SKIN CANCER DETECTION AND DIAGNOSIS USING IMAGE PROCESSING AND IMPLEMENTATION USING NEURAL NETWORKS AND ABCD PARAMETERS

SANTOSH ACHAKANALLI¹ & G. SADASHIVAPPA²

¹M.Tech Student, Digital Communication Engineering, R.V. College of Engineering, Bangalore, Karnataka, India
²Professor, Department of Telecommunication Engineering, Bangalore, Karnataka, India

ABSTRACT

Skin cancer is the most dangerous form of all cancers that occurs for human being. Among these skin cancers melanoma is the most deadliest form which accounts for more than 40% of all over the world. Even though melanoma is the deadliest if detected in early stages then it can be treated successfully. Day-by-Day many computer based automatic detection and diagnosis of skin cancer are developing. In this paper we present an improved method using image processing method which uses statistical features and dermoscopic features such as ABCD (Asymmetry, Border, Color and Diameter) for detection and diagnosis. The steps involved in this method are creation of database for dermoscopic images, filtering for noise removal, segmentation using thresholding, statistical feature extraction using Gray Level Co-occurrence Matrix (GLCM), Calculating Total Dermoscopy Score and then classification using Artificial Neural Network (ANN). The combined results of the neural network and the ABCD parameters indicate that the developed methodology proved effective and efficient for the skin cancer detection and diagnosis.

KEYWORDS: ABCD Parameters, Artificial Neural Network, Image Processing, Feature Extraction, Melanoma, Skin Lesion

INTRODUCTION

Due to uncontrolled growth of cells in human body leads to the cancer. Normally cells grow and divide to form new cells which is necessary for growth of a human body. When the cells in body becomes old or get damages, they usually die and a new cells take their place. But When new cells from the body doesn’t need them, the damaged cells don’t die as they should. This extra cell often forms a mass of tissue called a growth or tumor. Tumors on the skin can be benign (not cancer) or malignant Melanoma (cancer). Tumors on the skin can be benign (not cancer) or malignant Melanoma (cancer). Benign growths are not as harmful as malignant growths [20]. Melanoma is the deadliest form of skin cancer affecting the human skin and gradually spreads whole body. It arises from cancerous growth in pigmented skin area. It is a gradually spreading condition; this begins in the melanocytes (Types of cells) in skin [1]. If the skin cancer is not diagnosed at its early stages, it can cause death of the patient, so early detection of skin cancer is unavoidable.

The rate of detection of melanoma using dermoscopy is higher than detection only with unaided observation [8] as dermoscopy is a non-invasive diagnosis technique. The diagnostic accuracy of dermoscopy also depends on the expertism of the doctors, So that the Computer Aided diagnosis based on image processing technique will help for automatic diagnosis which is essential tool for learners and less experienced doctors. It considered to be a “Reference model” system where doctors take into consideration the information provided by computer before making decision [5]. In this paper a Computer Aided Diagnosis of skin cancer is proposed. This proposed methodology uses both artificial neural network and...
ABCD parameters for detection and diagnosis of skin cancer images. The neural network provides classification of dermoscopic image being a cancerous or non-cancerous[5]. Then the Total Dermoscopy Score of the same dermoscopy image is calculated by extracting medical features such as ABCD parameters. TDS is defined by $TDS = A*1.3 + B*0.1 + C*0.5 + D*0.5$, with a maximum score of 8.9, where: A is Asymmetry (0-2 points), B is equal to Border irregularity (0-8 points), C implies Color composition (0-6 points), D = Diameter or Differential structures (0-5 points). Higher the ABCD score the lesion is more likely to be a malignant melanoma ($TDS > 5.45$, [21]). This explained computer based automatic skin cancer diagnosis system is implemented in commonly available software - MATLAB® 2013(a).

PREAMBLE

Over the past years many research papers applying several techniques have been proposed for diagnosis and detection of melanoma a deadliest form of cancer. The dermoscope method was first described in the year 1987 [9]; which describes non-invasive diagnosis process which is based on using of oil immersion, magnifier and incident light. But the accuracy of this method depends on doctor’s experience. The clinical diagnosis of melanoma is not always easy even for those professing to have a special interest in the disease [Mihm 1997, Barnhill 1997]. A study carried out at St. George’s hospital bravely concluded that clinicians are only 50% correct when their diagnosis are compared with those of histopathology [Curley et al. 1989].

Another study of over 44,000 lesions from the university of Graz [Wolf et al. 1997] has shown a sensitivity of clinical diagnosis for melanoma of 70% - i.e. 1/3 of melanoma are missed by clinical criteria alone.

In year 2003, another work by M. Wiltgen, A. Gerger and J. Smolle uses a method of tissue counter analysis (TCA), which is based on the partition of whole image into square elements of equal size and then features, are calculated from these square elements of the image. The features, based on GLCM (Grey level co-occurrence matrix) and grey level histogram, allow the differentiation of homogeneous and high contrast or luminous tissue areas. The highest accuracy of classification obtained by this approach was 92.7%. Fatima, R. [11] introduces a multi-parameter extraction and classification system to aid an early detection skin cancer melanoma. Fasihi, N. [13] describes utilization of morphologic operators for segmenting a image and wavelet analysis to extract the feature which results in to better melanoma diagnosis system. Alcon, J.F. [12] has used pigmented skin lesion’s images, captured using digital camera for automatic melanoma diagnosis with an accuracy of 85%, sensitivity of 93% and specificity of 69%. PatwardhanJain, Y. K. [10] focuses on the development a skin cancer screening system that can be used by non-experts to classify normal from abnormal cases, using feature calculation and classification techniques. Here features are extracted using wavelet transform where as the classification is done using artificial neural networks. S. V. [14] uses wavelet transformation based skin lesion images classification system which utilizes a semantic representation of spatial frequency information contains in the skin lesion images.

In this proposed method we utilizing the above defined literature to detect and diagnose skin cancer severity. Here in this work we are proposing a neural network based classifier for cancer severity quantification of skin cancer image and also the clinical parameters such as ABCD for diagnosis. The proposed system will not only extract features from GLCM but along with three other features like ‘Skewness’, ‘Mean’ and ‘Kurtosis’, along with TDS range, which leads to overcome different problem cited by above literature.
METHODOLOGY

The proposed methodology of melanoma detection and diagnosis is shown in Figure 1. The method uses the following steps: Image acquisition from digital camera, Pre-processing involves noise removal, feature extraction both statistical and dermoscopic, feature selection, calculating TDS, classification using neural network and then evaluation.

![Figure 1: Flow Chart to Illustrate Set of Procedure for Skin Fault Analysis](image.png)

- **Image Acquisition**
  
  Image acquisition is defined as the process of capturing or retrieving an image from a camera, so it can be passed through various processes need to occur later. Image acquisition in image processing is the first step in the workflow sequence since, without an image, no further process is possible.

- **Image Pre-processing**
  
  The pre-processing is a modification of the image data that removes unwanted distortions or enhances some image features important for further processing. First, the image is resized to adjust images to a uniform scale (512x512) so that it supports feature classification with accuracy. After that, images are converted from RGB (Red, Green, and Blue) to grey level where the features are based on grey level co-occurrence matrix (GLCM).

- **Image Segmentation**
  
  Segmentation of image deals with the process of partitioning a digital image into multiple segments (sets of different pixels) the goal of this process is to simplify and/or change the representation of an image into something that is more meaningful and easier to analyze.[16]. Image segmentation is typically used to locate objects and boundaries (lines, curves, etc.) in images.
More accurately, image segmentation is the process of assigning a label to every pixel in an image such that each pixel with the same label share certain visual characteristics, using thresholding method for segmentation.

- **Feature Extraction**

  Feature extraction is to reduce the original data set by measuring certain features, or properties, that distinguish between input patterns. The features extracted in this work are based on texture analysis using GLCM. The GLCM is a powerful tool for image feature extraction by mapping the grey level co-occurrence probabilities based on spatial relations of pixels in different angular directions. The features extracted from GLCM are: Energy, Correlation, Homogeneity, Contrast, Along with other statistical parameters such as skewness, kurtosis and mean.

- **Classification Using Artificial Neural Network**

  The segmented image is classified based on the extracted features, such as, the color and texture. The design of classifiers were investigated in many studies, most of them are Artificial Neural Networks (ANN). The neural network classifier is used for classifying Malignant Melanoma from other skin diseases such as benign. Based on the computational simplicity ANN based classifier is used [17]. ANN is able to solve highly complex problems due to the nonlinear processing capabilities of its nodes (neurons). In this proposed system, a feed forward multilayer network is used and Backpropagation (BPN) Algorithm is used for training. The neural network structure consists of Input layer, in the middle Hidden layer and Output layer. The hidden layer and output layer adjusts weights value based on the error output in classification of different features. In Back Propagation algorithm, signal flow will be in forward direction. Each time output of the network is compared with desired output, if both do not match, An error signal is generated at the output. This error signal is propagated backwards and weights are adjusted at the middle layer so as to reduce the error. In this Back propagation algorithm, the hidden layer and output layer weights are initialized randomly at the beginning of training. Supervisory learning process is used for training. During forward pass of the input signal, according to the initial
weights and activation function used, the neural network gives an output and is compared with required output. If both do not match, an error occurs.

During reverse pass, the error signal is back-propagated and weights of both hidden and output layer are adjusted. The whole process of training will continues until error is zero. The neural network is trained with desired known values. After training, neural network can perform decision making.

In this proposed methodology, Eight Features were given as input to a multilayer feed forward network as shown in figure 4. Along with seven statistical features shown in figure the medical feature TDS is also given as input to neural network. There is one hidden layer between input and output layers having four hidden neurons. And at the output of neural network the output layer with one output neuron, gives output as 0 or 1, 0 represents non-cancerous and 1 represents cancerous.

![Figure 4: Structure of Artificial Neural Network](image)

Matlab is used for implementation including image preprocessing and for artificial neural network classification. The neural network is trained using the known statistical values and dermoscopy feature of melanoma images. The statistical values and dermoscopy feature of around 60 melanoma cancerous images and non-cancerous images are calculated and given to the classifier, which will give output as either 0 or 1.
STATISTICAL PARAMETERS

The statistical parameters considered are skewness, kurtosis and mean along with Gray Level Co-occurrence Matrix parameters such as Energy, contrast, homogeneity and correlation. Mean gives the average value of the pixels of image. The skewness describes the information of asymmetry distribution of pixels and kurtosis will show the peakedness value of the pixels. The Contrast is a measure of the intensity between neighboring pixel over the whole image and homogeneity returns a value that measures the closeness of the distribution of elements in the GLCM to the Gray Level Co-occurrence diagonal. These parameters are defined by the following equations:

\[
\text{mean}(\mu) = \frac{\sum_{x=1}^{M} \sum_{y=1}^{N} I(x,y)}{MN}
\]

\[
\text{Energy}(\sigma) = \frac{1}{MN} \sum_{x=1}^{M} \sum_{y=1}^{N} I^2(x,y)
\]

\[
\text{Skewness} = \frac{\sum_{x=1}^{M} \sum_{y=1}^{N} (I(x,y)-\mu)^3}{M \times N \times \sigma^3}
\]

\[
\text{Kurtosis}(k) = \frac{\sum_{x=1}^{M} \sum_{y=1}^{N} (I(x,y)-\mu)^4}{M \times N \times \sigma^4}
\]

This gray-level co-occurrence matrix (GLCM) is a gray-level spatial dependence matrix based calculations, which falls under the category of second-order statistics. The work conducted by Haralick et al. [18] suggested a set of 14 textual features which can be extracted from the Gray Level Co-occurrence matrix, which contains information about image textural characteristics such as contrast, homogeneity, and entropy.

The GLCM contains information about the positions of pixels in the image having similar gray level values. It is a two-dimensional array or matrix, \( P \), in which both the rows and the columns represent a set of possible image values. A gray level co-occurrence matrix \( P_d(i,j) \) is defined by first specifying a displacement vector \( d = (dx, dy) \) and counting all the pairs of pixels separated by ‘d’ having gray levels i and j. \( P_d(i,j) = n_{ij} \), where the variable \( n_{ij} \) is the number of occurrences of the pixel values (i, j) lying at distance d in the image. The grey level co-occurrence matrix \( P_d \) has dimension \( n \times n \), where \( n \) is the number of gray levels in the image. From this matrix we can derive the following statistics as texture features.

\[
\text{Contrast} = \sum_{i=1}^{n} \sum_{j=1}^{n} P_d(i,j)^2
\]

\[
\text{Homogeneity} = \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{P_d(i,j)}{1+|i-j|}
\]

GLCM correlation returns a measure of how correlated a pixel is to its neighbor over the whole image. Correlation is 1 or -1 for a perfectly positively or negatively correlated image.

MEDICAL FEATURE ANALYSIS
Melanoma skin cancer grows asymmetrically and presents a high variegation and some differential structures as a consequence of the irregular distribution of melanin across the lesion. This particular feature leads the classification of melanomas. Medical diagnostic algorithms based on the presence or absences of these features, such as the ABCD rule of dermoscopy [21], were developed to aid dermatologists.

In 1985, group from New York University [22] devised the ABCD acronym (Asymmetry, Border irregularity, Color variegation, Diameter > 6mm) in order to recognize melanoma in its early stages. It is one of the easiest symptoms to the most common signs of melanoma. Further, Stolz, W. [22] defined a new diagnosis scheme for dermatoscopic images known as the ABCD rule of dermatoscopy. The characteristics needed to diagnose a melanoma as malignant are:

Asymmetry (A): Cancerous lesions are checked for symmetry. If the lesion is Symmetric then it is benign (non-cancerous) and 0 value is assigned to A. For Cancerous (Melanoma) cases the lesion area is asymmetric and value 1 is assigned to A.

Border Irregularity (B): Most of the cancerous lesions edges are ragger, notched or blurred. Its value ranges 0 to 8. To verify Border irregularity the parameter called Compact Index is calculated (CI), it is measurement of the most popular form of barrier unanimous 2D objects. This can be determined by using the following equation:

\[ CI = \frac{P_L^2}{4\pi A_L} \]

Color (C): Cancerous skin lesion’s pigmentation is not uniform. The presence of up to six known colors must be detected - white, red, light brown, dark brown, blue-gray, and black giving 1 point to each for calculation of C score. Its value ranges 0 to 6. The above mentioned colors are described in RGB color space as follows in Table 1 [21].

<table>
<thead>
<tr>
<th>Color</th>
<th>RGB</th>
<th>rgb</th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>255,255,255</td>
<td>1.0,1.0,1.0</td>
</tr>
<tr>
<td>Black</td>
<td>0,0,0</td>
<td>0.0,0.0,0.0</td>
</tr>
<tr>
<td>Red</td>
<td>255,0,0</td>
<td>1.0,0.0,0.0</td>
</tr>
<tr>
<td>Light-Brown</td>
<td>205,133,63</td>
<td>0.80,0.52,0.25</td>
</tr>
<tr>
<td>Dark-Brown</td>
<td>101,67,33</td>
<td>0.40,0.26,0.13</td>
</tr>
<tr>
<td>Blue-Gray</td>
<td>0.134,139</td>
<td>0.0,0.52,0.54</td>
</tr>
</tbody>
</table>

In the above table 1, The RGB column describes colors using integers in the 0-255 range (8bits), while the third rgb column indicates colors in normalized form of colors in RGB column in the range of 0.0-1.0 using floating point representation (rgb = RGB/255). In this proposed algorithm, all the pixels constituting the lesion are scanned and the Euclidian distance in “rgb” coordinates D is calculated between their color and the six reference colors mentioned above. For example, for the first color, white r1g1b1 = [1.0, 1.0, 1.0] and thus:

\[ D = \text{SQRT} ((r-r1)^2 + (g-g1)^2 + (b-b1)^2) = \text{SQRT} ((r-1.0)^2 + (g-1.0)^2 + (b-1.0)^2). \]  

Diameter (D): Melanoma cancerous lesions are greater than 6mm wide. Differential structures with at least five patterns are relevant for specific types of lesions. Any other growth of a mole should be of concern. Its value ranges 0 to 5. To find the diameter, first find centroid then draw line from all the edge pixels to the pixel edges through the midpoint and averaged.

RESULTS
In this proposed work, many test images were collected from Internet and dermatology. These images were undergone filtering using Fast Median Filtering. After this process, Filtered images were segmented by the method of Maximum Entropy Threshold Segmentation. The statistical Feature Extraction technique used is GLCM. Feature extraction was done in MATLAB (2013) software. The obtained Features were given as inputs to Artificial Neural Network., which gives an output of either 0 or 1. Zero represents non-cancerous condition and one represents cancerous or malignant melanoma condition. The classifier neural network is designed using Matlab software. For classification, there were 60 cases were considered. The Neural Network Classifier classified the given data into cancerous and Noncancerous. Among that, 41 were classified as cancerous and 30 non-cancerous. There were 7 misclassifications. The Accuracy of this proposed system is 92 %.

The following steps explain the proposed methodology results:

**Step 1:** To Acquire Image from the Patient

![Figure 5: Acquired Melanocyte Image](image)

**Step 2:** Preprocessing: In this Step Test Image is Resized to 512x512 and Removal of Hair (If Any)

![Figure 6: Resized Melanocyte Image (512x512)](image)

**Step 3:** Segmentation is done Based on Thresholding in order to Get the Pixels of Cancerous Part

![Figure 7: Segmented Image to Get Cancerous Part](image)
Step 4: Converting Image to Gray Level in order to Obtain Statistical Parameters from the GLCM of the Image

![Grey image]

Figure 8: Grey Scale Converted Image

Step 5: Calculating Symmetry by Drawing to Diagonal Lines over a Lesion Area as Shown in Figure 9

![Lesion with Diagonals to Calculate Symmetry]

Figure 9: Lesion with Diagonals to Calculate Symmetry

Step 6: Obtaining statistical parameters of Grey scale image and feeding to neural network same result shown in table 2.

Step 7: To calculate Asymmetry (A) value: For the image in figure 9, The A value obtained is A= 0.34.

Step 8: To calculate Border irregularity (B) value: For the image in figure 9, The B value obtained is B=1.48.

Step 9: To calculate color(C) presence in the lesion area: for the image in figure 9,The C value is C=4,since four types of colors found adding 1 value to each color namely black, white, light brown and dark brown.

Step 10: To calculate diameter (D) value of the lesion area: For the image shown in figure 9, The diameter value is D=6.72mm.

Step 11: To calculate Total Dermoscopic Score (TDS) for the above ABCD parameters by using formula,

\[ TDS = A*1.3 + B*0.1 + C*0.5 + D*0.5 \]  (8)

For the above ABCD parameters the TDS value is TDS=5.85, which is greater than 5.45 hence given image is cancerous.

Thus for the entire input image the above mentioned steps from 1 to 11 has to be performed to conclude cancerous or non-cancerous. For the above image in figure 9, the over all parameters calculated is as shown in figure 10.
SIMULATIONS & RESULTS

In this work, the results obtained were validated with Diagnosis results Prepared by cancer organizations, using the conventional diagnosing Procedures. Among the collected database, 60 images given for classification, 41 were Cancerous and 19 non-cancerous according to the Dermatologist results which is taken as the reference for validation. The Neural Network based classifier and based on TDS gives the output of 38 cancerous and 22 non-cancerous conditions. There were 9 misclassifications. This obtained results show that the methodology has an accuracy of 92%. The error in classification will get reduced, as the number of sample images in the database taken for classification is increased. By including more features and improving the image processing techniques and algorithms for training network, the accuracy of classification can be increased.

Table 2: Statistical Parameters

<table>
<thead>
<tr>
<th>Mean</th>
<th>Skewness</th>
<th>Kurtosis</th>
<th>Correlation</th>
<th>Contrast</th>
<th>Energy</th>
<th>Homogeneity</th>
<th>Output Of ANN</th>
<th>Conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>155.9724</td>
<td>-3.74981</td>
<td>3.556303</td>
<td>-0.061125</td>
<td>0.1073</td>
<td>0.5445</td>
<td>0.9541</td>
<td>0</td>
<td>Non-Cancerous</td>
</tr>
<tr>
<td>264.6889</td>
<td>-0.954358</td>
<td>2.324605</td>
<td>0.072144</td>
<td>0.0977</td>
<td>0.8792</td>
<td>0.9555</td>
<td>1</td>
<td>Cancerous</td>
</tr>
<tr>
<td>95.032279</td>
<td>-0.878283</td>
<td>2.383074</td>
<td>0.512448</td>
<td>0.2278</td>
<td>0.1590</td>
<td>0.8809</td>
<td>0</td>
<td>Non-Cancerous</td>
</tr>
<tr>
<td>140.6542</td>
<td>0.309355</td>
<td>3.465311</td>
<td>0.178609</td>
<td>0.0470</td>
<td>0.4016</td>
<td>0.5842</td>
<td>1</td>
<td>Cancerous</td>
</tr>
<tr>
<td>245.1857</td>
<td>0.022998</td>
<td>2.331133</td>
<td>0.977072</td>
<td>0.0691</td>
<td>0.5987</td>
<td>0.9556</td>
<td>1</td>
<td>Cancerous</td>
</tr>
<tr>
<td>102.1157</td>
<td>-0.65</td>
<td>-0.5571</td>
<td>0.0881</td>
<td>0.1849</td>
<td>0.4088</td>
<td>0.0792</td>
<td>1</td>
<td>Cancerous</td>
</tr>
<tr>
<td>247.882</td>
<td>0.750</td>
<td>20.8333</td>
<td>0.9567</td>
<td>0.2480</td>
<td>0.0091</td>
<td>0.0040</td>
<td>0</td>
<td>Non-Cancerous</td>
</tr>
</tbody>
</table>
CONCLUSIONS

The skin cancerous detection and diagnosis using image processing, neural networks and medical features such as ABCD parameters is implemented. It proves to be a better diagnosis method than the conventional Biopsy method used by dermatologist. This image processing technique is having more advantageous to patients. The diagnosing Methodology uses Statistical features of image and medical features such as ABCD parameters to calculate TDS value. Artificial Neural Networks is used for the classification of Malignant Melanoma from non-cancerous. Dermoscopic images were collected and they are processed by various Image processing techniques. Cancerous region is segmented from healthy skin this method is called segmentation. The unique features such as statistical and dermoscopy of the segmented images were extracted using gray level co-occurrence matrix. Based on the features and the Total Dermoscop Score range the images were classified as Malignant or non-melanoma. This proposed methodology has got good accuracy of 92%. By varying the Image processing techniques and training algorithms of Neural Networks, the accuracy can be further improved for this system.

REFERENCES

2. Fikret Ercal, Anurag Chawla, William V. Stoecker, Hsi-Chieh Lee and Randy H. Moss, "Neural Network Diagnosis of Malignant Melanoma From Color Images", IEEE Transactions on Biomedical Engineering vol. 41, No. 9,1994


18. National Cancer Institute,” Melanoma and Other Skin Cancers”.