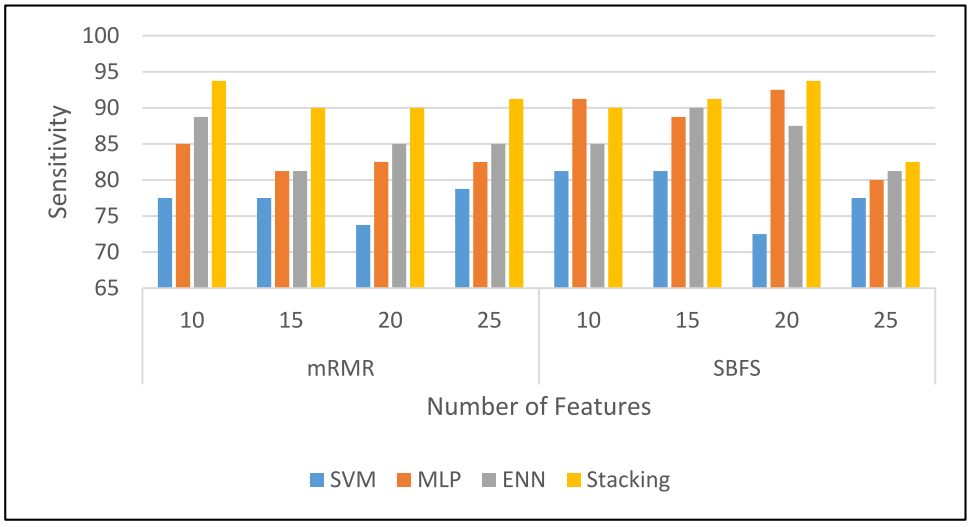

Fig. 18. Base Classifiers Sensitivity for Second Level of HSBS Method for Benign and Dysplastic Lesions Classification.

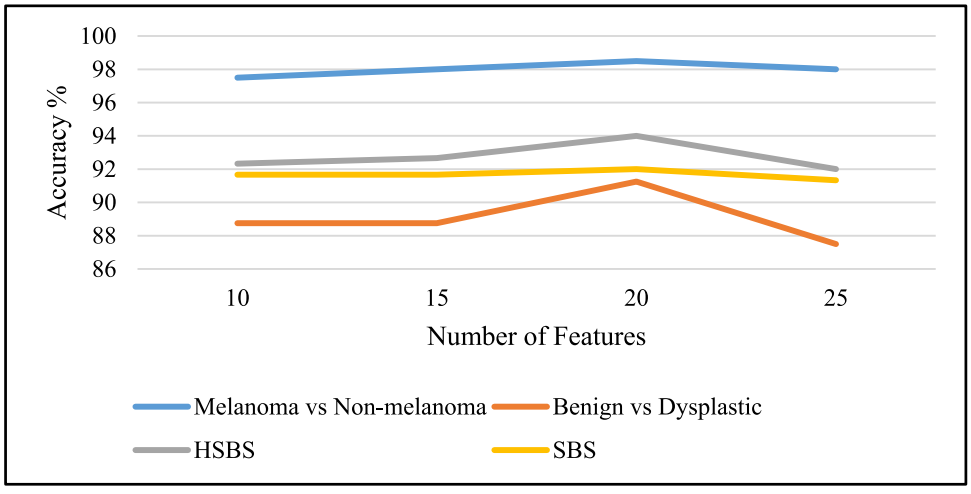

Fig. 19. Variation chart in Classification accuracy in terms of increasing the number of features based on the mRMR method for different Classification Modes.

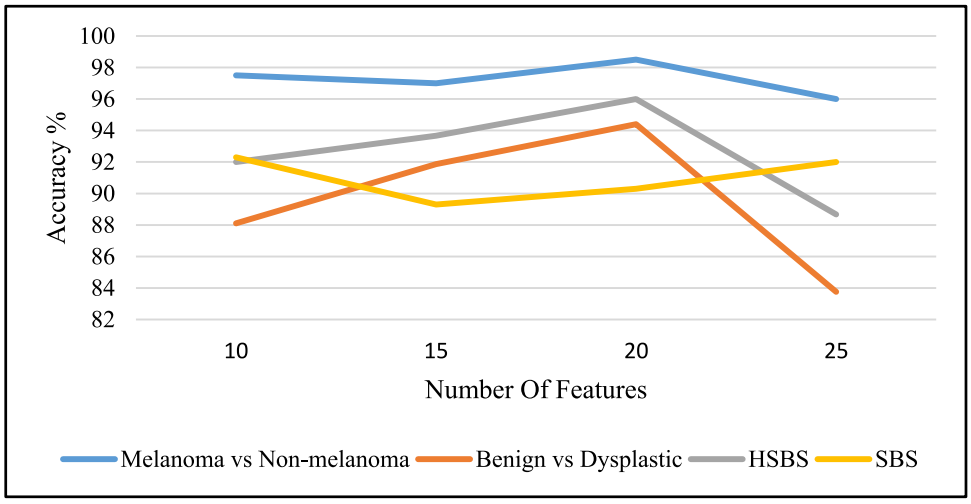

Fig. 20. Variation chart in Classification accuracy in terms of increasing the number of features based on the SBFS method for different Classification Modes.

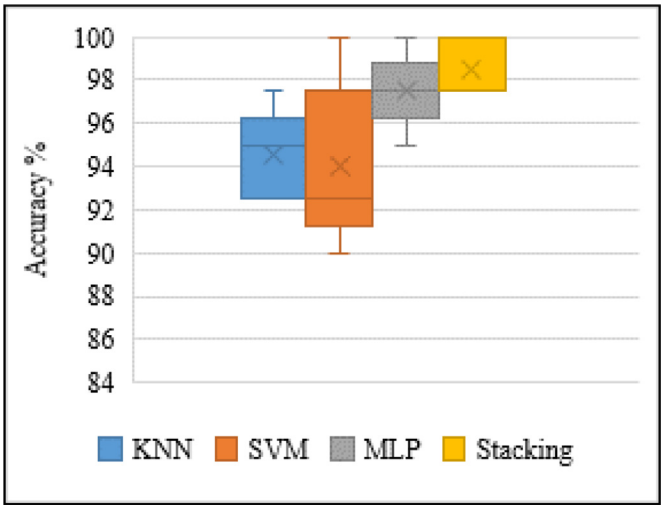

Fig. 21. Box Plots of performance variations of the different classification algorithms for the distinguish Melanoma from Non-melanoma Lesions based on the mRMR method.

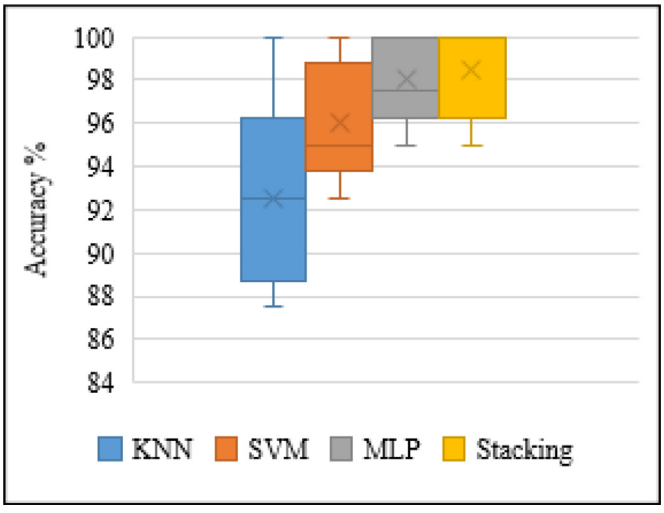

Fig. 22. Box Plots of performance variations of the different classification algorithms for the distinguish Melanoma from Non-melanoma Lesions based on the SBFS method.

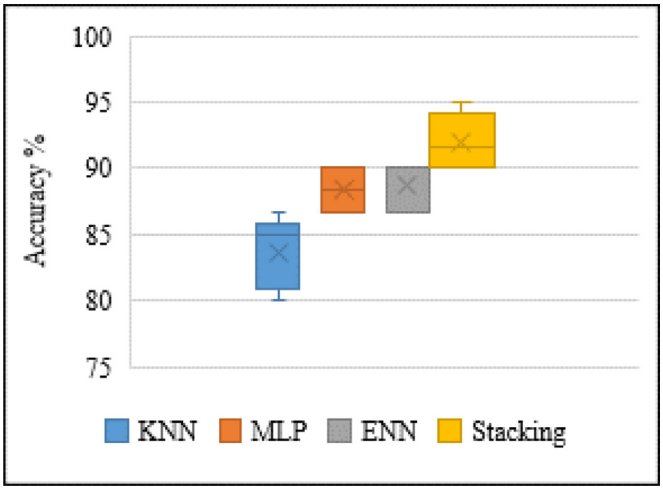

Fig. 23. Box Plots of performance variations of the different classification algorithms for the distinguish skin lesions to three classes based on the mRMR method.

SBFS method. This issue was also surveyed for skin lesions classification into the three classes according to Fig. 11. The drawen diagrams in Figs. 23 and 24 show that the selected features of the mRMR method, despite having a lower performance than the SBFS features, have a more stable performance.

In order to measure the generalizability of the proposed methods, the performance of the classification algorithms was evaluated by using the Ganster Dataset and based on the number of desirable features identified in the previous sections. The classification results for Melanoma and Non-melanoma lesions based on the 20 optimal features of the mRMR method were presented in Table 9 with an accuracy of 97.78%, sensitivity of 94.29%, and specificity of 99% that provides a better performance than the SBFS method. Additionally, the performance of the classification algorithm for the classification of the benign and dysplastic lesions show that the 20 optimal features based on the mRMR method and with 89.5% Accuracy, 90% Sensitivity and 89% Specificity yielded a better performance compared to the same number of features from the SBFS method. Finally, the performance of the proposed methods for Melanoma, Benign and Dysplastic lesions classification on the Ganster dataset was evaluated using the specified number of features. The results in Table 10 show that the HSBS classification algorithm by using the 20 selected features of the mRMR method

Table 9
Classification results of Stacking Method on Ganster dataset.

mRMR Method Evaluation Criteria			SBFS Method Evaluation Criteria		
Accuracy%	Sensitivity%	Specificity%	Accuracy%	Sensitivity%	Specificity%
Classification of Melanoma and Non-Melanoma Lesions					
97.78	94.29	99	97.41	95.71	98
Classification Of Benign and Dysplastic Lesions					
89.5	90	89	87	90	84

Table 10
Classification result of Skin Lesions to 3 classes in Ganster dataset.

mRMR Method Evaluation Criteria			SBFS Method Evaluation Criteria		
Accuracy%	Sensitivity%	Specificity%	Accuracy%	Sensitivity%	Specificity%
Structure Based on Stacking (SBS)					
87.9	81.85	90.9	89.1	83.70	91.85
Hierarchical Structure Based on Stacking (HSBS)					
93.33	90	95	92	88.15	94.07

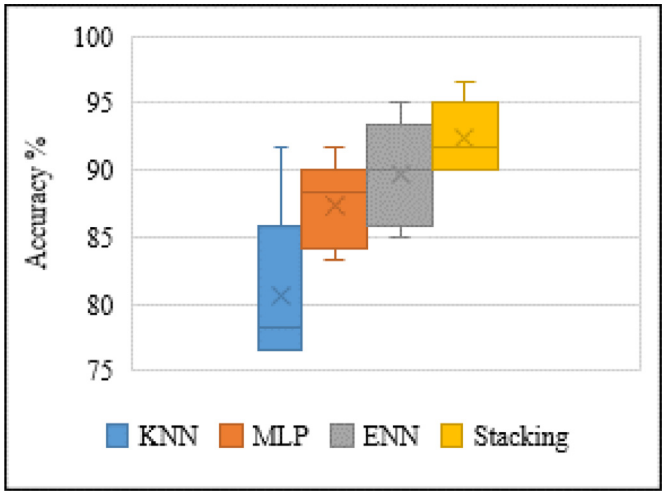

Fig. 24. Box Plots of performance variations of the different classification algorithms for the distinguish skin lesions to three classes based on the SBFS method.

yielded a better performance (an accuracy of 93.33%, sensitivity of 90%, and 95% specificity) compared to the selected features of the SBFS method. The results of the two classification approaches (SBS and HSBS) based on the two different datasets (PH^2 and Ganster) indicate that the proposed method has a good generalizability in the diagnosis of skin lesions.

The comparison of the results of the proposed method in this study with other related studies in recent years is presented in Table 11. The obtained results show that the proposed method has a better performance for the classification of skin lesions into melanoma and non-melanoma compared to most other similar works on the PH^2 dataset. In addition, the proposed method in this study achieved a comparable performance with the presented method by Alfed and Khelif (2017) with less than 0.3% difference in accuracy. However, the proposed method uses feature vectors with smaller dimensions, which significantly reduces the computational complexity. Moreover, the proposed hierarchical structure in classifying skin lesions into melanoma, dysplastic and benign has yielded favorable results. However, the dysplastic lesion has not been addressed in the other studies mentioned in Table 11. In addition, the results obtained from testing the proposed method on the second dataset showed that when the data change, the proposed method provides a stable and desirable performance.

6. Conclusion

In this study, an automatic skin cancer diagnostic system was developed. First, depending on the different stages of a diagnostic system, the pre-processing algorithm was applied to the dermoscopic images based on the Top-Hat Transform and the PDE inpainting methods in order to remove hair from the dermoscopic images. Then, the lesion area was segmented from other areas by the Otsu thresholding and HSV color space. Next, different types of features were extracted from the lesion area based on shape, complex networks, color and texture. Then, with respect to the large dimensions of the feature vector, two feature selection approaches were used based on Filter and Wrapper methods. Finally, a combination of classifiers were used instead of an independent classifier in order to classify the skin lesions. Thus, the Structure Based on Stacking method (SBS) was considered to combine the base classifiers. In order to improve the performance of the proposed method, a Hierarchical Structure Based on Stacking approach (HSBS) was introduced. According to this method, melanoma and non-melanoma lesions were distinguished from each other in the first stacking structure and then, in the second stacking structure, benign and dysplastic lesions were classified.

The proposed methods were implemented on PH^2 and Ganster Datasets by using the different numbers of the selected features from the two feature selection methods. The obtained results showed that the Stacking method using the 20 selected features derived from the Mutual information method had a good performance in the diagnosis of melanoma from non-melanoma with an accuracy of 98.5% for the PH^2 Dataset and 97.78% for the Ganster Dataset. The selected features of the mRMR method are mostly based on shape, complex networks, and color features, and few texture features were selected (Fig. 14). This suggests that the GLCM features had more redundancy than other features. The existence of redundancy in the features causes the classification algorithm to be unstable in the event of a change in the data.

To classify the lesions into benign, dysplastic and melanoma categories, the two classification methods of SBS and HSBS were evaluated. The results showed that the proposed HSBS algorithm using the 20 selected features from the SBFS method, obtained 96% accuracy for the PH^2 Dataset. However, the evaluation of the proposed algorithms using the Ganster Dataset showed that the 20 selected features of the mRMR method provided a better accuracy than the SBFS method. This fact suggests that using the selected features based on the Filter method, independent

Table 11
Comparative analysis between the results of different methods and proposed method for skin lesion classification.

Author (s)	No. Images	No. Features	Diagnosis	Acuuracy %	Sensitivity %	Specificity %
Xie et al. (2016)	PH^2+EDRA (200 + 160)	22	Melanoma Non-Melanoma	94.17	95	93.75
Alfed et al. (2017)	PH^2 200	–	Melanoma Non-Melanoma	98.79	99.41	98.18
	Dermofit 256			92.96	84.78	96.04
Barata, Celebi and Marques (2017)	PH^2 200	–	Melanoma Non-Melanoma	–	100	88.2
Ruela, Barata, Marques and Rozeira (2017)	PH^2 169	–	Melanoma Non-Melanoma	–	96	83
Satheesha et al. (2017)	PH^2 200	–	Melanoma In-Situ Melanoma Atypical Nevus Common Nevus	–	96	97
Nasir et al. (2018)	PH^2 200	–	Melanoma Non-Melanoma	97.5	97.7	96.7
Pathan et al. (2018)	PH^2 200	137	Melanoma Non-Melanoma	94.7	95.7	100
Hu et al. (2019)	PH^2 200		Melanoma Non-Melanoma	91.90	92.5	91.30
Warsi, Khanam, Kamya and Suárez-Araujo (2019)	PH^2 200	–	Melanoma Non-Melanoma	97.5	98.1	93.84
Gulati et al. (2020)	PH^2 200	–	Melanoma Non-Melanoma	94.5	82.5	97.5
Proposed Method (SBS)	PH^2 200	20	Melanoma Non-Melanoma	**98.5**	**97.5**	**98.75**
	Ganster 270			**97.78**	**94.29**	**97**
Proposed Method (HSBS)	PH^2 200	20	Melanoma Benign Dysplastic	**96**	**94**	**97**
	Ganster 270			**93.33**	**90**	**95**

of the classification algorithm, yields a more robust and stable performance in the event of data changes.

Examining the type of the 20 selected features derived from both of the feature selection methods showed that the selected features by the SBFS method for both levels of the classification algorithm were more based on the texture and color features, whereas the shape, comlpex networks and color-based features were mostly selected by the mRMR method. However, the results of the lesions classification showed that in the case of data changes, the selected features of the mRMR method offered a more stable performance. With this fact in mind, it is concluded that the color features along with the shape-based features have less redundancy than the GLCM texture features, and this makes the classification algorithm more stable. Therefore, the performance of the classification algorithm can be improved by using other texture descriptors such as Wavelet Transform Analysis (WTA), Local Binary Patterns (LBP), Histogram of Oriented Gradients (HOG) and their combination with shape and color features.

In this research, it was proved that the performance of the proposed Stacking algorithm is more favorable than other methods. However, it should be noted that the process of training and tuning of the parameters of the classifiers in this method is hard and time consuming. Additionally, in this stduy, there was no specific method for choosing classifiers for the first and second stages. Therefore, to improve the performance of the proposed method, providing an effective method for selecting the type of base classifires can be considered as a future line of research.

Declaration of Competing Interest

All authors have participated in (a) conception and design, or analysis and interpretation of the data; (b) drafting the article or revising it critically for important intellectual content; and (c) approval of the final version. This manuscript has not been submitted to, nor is under review at, another journal or other publishing venue. The authors have no affiliation with any organization with a direct or indirect financial interest in the subject matter discussed in the manuscript

Credit authorship contribution statement

Ghasem Shakourian Ghalejoogh: Conceptualization, Methodology, Software, Formal analysis, Investigation, Resources, Writing - original draft. **Hussain Montazery Kordy:** Conceptualization, Methodology, Validation, Formal analysis, Investigation, Supervision, Writing - review & editing, Visualization. **Farideh Ebrahimi:** Conceptualization, Methodology, Validation, Formal analysis, Investigation, Writing - review & editing, Visualization.

References

Abbas, Q., Garcia, I. F., Emre Celebi, M., & Ahmad, W. (2013). A feature-preserving hair removal algorithm for dermoscopy images. *Skin Research and Technology, 19*(1), e27–e36.

Abuzaghleh, O., Barkana, B. D., & Faezipour, M. (2015). Noninvasive real-time automated skin lesion analysis system for melanoma early detection and prevention. *IEEE Journal of Translational Engineering in Health and Medicine, 3*, 1–12.

Aggarwal, C. C. (Ed.). (2014). Data classification: algorithms and applications. CRC press.

Agrawal, N., Sinha, P., Kumar, A., & Bagai, S. (2015). Fast & dynamic image restoration using Laplace equation based image inpainting. *Journal of Undergraduate Research and Innovation, 1*(2), 115–123.

Ahmed, M., Rasool, A. G., Afzal, H., & Siddiqi, I. (2017). Improving handwriting based gender classification using ensemble classifiers. *Expert Systems with Applications, 85*, 158–168.

Alfed, N., & Khelifi, F. (2017). Bagged textural and color features for melanoma skin cancer detection in dermoscopic and standard images. *Expert Systems with Applications, 90*, 101–110.

Backes, A. R., Casanova, D., & Bruno, O. M. (2009). A complex network-based approach for boundary shape analysis. *Pattern Recognition, 42*(1), 54–67.

Barata, C., Celebi, M. E., & Marques, J. S. (2017). Development of a clinically oriented system for melanoma diagnosis. *Pattern Recognition, 69*, 270–285.

Cancer Fact & Figures (2018). https://cancer.org/research/cancer-facts-statistics/all-cancer-facts-figures/cancer-facts-figures-2018.html [Online; accessed 30-November-2018].

Cascinelli, N., Ferrario, M., Tonelli, T., & Leo, E. (1987). A possible new tool for clinical diagnosis of melanoma: The computer. *Journal of the American Academy of Dermatology, 16*(21), 361–367.

Celebi, M. E., Kingravi, H. A., Uddin, B., Iyatomi, H., Aslandogan, Y. A., Stoecker, W. V., et al. (2007). A methodological approach to the classification of dermoscopy images. *Computerized Medical Imaging and Graphics, 31*(6), 362–373.

Chandrashekar, G., & Sahin, F. (2014). A survey on feature selection methods. *Computers & Electrical Engineering, 40*(1), 16–28.

Ding, C., & Peng, H. (2005). Minimum redundancy feature selection from microarray gene expression data. *Journal of Bioinformatics and Computational Biology, 3*(02), 185–205.

Gandhi, S. A., & Kampp, J. (2015). Skin cancer epidemiology, detection, and management. *Medical Clinics, 99*(6), 1323–1335.

Ganster, H., Pinz, P., Rohrer, R., Wildling, E., Binder, M., & Kittler, H. (2001). Automated melanoma recognition. *IEEE Transactions on Medical Imaging, 20*(3), 233–239.

Garnavi, R., Aldeen, M., & Bailey, J. (2012). Computer-aided diagnosis of melanoma using border-and wavelet-based texture analysis. *IEEE Transactions on Information Technology in Biomedicine, 16*(6), 1239–1252.

Gulati, S., & Bhogal, R. K. (2020). Classification of melanoma from dermoscopic images using machine learning. In *Smart intelligent computing and applications* (pp. 345–354). Singapore: Springer.

Haralick, R. M., & Shanmugam, K. (1973). Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics, 6*, 610–621.

Hu, K., Niu, X., Liu, S., Zhang, Y., Cao, C., Xiao, F., et al. (2019). Classification of melanoma based on feature similarity measurement for codebook learning in the bag-of-features model. *Biomedical Signal Processing and Control, 51*, 200–209.

Jaworek-Korjakowska, J., & Tadeusiewicz, R. (2013). Hair removal from dermoscopic color images. *Bio-Algorithms and Med-Systems, 9*(2), 53–58.

Kasmi, R., & Mokrani, K. (2016). Classification of malignant melanoma and benign skin lesions: Implementation of automatic ABCD rule. *IET Image Processing, 10*(6), 448–455.

Kaur, G., & Joshi, K. (2015). A brief survey about existed segmentation techniques in automatic detection and segmentation of skin melanoma images. *International Journal of Emerging Research in Management*, 2278–9359.

Khalid, S., Jamil, U., Saleem, K., Akram, M. U., Manzoor, W., Ahmed, W., et al. (2016). Segmentation of skin lesion using Cohen–Daubechies–Feauveau biorthogonal wavelet. *SpringerPlus, 5*(1), 1603.

Mendonca, T. F., Celebi, M. E., Mendonca, T., & Marques, J. S. (2015). PH2: A public database for the analysis of dermoscopic images. *Dermoscopy image analysis*. CRC Press.

Nasir, M., Attique Khan, M., Sharif, M., Lali, I. U., Saba, T., & Iqbal, T. (2018). An improved strategy for skin lesion detection and classification using uniform segmentation and feature selection based approach. *Microscopy Research and Technique, 81*(6), 528–543.

Novovičová, J., Somol, P., Haindl, M., & Pudil, P. (2007, November). Conditional mutual information based feature selection for classification task. In *Iberoamerican congress on pattern recognition* (pp. 417–426). Berlin, Heidelberg: Springer.

Oliveira, R. B., Marranghello, N., Pereira, A. S., & Tavares, J. M. R. (2016). A computational approach for detecting pigmented skin lesions in macroscopic images. *Expert Systems with Applications, 61*, 53–63.

Oliveira, R. B., Pereira, A. S., & Tavares, J. M. R. (2017). Skin lesion computational diagnosis of dermoscopic images: Ensemble models based on input feature manipulation. *Computer Methods and Programs in Biomedicine, 149*, 43–53.

Pathan, S., Prabhu, K. G., & Siddalingaswamy, P. C. (2018a). Hair detection and lesion segmentation in dermoscopic images using domain knowledge. *Medical & Biological Engineering & Computing, 56*(11), 2051–2065.

Pathan, S., Prabhu, K. G., & Siddalingaswamy, P. C. (2018b). A methodological approach to classify typical and atypical pigment network patterns for melanoma diagnosis. *Biomedical Signal Processing and Control, 44*, 25–37.

Peng, H., Long, F., & Ding, C. (2005). Feature selection based on mutual information: Criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on Pattern Analysis & Machine Intelligence, 8*, 1226–1238.

Queen, L. (2017). Skin cancer: Causes, prevention, and treatment. A Senior Thesis submitted in partial fulfillment of the requirements for graduation in the Honors Program Liberty University Spring 2017.

Rastgoo, M., Garcia, R., Morel, O., & Marzani, F. (2015). Automatic differentiation of melanoma from dysplastic nevi. *Computerized Medical Imaging and Graphics, 43*, 44–52.

Rokach, L. (2010). Ensemble-based classifiers. *Artificial Intelligence Review, 33*(1–2), 1–39.

Ruela, M., Barata, C., Marques, J. S., & Rozeira, J. (2017). A system for the detection of melanomas in dermoscopy images using shape and symmetry features. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, 5*(2), 127–137.

Satheesha, T. Y., Satyanarayana, D., Prasad, M. G., & Dhruve, K. D. (2017). Melanoma is skin deep: A 3D reconstruction technique for computerized dermoscopic skin lesion classification. *IEEE journal of translational engineering in health and medicine, 5*, 1–17.

Scabini, L. F., Fistarol, D. O., Cantero, S. V., Gonçalves, W. N., Machado, B. B., & Rodrigues Jr, J. F. (2017). Angular descriptors of complex networks: A novel approach for boundary shape analysis. *Expert Systems with Applications, 89*, 362–373.

Schaefer, G., Krawczyk, B., Celebi, M. E., & Iyatomi, H. (2014). An ensemble classification approach for melanoma diagnosis. *Memetic Computing, 6*(4), 233–240.

Schaefer, G., Rajab, M. I., Celebi, M. E., & Iyatomi, H. (2011). Colour and contrast enhancement for improved skin lesion segmentation. *Computerized Medical Imaging and Graphics, 35*(2), 99–104.

Warsi, F., Khanam, R., Kamya, S., & Suárez-Araujo, C. P. (2019). An efficient 3D color-texture feature and neural network technique for melanoma detection. *Informatics in Medicine Unlocked*, 100176.

Wolpert, D. H. (1992). Stacked generalization. *Neural networks, 5*(2), 241–259.

Xie, F., Fan, H., Li, Y., Jiang, Z., Meng, R., & Bovik, A. (2016). Melanoma classification on dermoscopy images using a neural network ensemble model. *IEEE Transactions on Medical Imaging, 36*(3), 849–858.