J Med Syst (2016) 40:96
DOI 10.1007/s10916-016-0460-2

SYSTEMS-LEVEL QUALITY IMPROVEMENT

Novel Approaches for Diagnosing Melanoma Skin Lesions Through Supervised and Deep Learning Algorithms

J. Premaladha[1] · **K. S. Ravichandran**[1]

Received: 17 November 2015 / Accepted: 1 February 2016

Abstract Dermoscopy is a technique used to capture the images of skin, and these images are useful to analyze the different types of skin diseases. Malignant melanoma is a kind of skin cancer whose severity even leads to death. Earlier detection of melanoma prevents death and the clinicians can treat the patients to increase the chances of survival. Only few machine learning algorithms are developed to detect the melanoma using its features. This paper proposes a Computer Aided Diagnosis (CAD) system which equips efficient algorithms to classify and predict the melanoma. Enhancement of the images are done using *Contrast Limited Adaptive Histogram Equalization technique (CLAHE)* and *median filter*. A new segmentation algorithm called *Normalized Otsu's Segmentation (NOS)* is implemented to segment the affected skin lesion from the normal skin, which overcomes the problem of variable illumination. Fifteen features are derived and extracted from the segmented images are fed into the proposed classification techniques like *Deep Learning based Neural Networks* and *Hybrid Adaboost-Support Vector Machine (SVM)* algorithms. The proposed system is tested and validated with nearly 992 images (malignant & benign lesions) and it provides a high classification accuracy of 93 %. The proposed CAD system can assist the dermatologists to confirm the decision of the diagnosis and to avoid excisional biopsies.

Keywords Preprocessing · Segmentation · Classification · Artificial neural networks · Support vector machine · Deep learning · Adaboost

This article is part of the Topical Collection on *Systems-Level Quality Improvement*

✉ K. S. Ravichandran
raviks@it.sastra.edu

J. Premaladha
premaladha@sastra.edu

[1] School of Computing, SASTRA University, Tirumalaisamudram, Thanjavur 613401, Tamilnadu, India

Introduction

According to the statistics of American Cancer Society, "In the year of 2015, 73,870 new melanomas will be diagnosed and 9940 people are expected to die of melanoma" (http://www.cancer.org/cancer/skincancer-melanoma/detailedguide/melanoma-skin-cancer-key-statistics). Commonly, malignant melanoma is diagnosed using ABCD (Asymmetry, Border Irregularity, Color variation and Diameter) analysis. Melanoma patients have chances of survival, only if it is diagnosed at early stages. A computer aided diagnosis system is developed using image processing and machine learning algorithms to assist the dermatologists to diagnose the cancer at early stages. Many techniques were proposed in the literature to diagnose the melanoma at early stages. Significant techniques are as follows: Green et al., proposed a classification technique with features like size [1], shape, color and boundary of the skin lesion. Lee et al., developed a skin cancer diagnosis system with morphological features and classified using Artificial Neural Networks [2]. Aitken et al., classified using features like area, perimeter; irregularity features of border, statistical features of wavelengths and gradients of red, blue and green bands [3]. Chang et al., proposed a heuristic approach for feature extraction and lesion discrimination [4]. She et al., classified the melanoma skin lesions with the features like Asymmetry, Border, Color variation (Red, Blue, Green), and Diameter [5].

Published online: 12 February 2016

Fassihi et al., framed a classification procedure of melanoma with help the mean and variance of wavelet coefficients of skin lesion images [6]. In the thesis work, Computer-aided melanoma diagnosis by Garnavi, R., texture based features, border based features and geometrical features are used for classification [7, 8]. Amaliah et al., used common ABCD features and classified by the calculated TDV (Total Dermoscopy Value) [9]. Safi et al., used Asymmetry, Border irregularity, Color variation, Diameter and Elevation (ABCDE) [10]. Masood et al., presented a review of various techniques in each and every phase of the computer aided diagnosis of melanoma [11]. Focusing on feature extraction, the authors composed all possible features including asymmetry-based, color based [12], border irregularity based features, statistical features and also narrated the methods to extract those features. Premaladha J and Ravichandran KS proposed a methodology to detect the melanoma skin cancer using phylogenetic trees [13]. They also presented two methodologies, i) Scan-line method and ii) Fuzzy relational method to quantify the asymmetricity of the melanoma lesions by considering the borders of the melanoma image as fuzzy borders [14]. A rule based computer aided diagnosis is proposed by Shao et al. , which works on PC environment [15]. This paper presents novel methodologies to preprocess the images, to segment the affected lesion from the healthy skin, to extract the statistical and shape features and for classification based on the extracted features.

The remaining paper is organized as follows: Section-II presents the materials and methods. Section-III comprises the procedure for the automatic detection of melanoma and its phases. Section- IV gives the results and discussion of the proposed methodologies. Section – V drives home the conclusion and the future work.

Materials and methods

The main objective of the paper is to classify melanoma skin lesions using the features derived from the dermoscopy images The proposed methodology consists of the following steps: i) Image acquisition, ii) Preprocessing, iii) Segmentation, iv) Border extraction, v) Feature extraction and vi) Classification. The flow diagram of the implementation of the proposed methodology is given in the Fig. 1.

Dataset We have collected **992** dermoscopy images (http://www.bccancer.bc.ca/HPI/SkinCancerAtlas/Melanoma/default.htm, http://www.cancer.org/cancer/skincancer-melanoma/detailedguide/melanoma-skin-cancer-key-statistics, http://www.dermnet.org.nz/lesions/img/melanoma/mel-is.html, http://www.meddean.luc.edu/lumen/MedEd/medicine/dermatology/melton/content1.htm) from various repositories and validated through Chi-square distribution with 5 % level of significance and tested n-cross ratio, which shows that the data taken from the various repositories are accepted at 5 % level of significance. The dataset is comprised of both melanoma and other benign skin lesion images which are taken from different patients. The images are acquired through Dermatoscope [16] and also using digital camera Nikon D3 or Nikon D1x body and a Nikkor 2.8/105 mm micro lens [17].

Automatic detection of melanoma skin cancer

In this section, novel methodologies proposed in each phase of the computer aided diagnosis system shown in the Fig. 1 are discussed with the implementation details.

Image acquisition It is the first phase of the proposed system. The images of the skin are captured using the technique called Dermoscopy or Epiluminescence microscopy (http://en.wikipedia.org/wiki/Dermatoscopy). It is a technique used to examine the human skin and capture images with the help of equipment called Dermatoscope. Such images will be used to predict the melanoma occurrence.

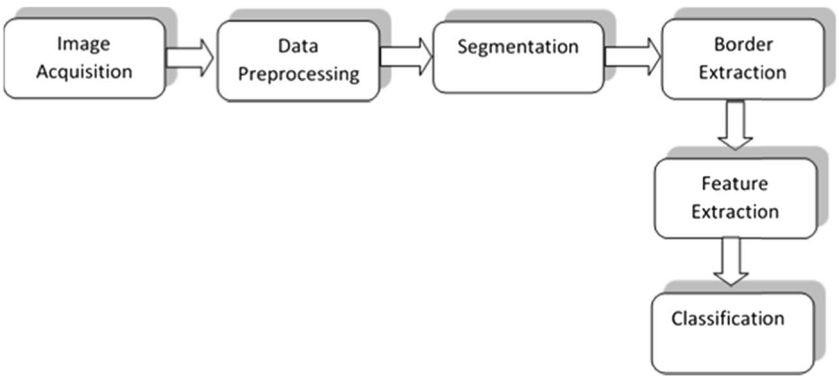

Fig. 1 Steps in computer aided diagnosis system

Preprocessing After the image acquisition, it will be processed before going to the third phase, segmentation. The dermoscopy images are enhanced by filtering the noise which occurs due to intrusion of small hairs, scars in the human skin. Some preprocessing techniques proposed in the literature [18] are given in Table 1.

Initially, the image is acquired as given in Fig. 2a. The CLAHE (http://radonc.ucsf.edu/research_group/jpouliot/tutorial/HU/Lesson7.htm) preprocessing technique proposed by [26], is an improvised method of Adaptive Histogram Equalization (AHE), specially designed to preprocess the medical images. It is applied on the image to get a contrast enhanced image, from which accurate features can be derived. CLAHE preprocessing is commonly applied for mammogram images and now it's used for the first time in this field to enhance the dermoscopy images. CLAHE technique processes each tile of the image and improves the contrast of each tile of the image. Hence, the output will be more precise than enhancing the contrast of an entire image. The enhanced image using CLAHE technique is shown in the Fig. 2b. In this paper median filtering technique [19, 23] is used for noise removal, as it is the most appropriate method for medical images. The noise is removed by applying median filter of size 11×11. Even the ragged edges are smoothened after filtering as shown in Fig. 2c.

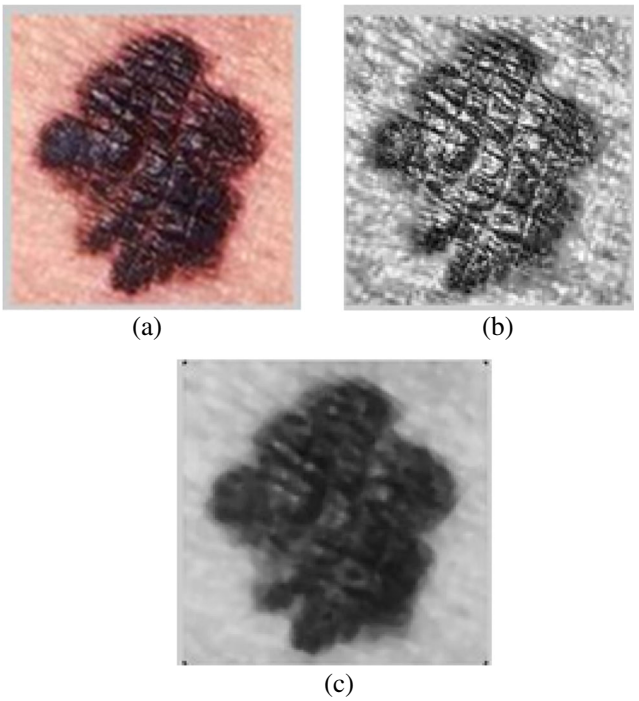

Fig. 2 **a** Original image. **b** Contrast Enhanced image. **c** Filtered image using CLAHE

Segmentation Segmentation is the process of dividing an image into meaningful specks, such that the required region is segmented, from which we can infer the necessary information for further processing. In case of dermoscopy images, the affected lesion area should be extracted from the normal skin and from the segmented region significant features extracted to distinguish the malignant lesion from benign one. Various segmentation techniques exist in the literature and some of them are given in the Table 2: Here, segmentation is done using the proposed *NOS* technique. This method is specially used to separate the foreground from the background. Input for the segmentation is the filtered image and the output is shown in the Figs. 3 and 4.

The procedure for the proposed NOS technique is as follows. The filtered image is normalized using Local–global Block Analysis [31] to remove variable illumination. Normalized image is shown in Fig. 3. It is then segmented using Otsu's Segmentation [32]. In Otsu's Segmentation, threshold is obtained. The pixels with intensity value greater than the fixed threshold are replaced by 1 and the rest are replaced by 0. The segmented images are sent for border extraction phase.

Table 1 Preprocessing techniques

Techniques	References
Black frame removal	Celebi, et al., [19]
Automatic colour equalization	Schaefer, et al., [20]
Hair removal technique	Capdehourat, et al., [21] Schmid-Saugeon, et al., [22]
Dull Razor, Karhunen-Loe've transform	Messadi, et al., [23]
Gaussian filter Pseudo random filter	Hance, et al., [24]
Non – skin masking	
Colour space transformation Contrast enhancement	Celebi, et al., [25]

Table 2 Segmentation techniques

Techniques	Reference
Iterative segmentation	Rajab, et al.,[27] Palus, et al., [28]
Cooperative neural network segmentation	Schaefer, et al., [20]
Adaptive thresholding Gradient vector flow	Silveira, et al., [29]
Fuzzy based split and merge	
Fuzzy C – means clustering Principle component transform	Hance, et al., [24]
Region fused band and narrow band graph partition	Yuan, et al., [30]

Fig. 3 Normalized image

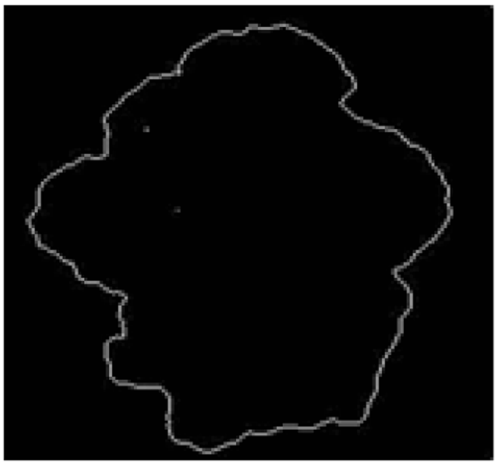

Fig. 5 Border of the affected lesion

Border extraction The segmented lesion is the input for the border extraction phase. To determine the exact border from the segmented lesion image, different methodologies are followed and also proposed by the researchers. Erkol et al., used Gradient Vector flow (GVF) snakes [33], Celebi et al., proposed the statistical region merging technique [34]. Celebi et al., developed a fusion of several thresholding methods to detect the border automatically [35]. Here, the border of the dermoscopy is found by connecting the non-zero pixels to the neighborhood non-zero pixels. Then, the border or outline of the image is drawn as shown in the Fig. 5.

Feature extraction The next important phase is feature extraction. It is the process of deriving meaningful features from the image for classification. Researchers [36] used different techniques to extract the features. They are: Principle Component Analysis (PCA), Wavelet Packet Transform (WPT) [37, 38] Grey Level Co-occurrence Matrix (GLCM) [39], Fourier Power Spectrum (FPS) [40], Gaussian Derivative Kernels [41] and Decision Boundary Feature Extraction [42] to reduce the data dimensionality. In this work, few features are extracted with GLCM matrix and few features are from geometrical based features. Table 3 presents the list of statistical and shape features [43].

Fig. 4 Segmented image

These features are derived from skin lesion images for classification. Mean value of an image: $f(x,y) = \frac{1}{mn}\sum_{(r,c)\in W} g(r,c)$, where f(x, y) = Mean value, m × n = Size of window, g(r, c) = Input image, (r,c) = rows and columns is the average of sum of all the pixel values of the image matrix. Standard deviation $\tilde{f}(x,y) = \sqrt{\frac{1}{mn-1}\sum_{(r,c)\in W}\left(g(r,c)\frac{1}{mn-1}\sum_{(r,c)\in W}g(r,c)\right)^2}$, where m × n = Size of window, $g(f, c)$ = Input image, (r,c) = rows and columns represents the "dispersion" occurs from the average pixel value. Variance $\tilde{f}(x,y) = \frac{1}{mn-1}\sum_{(r,c)\in W}\left(g(r,c)\frac{1}{mn-1}\sum_{(r,c)\in W}g(r,c)\right)^2$ says how far the pixel points are spread out from the mean

Table 3 Features derived from skin lesions

Statistical features	Shape features
1. Mean	9. Area
2. Standard deviation	10. Perimeter
3. Variance	11. Diameter
4. Entropy	12. Asymmetry index
5. Contrast/Inertia	13. Circularity index
6. Homogeneity	14. Fractal dimension
7. Energy	15. Compactness index
8. Correlation	

value. Entropy value of an image $-\sum_i P(x_i)\log_2 P(x_i)$ gives the randomness measure to characterize the texture of the image, where p contains the histogram counts. Intensity contrast between a pixel and its neighborhood pixels over the entire image is defined as the contrast $=\sum_{i,j}|i-j|^2 p(i,j)$. Homogeneity value $\sum_{i,j}\frac{p(i,j)}{1+|i-j|}$ measures the closeness of the distribution of elements in the GLCM to the GLCM diagonal. Energy $\sum_{i,j}p(i,j)^2$ gives the sum of squared values of the image matrix. The measure of correlation between pixels is given as $Correlation=\sum_{i,j}\frac{(i-\mu i)(j-\mu j)p(i,j)}{\sigma_i\sigma_j}$. Area is the measure of all pixels in an image. Perimeter of an image is drawn by connecting the non-zero pixel with at least one non-zero valued pixel. Diameter is the equivalent circular diameter $\sqrt{4A\pi}$ calculated for the image with its area A.

Asymmetry Index is calculated from the relation, $AI=\frac{\Delta A}{A}\times 100$, where ΔA = Difference between the total area and the lesion area. The circularity index is calculated from the relation, $=\frac{4A\pi}{P^2}$, where P is the Perimeter. Fractal dimension of the lesion is quantified by Hausdorff dimension using box counting method and it is defined as follows: $N(r)\sim r^{-D}$ where $-D=\lim_{r\to 0}\frac{\log N(r)}{\log r}$, r = size of Square boxes, N(r) = Number of non-empty boxes of size r. Compactness index is defined as $CI=\frac{4\pi P^2}{A}$. The extracted features are normalized based on the mean and standard deviation $Z=\frac{X-\mu}{\sigma}$, where X = Value to be Normalized μ = Mean of the samples, σ = Standard deviation of the samples, and Z = Normalized value; and then given as input for classifiers.

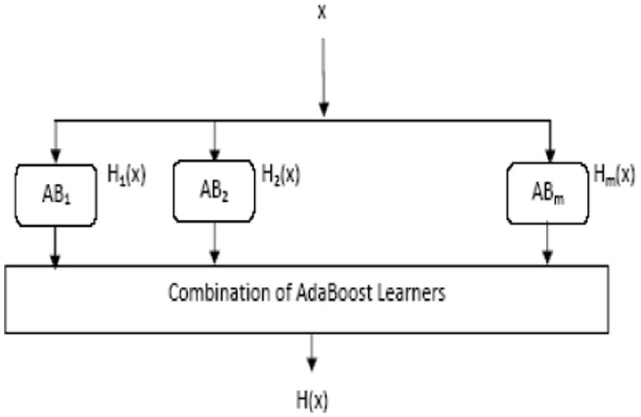

Fig. 7 A typical ensemble of classifiers

Classification techniques

The final phase is the classification [44] and in this section we proposed *classification* techniques, like deep learning based neural network, support vector machine [45], hybrid AdaBoost algorithms; and then the results are compared with the state of the art techniques discussed in recent literature. Only few researchers are working on the parameter based classification techniques. They are: In 2004, Sikorski [37] used Wavelet Analysis with the help of 8 parameters which are fragmentation index, circularity

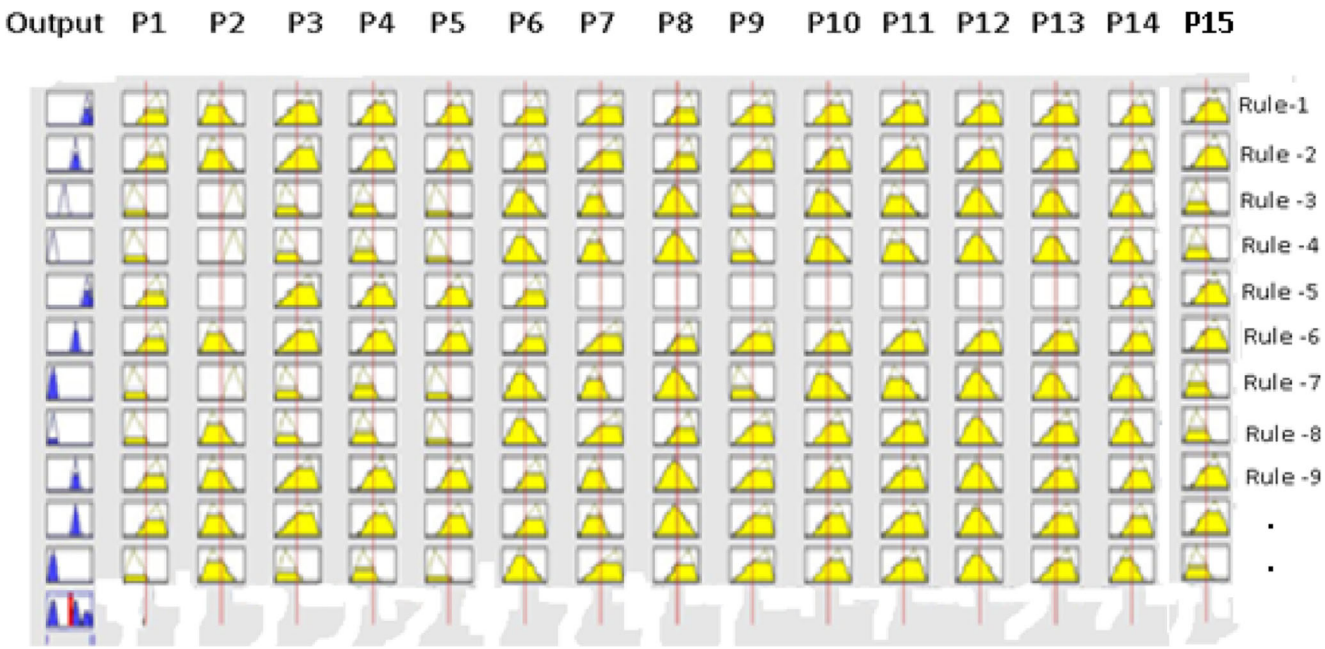

Fig. 6 Rule view of the fuzzy inference system

factor, asymmetry index, mean color variations, standard deviation of color variations, skewness of color variations, texture and boarder irregularity index. The accuracy is claimed as 83 %. Garnavi et al., identified diseases through decision tree with the help of 18 features [8] namely, energy, standard deviation, skewness, kurtosis, norm, entropy, average-energy, area, perimeter, greatest and lowest diameter, irregularity index, irregularity indices A to D, minor and major asymmetry indices; and asymmetry index. They claimed accuracy as 88.3. Lau et al., have developed wavelet based back propagation neural network with seven parameters which are mean, maximum, minimum asymmetry, mean, standard deviation, median and variance [46]. They have achieved nearly 90 % of accuracy. The same level of accuracy has been also achieved by Fassihi et al., and Garnavi et al., by using logistic model tree classification techniques and threshold based artificial neural network respectively [6, 8].

Deep learning based neural networks (DLNN)

The architecture of deep learning consists of a multilayered network in which the optimization routine is applied on it. Each hidden layer is connected to many other preceding hidden layers and it is going further the network in a non-linear combination of its preceding layers. When the optimization

Table 4 Pseudocode for Hybrid- AdaBoost algorithm

Pseudo Code for Hybrid- AdaBoost Algorithm

Input : Z = { z1, z2, …, zN} with zk= (xk, yk) for k = 1,2, .., N
M = Number of combined classifier

Output : H(x) - it is suited for training data set

1. Initialize data weight: $wi \leftarrow \frac{1}{N}$ for i ε {1, 2, .., N} where wi is the weight
2. Initialize standard deviation (σ), minimum standard deviation (σmin), and step-size (σStep)
3. do
   begin
   3a. Using Radial Basis Function kernel in SVM to train the given weighted sample Data
   3b. Calculate the training error of Hm say Ω m, where

$$\Omega_m = \sum_{i=1}^{N} w_i^T y_i \neq Hm(x)$$

   3c. if $\Omega_m > 0.5$ then
      begin
         σ ← σ – σ Step;
         Taking the next set of training sample data and
         Then goto step-3
      end;

   3d. Set the weight of the classifier H_m: $\delta_m = \frac{1}{2} \log\left(\frac{1-\Omega_i}{\Omega_i}\right)$
   3e. Update the weight vector

$$w_i^{m+1} = \frac{w_i^m \exp\{-\delta_m y_i H_m(x_i)\}}{B_i}$$; where Bi is the normalization constant and normalized to $\sum_{i=1}^{N} w_i^{m+1} = 1$

   end;
   while ($\sigma > \sigma_{min}$);
4. Compute the output: $H(x) = \text{sgn}(\sum_{m=1}^{M} \Omega_m . H_m(x))$

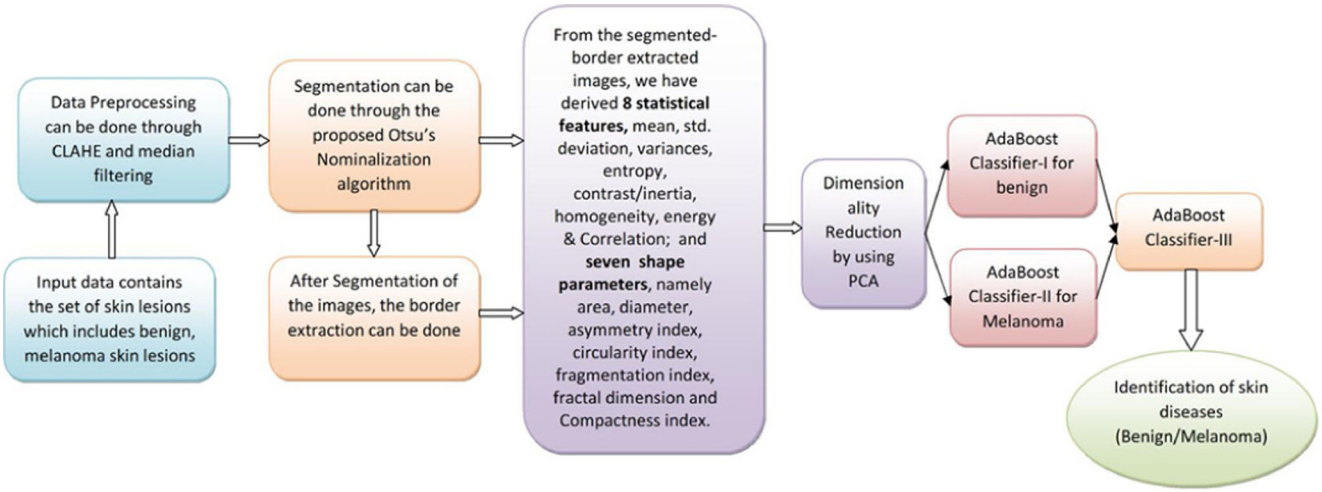

Fig. 8 Proposed architecture for ensemble based Hybrid-AdaBoost algorithm

routine is applied to the network, each hidden layer becomes an optimally weighted, non-linear combination of the layer below it. Again, the number of hidden neurons of each hidden layers are arranged in monotonically decreasing, that is, the number of neurons in each hidden layers are $m_1 \geq m_2 \geq m_3 \ldots\ldots \geq m_n$, where m_i is the number of neurons in i^{th} hidden layer. These arrangements of neurons in each hidden layer are called a lower dimensional projection of the layer below it, because of all the combining and recombining of the outputs from all the previous layers in combination with their activation functions. Hence, the information from the below layer is categorically summarized by a non-linear, optimally weighted and lower dimensional projection in each subsequent layer of the deep learning network. In the state of art, many papers implemented Artificial Neural Networks (ANN), but in this paper, we have proposed DLNN for classification of the diseases. It is a branch of Machine learning algorithms and it will overcome the limitations of the existing ANN and it provides good classification accuracy. The merits and demerits of ANN and DLNN are summarized below:

a) Back propagation learning in ANN contains single hidden layer and it provides good accuracy; however it requires numerous data set for training with better accuracy. Whereas, DLNN contains many hidden layers and it provide better accuracy by using non linearity, optimally

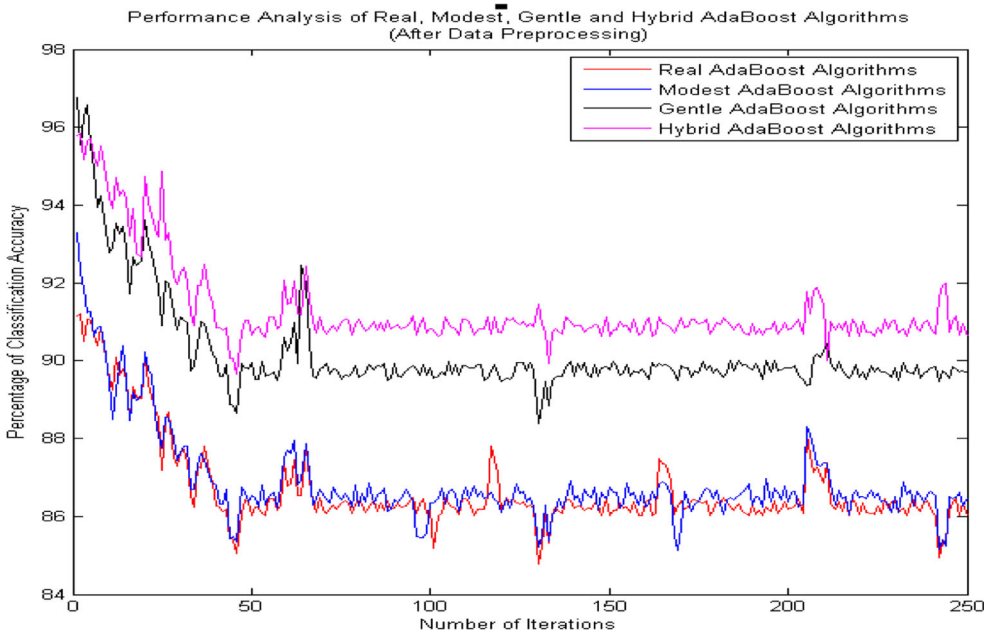

Fig. 9 The performance analysis of AdaBoost (Real, Gentle, Modest) and Hybrid-AdaBoost algorithms

Table 5 Accuracy acquired in the state of the art literature

S.No.	Name of the authors	Preprocessing	Segmentation	No. of features	Classifier	Classification accuracy literature (%) (a)	Accuracy for the taken dataset (%) (b)		Accuracy for the proposed preprocessing, segmentation + State of the art classifier (%) (c)	
							Sensitivity (%)	Specificity (%)	Sensitivity (%)	Specificity (%)
1.	Sikorski, 2004 [37]	RGB to Indexed Image	Thresholding	8	ANN	87.45	87.23		90.12	
							89.69	83.91	92.65	86.86
2.	Maglogiannis, et al., 2006 [39]	Mean Filter	Local Thresholding	8	SVM	87.98	88.04		90.44	
							90.26	85.01	92.65	86.86
3.	Lau, et al., 2009 [46]	Wavelet De-noising + Median Filter	Thresholding + SRM	13	Back Propagation Neural Networks	88.54	88.09		89.95	
							90.42	91.00	92.65	86.86
4.	Fassihi, et al., 2011 [6]	Mean Filter	Morphological operators	8	Artificial Neural Networks	89.56	89.92		90.13	
							92.32	86.80	92.65	86.86
5.	Garnavi, et al., 2012 [8]	Mean filter	Global Thresholding + Adaptive Histogram Thresholding	23	Logistic Model Tree	89.67	89.41		90.32	
							91.80	83.31	92.65	86.86

Table 6 Performance analysis of the proposed methodologies

S.no	Name of the authors	Preprocessing	Segmentation	Features	Classifer	Accuracy (%)	Sensitivity (%)	Specificity (%)	Kappa value
1.	Proposed Methods	CLAHE and Median Filter	Otsu's Normalized algorithm (Proposed)	15	ANN	90.12	92.65	86.86	0.798
2.					ANFIS	90.39	92.68	87.3	0.802
3.					SVM	90.44	92.65	86.86	0.805
4.					DLNN	92.89	94.83	90.46	0.856
5.					Real AdaBoost	86.52	89.16	56.76	0.726
6.					Modest AdaBoost	86.84	89.5	83.49	0.732
7.					Gentle AdaBoost	89.91	92.32	86.8	0.794
8.					Hybrid AdaBoost	91.73	94.08	88.71	0.831

weighted and lower dimensional production mechanisms as discussed earlier.

b) Even with the less amount of dataset, DLNN provides good classification accuracy.

c) In case of DLNN, the network is not trained with the target output, instead it is trained as equal as the given input dataset. This will indirectly lead to better accuracy, as it depends on the input rather than the target output.

The proposed DLNN is constructed with 15 input features of 992 samples of both malignant and benign lesions. The network consists of 15 input neurons, 3 hidden layers and 2 output neurons. The network has been trained and the classification accuracy reached through DLNN is 92.89 %. Whereas, in the same dataset, back propagation learning of ANN is implemented by taking 70 % of dataset for training, 15 % of data for testing and 15 % for validation. Through ANN, 90.1 % of classification accuracy is acquired.

Adaptive neuro-fuzzy inference system (ANFIS)

In the same dataset, a Mamdani based Adaptive Neuro-Fuzzy Inference System (ANFIS) has been constructed with 15 inputs and two outputs. Fifteen features of each image is given as input and the output is predicted accordingly. For both the input and the output variables, triangular membership functions are used to fix the range with three linguistic levels namely Low, Medium and High and totally we derived 8^{15}rules and some sample rules are given below and its rule view are given in Fig. 6

1. If (mean is low) and (std_dev is medium) and (variance is medium) and (correlation is low) and (entropy is low) and (contrast is medium) and (homogeneity is low) and (energy is medium) and (area is low) and (perimeter is medium) and (diameter is low) and (asymmetry_index is low) and (circularity_index is medium) and (fractal_dim is low) and (compactness_index is medium) then (prediction is benign_initial)
2. If (mean is low) and (std_dev is high) and (variance is high) and (correlation is medium) and (entropy is medium) and (contrast is low) and (homogeneity is high) and (energy is high) and (area is low) and (perimeter is high) and (diameter is high) and (asymmetry_index is medium) and (circularity_index is low) and (fractal_dim is medium) and (compactness_index is medium) then (prediction is benign_advanced)
3. If (mean is high) and (std_dev is high) and (variance is high) and (correlation is medium) and (entropy is high) and (contrast is low) and (homogeneity is medium) and (energy is high) and (area is high) and (perimeter is high) and (diameter is high) and (asymmetry_index is high) and

(circularity_index is high) and (frac_dim is high) and (compactness_index is high) then (prediction is melanoma_advanced)

The proposed ANFIS is tested and implemented with the same dataset and we got 90.3 % of classification accuracy.

AdaBoost algorithm

AdaBoost based classification technique for prediction of melanoma skin lesions is proposed in this section. It is used to designing an ensemble [47, 48] of classifiers as shown in Fig. 7 to predict whether the skin lesion image is malignant or benign. It classifies the features based on the weights of the parameters rather than sampling (with or without replacement). The output of the M weak classifiers is combined and from that, the output of the ensemble classifier H(x) - strong classifier is obtained. The value of ensemble classifier H(x) is defined as $H(x) = \mathrm{sgn}\left(\sum_{n=1}^{m} \alpha_n H_n(x)\right)$ where α_n = weight vector of each weak classifier. The ground truth of the algorithm is, all the weak classifiers are trained sequentially and the weight vector is updated (increment/decrement) based on the difference between the training set pattern accuracy of the n^{th} classifier and $(n\text{-}1)^{th}$ classifier. The difference is the error rate occurs between the classifiers and the weight is incremented or decremented by (Δx) based on the error percentage.

The variants of AdaBoost algorithms which adopts supervised learning are Real AdaBoost (1999), Margin AdaBoost (2000), Modest AdaBoost (2000), Gentle AdaBoost (2000), Any Boost (2000), Margin Boost (2000), KL Boost (2003), Float Boost (2004), Active Boost (2004), Jensen-Shannon Boost(2005), Emphasis Boost(2006), Entropy Boost(2007), Reweight Boost (2007) and so on. Each AdaBoost algorithm has its own merits and demerits and for the verification purpose we have tested and implemented the same data with three AdaBoost algorithms namely Real, Gentle and Modest AdaBoost algorithms. Their classification accuracies are given in the Table 6.

Hybrid AdaBoost algorithm (SVM-AdaBoost)

The Support Vector Machine (SVM) is one of the supervised learning algorithm, which is used to solve the classification problems in different research domains. It is a model-free and data-driven model which suits best for datasets with higher dimensions as similar to our proposed problem. In the hybrid AdaBoost algorithm, SVM is considered as a base classifier and it is trained by the Gentle AdaBoost classifier. In the training part, the AdaBoost classifier updates the kernel function σ periodically and it will boost the weight vector of the training dataset. The pseudocode representation for the hybrid AdaBoost algorithm is given in Table 4.

The proposed architecture for ensemble based hybrid AdaBoost algorithm is given in Fig. 8. The proposed Hybrid-AdaBoost algorithm is tested and implemented with the same dataset and the classification accuracy is raised from 89.1 % (Gentle AdaBoost) to 91.7 % (Hybrid AdaBoost). The performance of three AdaBoost algorithms and hybrid AdaBoost algorithm is given Fig. 9.

Results and discussion

The proposed CAD system uses the combination of CLAHE and median filtering, Normalized Otsu's Segmentation, and machine learning algorithms. All the proposed machine learning algorithms are tested with 15 input features and two output features for 992 samples. In literature, the researchers used different data sets for classification purpose. Here, for the sake of uniformity we tested the state of the art algorithms in our dataset (992 samples) and the classification accuracy is obtained for comparison. In Table 5, column (a) shows the accuracy obtained with their dataset; column (b) is the accuracy obtained by implementing the state of the art algorithms in our dataset and column (c) presents the accuracy acquired by implementing the proposed preprocessing and segmentation techniques with the state of the art classification technique.

The detailed performance analysis of the literature is shown in Table 5. Based on the intensity of the skin lesions, the proposed GUI will provide any one of the following results: (1). You are out of danger, just a benign lesion (2). Your lesion is transforming from benign to melanoma, consult a doctor, and (3). Please consult the doctor immediately, it is melanoma. Table 6 shows performance analysis of the proposed machine learning algorithms using its classification accuracy, sensitivity, specificity and kappa value.

From the Table 6, it is concluded that

- CLAHE + median filter & Normalized Otsu's segmentation algorithm helped to acquire the affected region exactly.
- From the extracted region, significant features are extracted and fed for the classification.
- When we are combining SVM + AdaBoost (Hybrid-AdaBoost), the resultant architecture provides better classification accuracy than the other supervised machine learning algorithms.
- DLNN requires minimal processing time (234 s) and provides higher classification accuracy when compared to other machine learning algorithms.

During the proposed study, we observed that some

data-overfitting problem occurs and it impacts the classification accuracy. It is because of considering larger number of decision parameters and datasets. There is a possibility of improving the classification accuracy by adjusting the training, testing and validation datasets. Another proven way of improving the percentage of classification accuracy, is representing the datasets as tensor instead of vector and it is our future work.

Conclusion

Thus, the paper concludes that the proposed combination of techniques composes a computer aided diagnosis system that provides better classification accuracy with less computational time. As it includes both shape and statistical features, the finer details of the images are acquired. The different datasets are taken from various standard repositories like Mednode, PH^2 database [16] and its validities of the datasets are checked with 10-fold cross validation method with 95 % level of confidential limits. From the collected images, we have derived 15 input features and it is used to evolve the classification accuracy with the help of different machine learning algorithms like ANN, ANFIS, SVM, DLNN and Hybrid-AdaBoost algorithms. Finally, it is concluded that DLNN and Hybrid-AdaBoost algorithms are performed well when compared to the other machine learning algorithms. This work can be extended to any domain of images, like Satellite images, Industrial images, CT images, MRI images, etc. and these predictions which are estimated only on the features of the images.

Acknowledgments We, the authors sincerely thank the Department of Science and Technology, India for providing the INSPIRE fellowship (IF120649) to carry out this research work. Our earnest thanks to the SASTRA University for providing all the facilities to proceed with the research.

References

1. Green, A., Martin, N., Pfitzner, J., O'Rourke, M., and Knight, N., Computer image analysis in the diagnosis of melanoma. *J. Am. Acad. Dermatol.* 31(6):958–964, 1994.
2. Lee, H. C., *Skin cancer diagnosis using hierarchical neural networks and fuzzy logic*. Department of Computer Science, University of Missouri, Rolla, 1994.
3. Aitken, J. F., Pfitzner, J., Battistutta, S. O., Rourke, P. K., Green, A. C., and Martin, N. G., Reliability of computer image analysis of pigmented skin lesions of Australian adolescents. *J. Cancer* 78(2): 252–257, 1996.
4. Chang, Y., Stanley, R. J., Moss, R. H., and Van Stoecker, W., A systematic heuristic approach for feature selection for melanoma discrimination using clinical images. *Skin Res. Technol.* 11(3): 165–78, 2005.
5. She, Z., Liu, Y., and Damatoa, A., Combination of features from skin pattern and ABCD analysis for lesion classification. *Skin Res. Technol.* 13(1)25–33, 2007, which has been published in final form at http://onlinelibrary.wiley.com.
6. Fassihi, N., Shanbehzadeh. J., Sarafzadeh, A., and Ghasemi, E., Melanoma diagnosis by the use of wavelet analysis based on morphological operators. *Proceedings of the International Multiconference of Engineers and Computer Scientists*. 16–18, 2011.
7. Garnavi, R., Computer-aided diagnosis of melanoma. PhD thesis. 2011.
8. Garnavi, R., Aldeen, M., and Bailey, J., Computer-aided diagnosis of melanoma using border-and wavelet-based texture analysis. *IEEE Trans. Inf. Technol. Biomed.* 16(6):1239–1252, 2012.
9. Amaliah, B., Fatichah, C., and Widyanto, M. R., ABCD feature extraction of image dermatoscopic based on morphology analysis for melanoma skin cancer diagnosis. *J. Comput. Inf.* 3(2):82–90, 2012.
10. Safi, A., Baust, M., Pauly, O., Castaneda, V., Lasser, T., Mateus, D., Navab, N., Hein, R., and Ziai, M., Computer–aided diagnosis of pigmented skin dermoscopic images. *MCBR-CDS 2011, LNCS 7075*, 105–115. Springer-Verlag Berlin Heidelberg. 2012.
11. Masood, A., and Al-Jumaily, A. A., Computer aided diagnostic support system for skin cancer: a review of techniques and algorithms. *Int. J. Biomed. Imaging*, 2013. doi:10.1155/2013/323268.
12. LeAnder, R., Chindam, P., Das, M., and Umbaugh, S. E., Differentiation of melanoma from benign mimics using the relative-color method. *Skin Res. Technol.* 16(3):297–304, 2010.
13. Premaladha, J., and Ravichandran, K. S., Detection of melanoma skin lesions using phylogeny. *Natl. Acad. Sci. Lett.* 38(4):333–338, 2015.
14. Premaladha, J., and Ravichandran, K. S., Quantification of fuzzy borders and fuzzy asymmetry of malignant melanomas. *Proc. Natl. Acad. Sci. India Sect. A. Phys. Sci.* 85(2):303–314, 2015.
15. Shao, S., and Grams, R. R., A proposed computer diagnostic system for malignant melanoma (CDSMM). *J. Med. Syst.* 18(2):85–96, 1994.
16. Mendonça, T., Ferreira, P. M., Marques, J., Marcal, A. R. S., and Rozeira, J., PH^2 - A dermoscopic image database for research and benchmarking. *35th International Conference of the IEEE Engineering in Medicine and Biology Society*: Osaka, Japan. 2013.
17. Giotis, I., Molders, N., Land, S., Biehl, M., Jonkman, M. F., and Petkov, N., MED-NODE: a computer-assisted melanoma diagnosis system using non-dermoscopic images. *Expert Syst. Appl.* 42: 6578–6585, 2015.
18. Premaladha, J., Sujitha, S., Lakshmi Priya, M., and Ravichandran, K. S., A survey on melanoma diagnosis using image processing and soft computing techniques. *Res. J. Inf. Technol.* 6(2):65–80, 2014.
19. Celebi, M. E., Iyatomi, H., Stoecker, W. V., Moss, R. H., Rabinovitz, H. S., Argenziano, G., and Peter, H., Automatic detection of blue-white veil and related structures in dermoscopy images. *Comput. Med. Imaging Graph.* 32:670–677, 2008.
20. Schaefer, G., Rajab, M. I., Celebi, M. E., and Iyatomi, H., Colour and contrast enhancement for improved skin lesion segmentation. *Comput. Med. Imaging Graph.* 35:99–104, 2010.
21. Capdehourat, G., Corez, A., Bazzano, A., Alonso, R., and Musé, P., Toward a combined tool to assist dermatologists in melanoma detection from dermoscopic images of pigmented skin lesions. *Pattern Recogn. Lett.* 32:2187–2196, 2004.
22. Schmid-Saugeon, P., Guillod, J., and Thiran, J. P., Towards a computer-aided diagnosis system for pigmented skin lesions. *Comput. Med. Imaging Graph*. 65–78, 2003.
23. Messadi, M., Bessaid, A., and Taleb-Ahmed, A., Extraction of specific parameters for skin tumour classification. *J. Med. Eng. Technol.* 33(4):288–295, 2009.

24. Hance, G. A., Umbaugh, S. E., Moss, R. H., and Stoecker, W. H., Unsupervised color image segmentation. *IEEE Eng. Med. Biol.* 15(1):104–111, 1996. doi:10.1109/51.482850.
25. Celebi, M. E., Iyatomi, H., Schaefer, G., and Stoecker, W. V., Lesion border detection in dermoscopy images. *Comput. Med. Imaging Graph.* 33:148–153, 2009.
26. Bandyopadhyay, S. K., Preprocessing of mammogram images. *Int. J. Eng. Sci. Technol.* 2(11):6753–6758, 2010.
27. Rajab, M. I., Woolfson, M. S., and Morgan, S. P., Application of region-based segmentation and neural network edge detection to skin lesions. *Comput. Med. Imaging Graph.* 28:61–68, 2004.
28. Palus, H., and Bogdanski, M., Clustering techniques in colour image segmentation. *AI-METH Artif. Intell. Methods.* 5–7, 2003.
29. Silveira M., Nascimento, J. C., Marques, J. S., Marçal, A. R. S., Mendonça, T., Yamauchi, S., Maeda, J., and Rozeira, J., Comparison of segmentation methods for melanoma diagnosis in dermoscopy images. *IEEE J. Sel. Top. Sign. Process*. 3(1), 2009.
30. Yuan, X., Situ, N., and Zouridakis, G., A narrow band graph partitioning method for skin lesion segmentation. *Pattern Recogn.* 42:1017–1028, 2009.
31. Saripan, Azmi, and Abdullah, et al., Illumination compensation in pig skin texture using local-global block analysis. *Mod. Appl. Sci*. 3(2), 2009.
32. Cai, Yang, and Cao, et al., A new iterative triclass thresholding technique in image segmentation. *IEEE Trans. Image Process*. 23(3), 2014.
33. Erkol, B., Moss, R. H., Stanley, R. J., Stoecker, W. V., and Hvatum, E., Automatic lesion boundary detection in dermoscopy images using gradient vector flow snakes. *Skin Res. Technol.* 11(1):17–26, 2005.
34. Celebi, M. E., Hassan, A. K., Iyatomi, H., Lee, J. K., Aslandogan, Y. A., Stoecker, W. V., Moss, R., Joseph, M. M., and Marghoob, A. A. Fast and accurate border detection in dermoscopy images using statistical region merging. *Med. Imaging*. 65123V–65123V, 2007.
35. Celebi, M. E., Hwang, S., Hitoshi, I., and Schaefer, G. Robust border detection in dermoscopy images using threshold fusion. *17th IEEE International Conference on Image Processing (ICIP)*. 2541–2544, 2010.
36. Celebi, M. E., Kingravi, H. A., Iyatomi, H., Aslandogan, Y. A., Stoecker, W. V., Moss, R. H., Malters, J. M., Grichnik, J. M., Marghoob, A. A., Rabinovitz, H. S., and Menzies, S. W., Border detection in dermoscopy images using statistical region merging. *Skin Res. Technol.* 14(3):347–353, 2008.
37. Sikorski, J., Identification of malignant melanoma by wavelet analysis. *Proceedings of Student/Faculty Research Day, Pace University*. 2004.
38. Chiem, A., Al-Jumaily, A., and Khushaba, N. R., A novel hybrid system for skin lesion detection. *Proceedings of the 3rd International Conference on Intelligent Sensors. Sensor Networks and Information Processing (ISSNIP'07)*. 567–572, 2007.
39. Maglogiannis, I., Zafiropoulos, E., and Kyranoudis, C., Intelligent segmentation and classification of pigmented skin lesions in dermatological images. In: *Advances in artificial intelligence*. Springer, Berlin, pp. 214–223, 2006.
40. Tanaka, T., Torii, S., Kabuta, I., Shimizu, K., Tanaka, M., and Oka, H., Pattern classification of nevus with texture analysis. *Proceedings of the 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC'04)*. 1459–1462, 2004.
41. Zhou, H., Chen, M., and Rehg, J. M., Dermoscopic interest point detector and descriptor. *Proceedings of the 6th IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI'09)*. 1318–1321, 2009.
42. Lee, C., and Landgrebe, D. A., Feature extraction based on decision boundaries. *IEEE Trans. Pattern Anal. Mach. Intell.* 15(4):388–400, 1993.
43. Anuradha, K., and Sankaranarayanan, K., Statistical Feature extraction to classify oral cancers. *J. Glob. Res. Comput. Sci.* 4(2):8–12, 2013.
44. Duda, R., Hart, P., and Stork, D., *Pattern classification*, 2nd edition. Wiley, New York, 2001.
45. Vanitha, L., and Venmathi, A. R., Classification of medical images using support vector machines. *Int. Conf. Inf. Netw. Topol*. 4, 2011.
46. Lau, H. T., and Al-Jumaily, A., Automatically early detection of skin cancer: study based on neural network classification. *Int. Conf. Soft Comput. Pattern Recognit*. 375–380, 2009.
47. Kilic, N., and Hosgormez, E., Automatic estimation of osteoporotic fracture cases by using ensemble learning approaches. *J. Med. Syst.* 40(3):61, 2016. doi:10.1007/s10916-015-0413-1. **Epub 2015 Dec 12**.
48. Mandal, I., and Sairam, N., Accurate prediction of coronary artery disease using reliable diagnosis system. *J. Med. Syst.* 36(5):3353–3373, 2012.