AUTOMATIC SKIN CANCER IMAGES CLASSIFICATION

Mahmoud Elgamal

Institute of Hajii and umra., Um Alqura Uni.

Saudii Arabia

Email: mahmoud_40@yahoo.com

Abstract—**Early detection of skin cancer has the potential to reduce mortality and morbidity. This paper presents two hybrid techniques for the classification of the skin images to predict it if exists. The proposed hybrid techniques consists of three stages, namely, feature extraction, dimensionality reduction, and classification. In the first stage, we have obtained the features related with images using discrete wavelet transformation. In the second stage, the features of skin images have been reduced using principle component analysis to the more essential features. In the classification stage, two classifiers based on supervised machine learning have been developed. The first classifier based on feed forward back-propagation artificial neural network and the second classifier based on k-nearest neighbor. The classifiers have been used to classify subjects as normal or abnormal skin cancer images. A classification with a success of 95% and 97.5% has been obtained by the two proposed classifiers and respectively. This result shows that the proposed hybrid techniques are robust and effective.**

Keywords: **Skin cancer images, wavelet transformation, principle component analysis, feed forward back-propagation network, k-nearest neighbor classification.**

I. INTRODUCTION

Skin cancer is a major public health problem in the light-skinned population. Skin cancer is divided into *non melanoma skin cancer* (NMSC) and *melanoma skin cancer* (MSC) figure (1.1). Non melanoma skin cancer (MMSC) is the
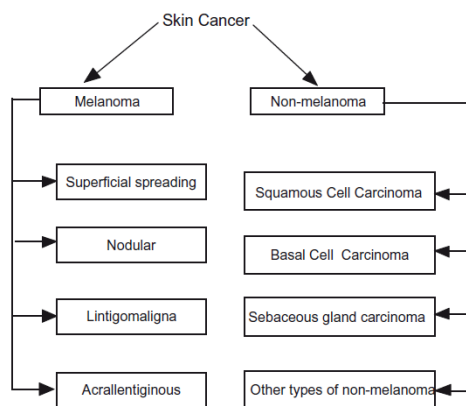

Fig. 1.1. Types of skin cancer.

most prevalent cancer among light-skinned population. It is divided into basal cell carcinoma (BCC) (75%), squamous cell carcinoma (SCC) (24%), and other rare types (1%) such as sebaceous carcinoma[19]. The critical factor in assessment of patient prognosis in skin cancer is early diagnosis. More than 60,000 people in the United States were diagnosed with invasive melanoma in recent years, and more than 8000 Americans died of the disease. The single most promising strategy to cut acutely the mortality rate from melanoma is early detection. Attempts to improve the diagnostic accuracy of melanoma have spurred the development of innovative in-vivo imaging modalities, including total body photography, dermoscopy, automated diagnostic system and reflectance confocal microscopy. The use of computer technology in medical decision support is now widespread and pervasive across a wide range of medical area, such as cancer research, gastroenterology, hart diseases, brain tumors etc. [10], [12]. Recent work [19] has shown that skin cancer recognition from images is possible via supervised techniques such as artificial neural networks and fuzzy systems combined with feature extraction techniques. Other supervised classification techniques, such as k-nearest neighbors ($k-NN$) also group pixels based on their similarities in each feature image [5], [8], [9] can be used to classify the normal/abnormal images. Therefore image processing become our choice for an early detection of the skin cancer, as it is non-expensive technique. The identification of the edges of an object in an image scene is an important aspect of the human visual system because it provides information on the basic topology of the object from which an interpretative match can be achieved. In other words, the segmentation of an image into a complex of edges is a useful prerequisite for object identification. However, although many low-level processing methods can be applied for this purpose, the problem is to decide which object boundary each pixel in an image falls within and which high level constraints are necessary. In this paper, supervised machine learning techniques, i.e., Feedforawd-Backpropagation Neural Network($FP-ANN$) used as a classifier. Also unsupervised classification techniques such as $k-NN$ used as a classify images into normal/abnormal as will be discussed in the subsequent sections. Wavelet transform is an effective tool for feature extraction, because they allow analysis of images at various levels of resolution. This technique requires large storage and is computationally more expensive [12]. In order to reduce the feature vector dimension obtained from wavelet transformation and increase the discriminative power, the principal component

analysis (PCA) has been used; PCA reduces the dimensionality of the data and therefore reduces the computational cost of analyzing new data.

This paper is organized as follows: a short description of the images preprocessing section (II), details of the proposed hybrid techniques sections (III) and (IV), section (V) contains simulation and results and finally the conclusion.

II. SKIN CANCER IMAGES PREPROCESSING

There are many types of the skin cancer, each type has a different color, size and features. Many skin features may have impact on digital images like hair and color, and other impacts such as lightness, and type of the scanner or digital camera. In the preprocessing step, the border detection procedure namely, color space transformation, contrast enhancement, and artifact removal, treated as follow [2]:

i) Color space transformation Dermoscopy images are commonly acquired using a digital camera with a dermoscope attachment. Due to the computational simplicity and convenience of scalar (single channel) processing, the resulting RGB (red-green-blue) color image is often converted to a scalar image using one of the following methods:

- Retaining only the blue channel (lesions are often more prominent in this channel).
- Applying the luminance transformation, i.e.

$$Luminance = 0.299 \times Red + 0.587 \times Green + 0.114 \times Blue.$$

- Applying the *Karhunen-Love* (KL) transformation [18] and retaining the channel with the highest variance.

ii) Contrast enhancement Delgado et al. [6], proposed contrast enhancement method, based on *independent histogram pursuit* (IHP). This algorithm linearly transforms the original RGB image to a decorrelated color space in which the lesion and the background skin are maximally separated. Border detection is then performed on these contrast-enhanced images using a simple clustering algorithm.

iii) Artifact removal Dermoscopy images often contain artifacts such as such as black frames, ink markings, rulers, air bubbles, as well as intrinsic cutaneous features that can affect border detection such as blood vessels, hairs, and skin lines. These artifacts and extraneous elements complicate the border detection procedure, which results in loss of accuracy as well as an increase in computational time. The most straightforward way to remove these artifacts is to smooth the image using a general purpose filter such as the *Gaussian*(GF), *median*(MF), or *anisotropic diffusion filters*(ADF).

Images are processed to have the following starting features: same size, color segmentations to remove hair if any as discussed in the next subsection.

A. Segmentation

Segmentation refers to the partitioning of an image into disjoint regions that are homogeneous with respect to a chosen property such as luminance, color, texture, etc. ([3] and [20]). Segmentation methods can be roughly classified into the following categories:

- Histogram thresholding: These methods involve the determination of one or more histogram threshold values that separate the objects from the background.
- Clustering: These methods involve the partitioning of a color (feature) space into homogeneous regions using unsupervised clustering algorithms.
- Edge-based: These methods involve the detection of edges between the regions using edge operators.
- Region-based: These methods involve the grouping of pixels into homogeneous regions using region merging, region splitting, or both.
- Morphological: These methods involve the detection of object contours from predetermined seeds using the watershed transform.
- Model-based: These methods involve the modeling of images as random fields whose parameters are determined using various optimization procedures.
- Active contours (snakes and their variants): These methods involve the detection of object contours using curve evolution techniques.
- Soft computing: These methods involve the classification of pixels using soft-computing techniques including neural networks, fuzzy logic, and evolutionary computation. Figure (2.1) shows a segmentation of *Superficial Spreading Melanoma* (SSM) image.

III. THE PROPOSED HYBRID TECHNIQUES

The proposed hybrid techniques based on the following techniques, discrete wavelet transforms DWT, the principle components analysis PCA, FP-ANN, and k-NN. It consists of the following phases: feature extraction, feature reduction, and classification phase as illustrated in Fig. (3.1).
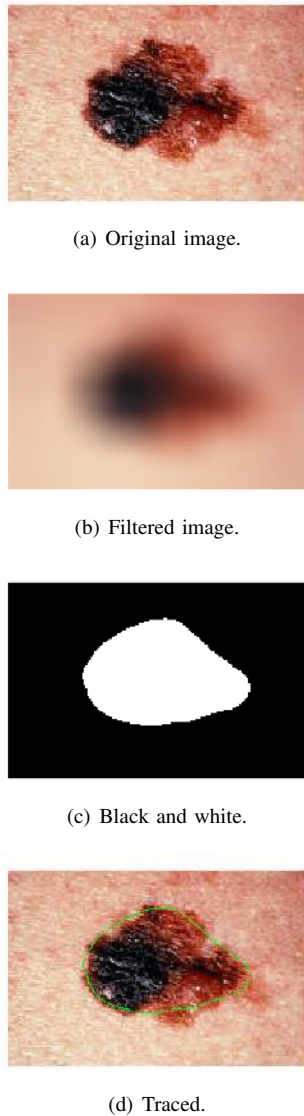
(a) Original image.

(b) Filtered image.

(c) Black and white.

(d) Traced.

Fig. 2.1. Superficial spreading image segmentation.

Fig. 3.1. Flow-diagram of the proposed technique.

A. DWT based feature extraction

The rationale of the approach proposed here implies two steps. In the first step, signals (tissue samples) are transformed into time-scale domain by DWT. In the second step features are extracted from the transformed data by a principle component analysis(PCA). The continuous wavelet transform of a signal $x(t)$ is defined as the sum over all time of the signal multiplied by wavelets $\psi_{ab}(t)$ (see Eq. (2)), the scaled (stretched or compressed) and shifted versions of the mother wavelet $\psi(t)$:

$$WT_x(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t)\psi^*\left(\frac{t-b}{a}\right)dt, \qquad (1)$$

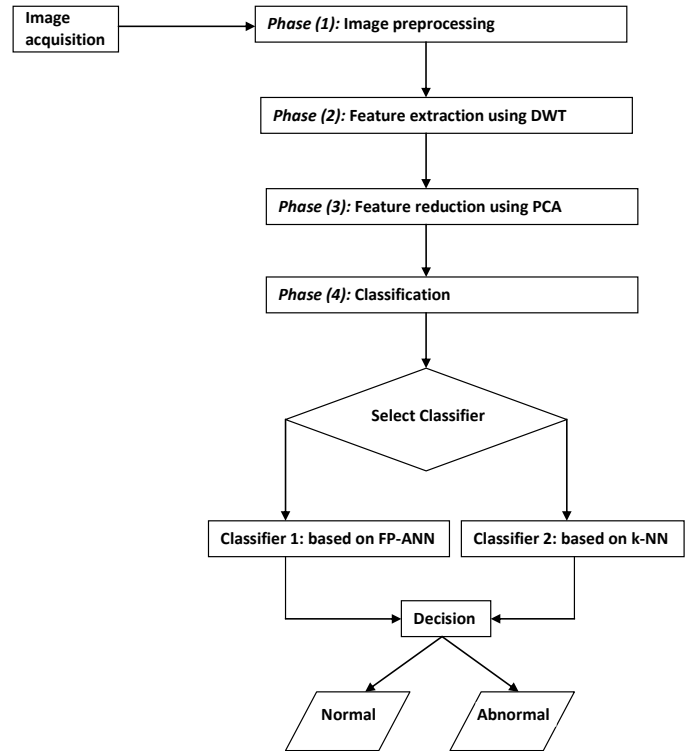$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right), \ a \in R^+, \ b \in R, \qquad (2)$$

where $\psi^*(.)$ represents the conjugated transpose of the wavelet function$\psi(.)$. The time-wavelet resolution depends on the scaling parameter a. For smaller c, $\psi_{ab}(t)$ has a narrow time-support and therefore a wider frequency support. When parameter a increases, the time support of $\psi_{ab}(t)$ increases as well and the frequency support becomes narrower. The translation parameter b determines the localization of $\psi_{ab}(t)$ in time. The DWT is defined taking discrete values of a and b. The full DWT for signal $x(t)$ can be represented as

$$x(t) = \sum_{k \in Z} \mu_{j_0,k}\phi_{j_0,k}(t) + \sum_{j=-\infty}^{j_0} \sum_{k \in Z} \omega_{j,k}\psi_{j,k}(t), \quad (3)$$

where $\phi_{j_0,k}(t)$ and $\psi_{j,k}(t)$ are the flexing and parallel shift of the basic scaling function $\phi(t)$ and the mother wavelet function $\psi(t)$, and $\mu_{j_0,k}(j < j_0)$ and $\omega_{j,k}$ are the scaling coefficients and the wavelet coefficients, respectively. Generally, scales and positions are based on powers of 2, which is the dyadic DWT. Once a mother wavelet is selected, the wavelet transform can be used to decompose a signal according to scale, allowing separating the fine-scale behavior (detail) from the large-scale behavior (approximation) of the signal [1]. The relationship between scale and signal behavior is: low scale corresponding to compressed wavelet, and to rapidly changing details, namely high frequency; high scale corresponding to stretched wavelet, and to slowly changing coarse features, namely low frequency. Signal decomposition is typically done in an iterative fashion using the scales $a = 2, 4, 8, \cdots, 2^L$, with successive approx-
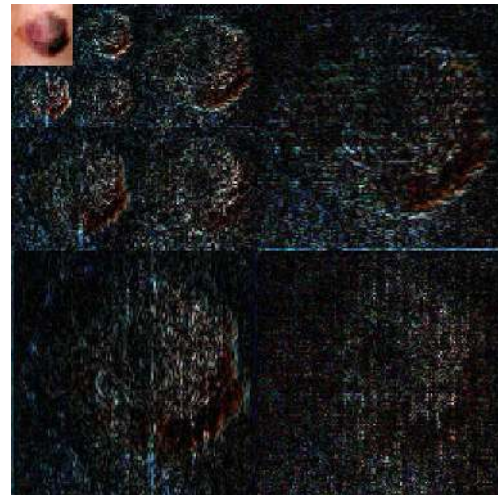
(a) forward 2D DWT.
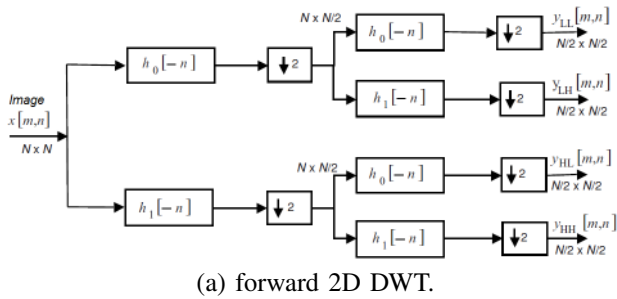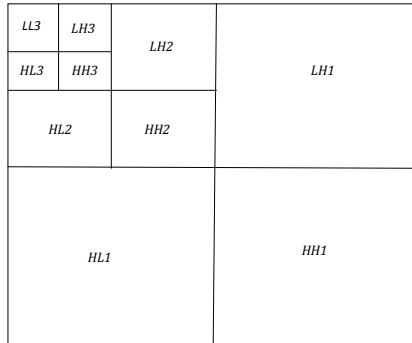
(b) approximation and detail coefficient placement.

Fig. 3.2. Computation of a three-level 2D DWT via subband coding scheme.

(a) Superficial spreading image.

(b) DWT of the image.

Fig. 3.3. A three-level 2D DWT of the tissue image.

imations being split in turn, so that one signal is broken down into many lower resolution components. In practice, signal decomposition can be implemented in a computationally efficient manner via the fast wavelet transform developed by Mallat [15], behind which the basic idea is to represent the wavelet basis as a set of high-pass and low-pass filters in a filter bank, as shown in figure (3.2). Figure (3.3) shows the three-level DWT of SSM image. The first level consists of LL, HL, LH, and HH coefficients. The HL coefficients correspond to high-pass in the horizontal direction and low-pass in the vertical direction. Thus, the HL coefficients follow horizontal edges more than vertical edges. The LH coefficients follow vertical edges because they correspond to high-pass in the vertical direction and low-pass in the horizontal direction. As a result, there are 4 sub-band (LL, LH, HH, HL) images at each scale. The sub-band LL is used for the next 2D DWT. The LL subband can be regarded as the approximation component of the image, while the LH, HL, and HH subbands can be regarded as the detailed components of the image. As the level of the decomposition increased, we obtained more compact yet coarser approximation components. Thus, wavelets provide a simple hierarchical framework for interpreting the image information. In our algorithm, level-3 decomposition via Haar wavelet was utilized to extract features.

B. PCA based feature reduction

The Principal Component Analysis (PCA) is one of the most successful techniques that have been used in image recognition and compression. The purpose of PCA is to reduce the large dimensionality of the data space (observed variables) to the smaller intrinsic dimensionality of feature space (independent variables), which are needed to describe the data economically. This is the case when there is a strong correlation between observed variables. The jobs which PCA can do are prediction, redundancy removal, feature extraction, data compression, etc. Because PCA is a classical technique which can do something in the linear domain, applications having linear models are suitable, such as signal processing, image processing, system and control theory, communications, etc. Given a set of data, PCA finds the linear lower-dimensional representation of the data such that the variance of the reconstructed data is preserved ([13] and [14]). Using a system of feature reduction based on a combined PCA on the feature vectors that calculated from the wavelets limiting the feature vectors to the component selected by the PCA should lead to an efficient classification algorithm utilizing supervised approach. So, the main idea behind using PCA in our approach is to reduce the dimensionality of the wavelet coefficients. This leads to more efficient and accurate classifier. In figure(3.4) given below an outline of PCA-algorithm.

Input: Data set(matrix) $X_{m \times n}$, ($m \equiv$ # parameters and $n \equiv$ # samples).
Output: $Y = PX$, $P \equiv$ transformation matrix from X to Y; basis for X, s.t.,
the covariance matrix in this base is diagonalized.

Step 1: Calculate the covariance matrix in the new basis $C_y = \frac{1}{n-1} PAP^T$,
$A = XX^T$; A is symmetric.
Step 2: Diagonalize the cov. matrix, i.e., $A = EDE^T$.
Step 3: Find the eigenvalues and eigenvectors from the previous step;
$A = EDE^T$, where
D = diagonal matrix contains the eigenvalues of A,
E = matrix contains the eigenvectors of A.
Step 4: Dimension reduction:
• rearrange the eigenvectors and eigenvalues, i.e., sort the columns of the
eigenvector matrix E and eigenvalue matrix D in order of decreasing
eigenvalue.
• select a subset of the eigenvectors with higher values as basis vectors,
project the data onto the new basis; of lower dimensions.

Fig. 3.4. PCA algorithm.

Therefore, the feature extraction process was carried out through two steps: firstly the wavelet coefficients were extracted by the DWT and then the essential coefficients have been selected by the PCA (see Fig. 3.5).
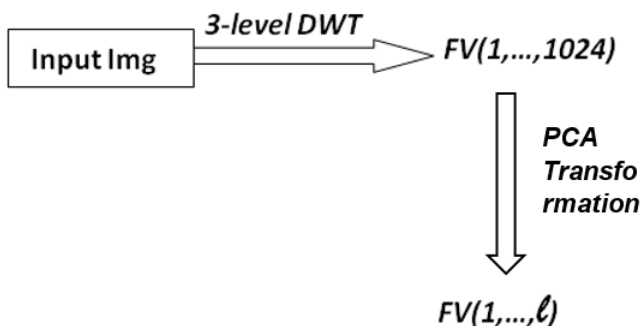

Fig. 3.5. Feature extraction of image using DWT and reduction using PCA(ℓ is new feature vector dimension).

IV. SUPERVISED LEARNING CLASSIFICATION

Supervised machine learning is the search for algorithms that reason from externally supplied instances to produce general hypotheses, which then make predictions about future instances. In other words, the goal of supervised learning is to build a concise model of the distribution of class labels in terms of predictor features. The resulting classifier is then used to assign class labels to the testing instances where the values of the predictor features are known, but the value of the class label is unknown. In this paper two supervised learning algorithms are employed for classification, the first is *k-Nearest neighbors* and *Artificial neural network*, a details of both will be given below.

A. k-Nearest neighbors(k–NN):

One of the most straightforward instance-based learning algorithms is the nearest neighbor algorithm. k–NN is based on the principle that the instances within a dataset will generally exist in close proximity to other instances that have similar properties [4]. If the instances are tagged with a classification label, then the value of the label of an unclassified instance can be determined by observing the class of its nearest neighbors. The $k - NN$ locates the k nearest instances to the query instance and determines its class by identifying the single most frequent class label. Figure (4.1) shows a description of the algorithm. The training phase for $k - NN$ consists of simply storing all known instances and their class labels. A tabular representation can be used, or a specialized structure such as a kd-tree. If we want to tune the value of k and/or perform feature selection, n-fold cross-validation can be used on the training dataset. The testing phase for a new instance t, given a known set I is as follows:

1) Compute the distance between t and each instance in I.
2) Sort the distances in increasing numerical order and pick the first k elements.
3) Compute and return the most frequent class in the k nearest neighbors, optionally weighting each instance's class by the inverse of its distance to t.

Fig. 4.1. k–NN algorithm.

B. Artificial Neural Network (ANN):

An artificial neural network (ANN), usually called neural network (NN), is a mathematical model or computational model that is inspired by the structure and/or functional aspects of biological neural networks. A neural network consists of an interconnected group of artificial neurons, and it processes information using a connectionist approach to computation. In most cases an ANN is an adaptive system that changes its structure based on external or internal information that flows through the network during the learning phase [21]. The feed-forward neural network was the first and arguably simplest type of artificial neural network devised. In this network, the

information moves in only one direction, forward, from the input nodes, through the hidden nodes (if any) and to the output nodes figure(4.2). There are no cycles or loops in the network. The neural network which was employed as the
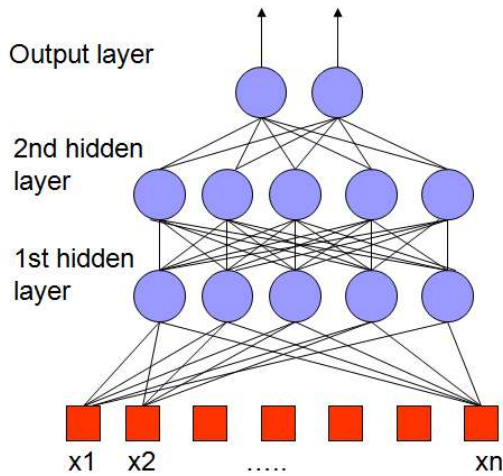

Fig. 4.2. Feed forward neural network.

classifier required in this study had three layers, as shown in figure (4.2). The first layer consists of eight input elements in accordance with the size of the feature vectors that selected from the wavelet coefficients by the PCA. The number of neurons in the hidden layer was five. The single neuron in the output layer was used to represent normal and abnormal tissue image (see Fig. 4.2). The most frequently used training algorithm in classification problems is the back-propagation (BP) algorithm, which is used in this paper too. The details of back-propagation (BP) algorithm are well documented in the literature [7]. The neural network has been trained to adjust the connection weights and biases in order to produce the desired mapping. At the training stage, the feature vectors are applied as input to the network and the network adjusts its variable parameters, the weights and biases, to capture the relationship between the input patterns and outputs [7]. The performance is measured by mean square error(MSE):

$$MSE = \frac{1}{n} \sum^{m} (y_t - y_o)^2, \qquad (4)$$

here y_t is the target value, y_o is the actual output, and m is the number of training data, figure(4.3) shows ANN Back-propagation algorithm.

V. SIMULATION:

In this section the proposed hybrid techniques are used for the classification of malignant melanoma. The data set consists of total 40 images (20 normal and 20 abnormal) images downloaded from [11]. Figure (5.1) shows some samples from the used data. The proposed approach was implemented in Matlab 7.12[16], on a PC with the following configurations, processor 2GHz; 2 GB of ram; run under MSWin. 7 operating system. Figure (6.3) shows GUI of the implementation. The

Initialize the weights in the network (often randomly).

   Do For each example e in the training set:

   a) O = neural-net-output(network, e) ; forward pass,
   b) T = teacher output for e,
   c) Calculate error $(T - O)$ at the output units,
   d) Compute δ_{wh} for all weights from hidden layer to output layer ; backward pass,
   e) Compute δ_{wi} for all weights from input layer to hidden layer ; backward pass continued,
   f) Update the weights in the network.

   Until all examples classified correctly or stopping criterion satisfied.

Return the network.

Fig. 4.3. ANN Back-propagation algorithm.

images resized to smaller size of pixels 8×8, Haar wavelet with the third level used, and the feature vector reduced to just eight components. With those settings, the hybrid classifiers; FP-ANN and KNN were run and an instance of the output shown in Fig. (6.3).
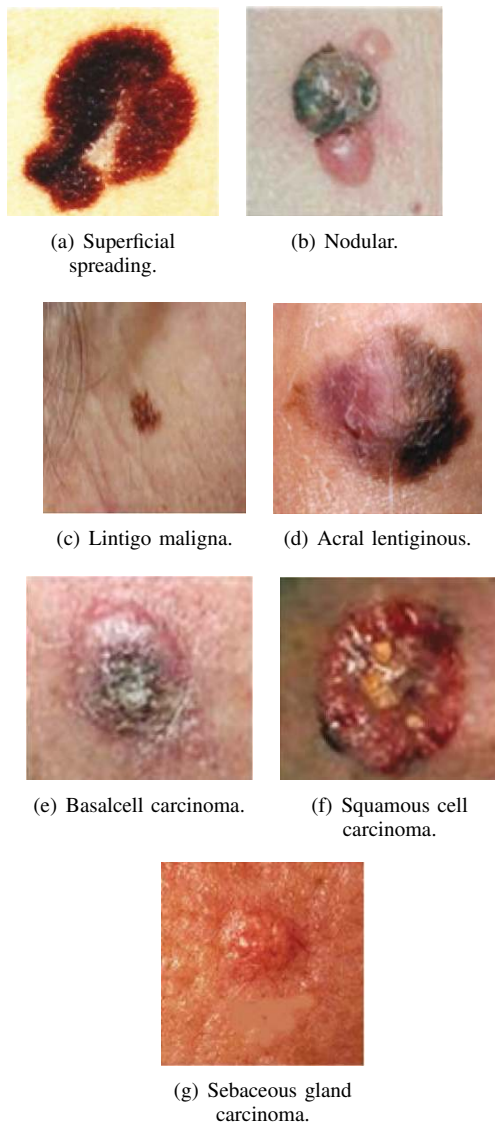
(a) Superficial spreading.

(b) Nodular.

(c) Lintigo maligna.

(d) Acral lentiginous.

(e) Basalcell carcinoma.

(f) Squamous cell carcinoma.

(g) Sebaceous gland carcinoma.

Fig. 5.1. Samples of skin cancer.

A. Performance Evaluation

In this section, the performance of the proposed techniques are evaluated for the skin cancer images. The proposed techniques performance evaluated in terms of confusion matrix, sensitivity, specificity, and accuracy. The three terms defined as follow [17]:

a) Sensitivity(true positive fraction): the result indicates positively(disease).

$$Sen. = \frac{TP}{(TP + FN)}$$

b) Specificity(true negative fraction): the result indicates negatively(non-disease).

$$Spec. = \frac{TN}{(TN + FP)}$$

c) Accuracy: the probability that the diagnostic test is performed correctly.

$$Accur. = \frac{(TP + TN)}{(TP + TN + FP + FN)}.$$

Where,
TP (True Positives) $\equiv$ Correctly classified positive cases,
TN (True Negative)$\equiv$ Correctly classified negative cases,
FP (False Positives) -Incorrectly classified negative cases, and
FN (False Negative)$\equiv$ Incorrectly classified positive cases.

Prop. Tech.	TP	TN	FP	FN	Sen.(%)	Spec.(%)	Accur.(%)
ANN	19	19	1	1	95	95	95
K-NN	20	19	1	0	100	95	97.5

Fig. 5.2. Performance of the used techniques.

VI. CONCLUSION

In this paper, an automated medical decision support system for skin cancer developed with normal and abnormal classes. First the discrete wavelet transformation were applied on the images to get the feature vectors, as the dimensionality of the vectors quite large, one needed to reduce it through the principle component analysis. The resulting feature vectors have a few components; means, less time and memory requirements. Afterwards, those vectors were used for classification either with feed-forward neural network or k-nearest neighbor algorithm. The results of the deployed techniques were promising as, we got 100% for sensitivity, 95% for specificity , and 97.5% for accuracy.

REFERENCES

[1] L.M. Bruce, C.H. Koger, and J. Li, *Dimensionality reduction of hyperspectral data using discrete wavelet transform feature extraction*, IEEE Trans. Geosci. Remote Sensing 40 (10) (2002) 10.

[2] M. E. Celebi, H. Iyatomi, G. Schaefer, and W. V. Stoecker, *Lesion Border Detection in Dermoscopy Images*, Computerized Medical Imaging and Graphics 33 (2009).

[3] M. E Celebi, H. A. Kingravi, and H.Iyatomi, *Border detection in dermoscopy images using statistical region merging*, Skin Research and Technology, 2008;14(3):34753.

[4] T. Cover, P. Hart, *Nearest neighbor pattern classification*, IEEE Transactions on Information Theory, 1967; 13(1): 217.

[5] I. Daubechies, *Ten Lectures on Wavelets*, Regional Conference Series in Applied Mathematics, SIAM, Philadelphia, 1992.

[6] D. Delgado, C. Butakoff, BK. Ersboll, and WV. Stoecker,*Independent histogram pursuit for segmentation of skin lesions*, IEEE Trans on Biomedical Engineering 2008;55(1):15761.

[7] Jerome Feldman, *Neural Networks - A Systematic Introduction*, Springer-Verlag, Berlin, New-York, 1996.

[8] P.S.Hiremath, S. Shivashankar, and J. Pujari,Wavelet Based Features for Color Texture Classification with Application to CBIR, IJCSNS International Journal of Computer Science and Network Security, 6 (2006), pp. 124- 133.

[9] K. Karibasappa, S. Patnaik, *Face Recognition by ANN using Wavelet Transform Coefficients*, IE(I) Journal-CP, 85,(2004), pp. 17-23.

[10] F. Gorunescu, *Data Mining Techniques in Computer-Aided Diagnosis: Non-Invasive Cancer Detection*, PWASET Volume 25 November 2007 ISSN 1307-6884, PP. 427-430.

[11] Is it skin cancer? 38 photos that could save your life, *http://www.cbsnews.com/2300-204_162-10006772.html*.

[12] S. Kara, F. Dirgenali, *A system to diagnose atherosclerosis via wavelet transforms, principal component analysis and artificial neural networks*, Expert Systems with Applications 32 (2007), pp. 632-640.

[13] M. Kocionek, A. Materka, M. Strzelecki, and P. Szczypinski, *Discrete wavelet transform derived features for digital image texture analysis*, Proc. of Interational Conference on Signals and Electronic Systems, 18-21 September 2001, Lodz, Poland, pp. 163-168.

[14] F. Latifoglu, K. Polat, S. Kara, and S. Gunes, it Medical diagnosis of atherosclerosis from Carotid Artery Doppler Signals using principal component analysis (PCA), k-NN based weighting pre-processing and Artificial Immune Recognition System (AIRS), Journal of Biomedical Informatics 41 (2008), pp. 15-23.

[15] S.G. Mallat, *A theory for multiresolution signal decomposition: the wavelet representation*, IEEE Pattern Anal. Machine Intell. 11 (7) (1989) 674693.

[16] MatLab, *http://www.mathworks.com/*.

[17] K. Polat, B. Akdemir, S. Gnes, *Computer aided diagnosis of ECG data on the least square support vector machine*, Digital Signal Processing 18 (2008), pp. 25-32.

[18] Pratt WK. ,*Digital image processing: PIKS inside.*, Hoboken, NJ: John Wiley & Sons; 2007.

[19] B. Salah, M. Alshraideh, R. Beidas and F. Hayajneh, *Skin Cancer Recognition by Using a Neuro-Fuzzy System*, Cancer Informatics, 2011.

[20] M Sonka, V Hlavac, and R. Boyle, *Image processing, analysis, and machine vision*, Cengage-Engineering 2007.

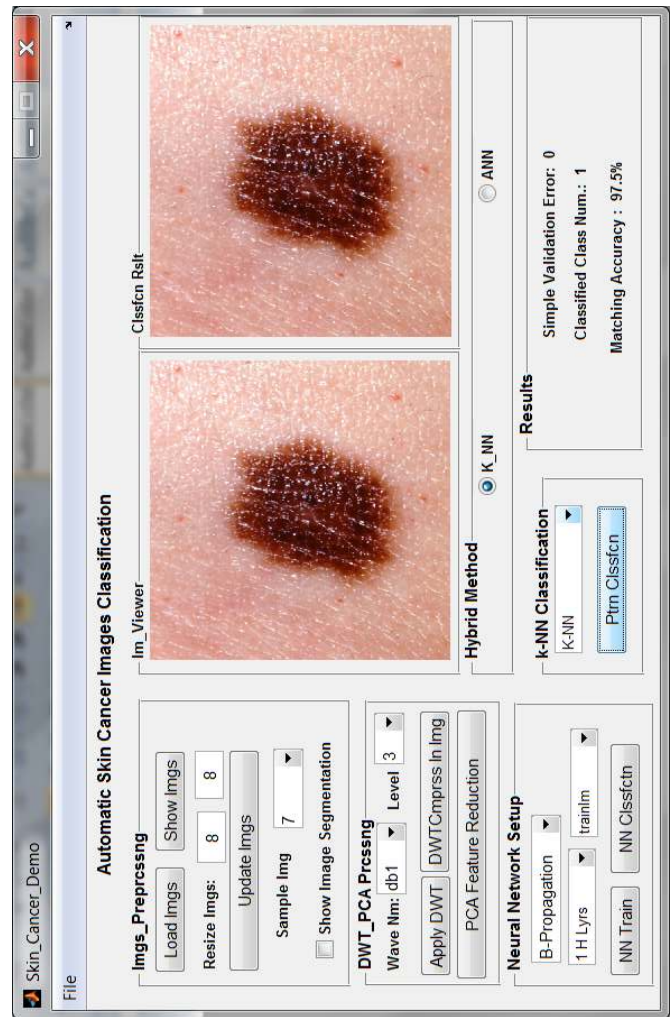[21] *http://en.wikipedia.org/wiki/Artificial_Neural_Network*, Wikipedia 2012.

Fig. 6.3. GUI of the proposed techniques.