MACHINE LEARNING ALGORITHM FOR AUTOMATED SKIN CANCER PREDICTION AND PROGNOSIS

Saranya N[1], Nirmala M[2], Kiruthika P[3]
[1]UG Student, Department of ECE, K.S. Rangasamy College of Technology,
[2,3]Assistant Professor, Department of ECE, K.S. Rangasamy College of Technology

Abstract
Skin diseases are one of the most common diseases in human and its incidence is increasing dramatically. Therefore early diagnosing is a crucial issue. Even experienced doctor is not able to classify the skin diseases and its source for causing disease, so the computer based skin disease detection is necessary to provide recommendation for non-specialized user. It is well-known that early finding and treatment of skin diseases can reduce the mortality and morbidity of patients. Digital Dermoscopy is widely considered as one of the most cost effective means to identify and classify skin-diseases. An automated medical image analysis system has three stages: (1) Image Segmentation, (2) Feature Extraction and Selection (3) Classification. Enhancement is the most important, since it affects the precision of the subsequent steps. Supervised enhancement is somewhat easy to implement by varying its parameters for variety of lesion shapes, sizes, and colors along with diverse skin types and textures. During the past few years, various contributions have been made in literature regarding the application of pattern recognition techniques for dead skin diagnosis in cell level. This paper proposes a k-means segmentation technique, Co-occurrence matrix for feature extraction and classification using if else

Index Terms: k-means clustering Segmentation, Feature Extraction using Co-occurrence matrix, Classification, median filter, Lesion Segmentation

1 INTRODUCTION

Image processing is a method to convert an image into digital form and perform some operations on it, in order to get an enhanced image or to extract some useful information from it. It is a type of signal dispensation in which input is image, like video frame or photograph and output may be image or characteristics associated with that image. Usually Image Processing system includes treating images as two dimensional signals while applying already set signal processing methods to them. It is among rapidly growing technologies today, with its applications in various aspects of a business. Image Processing forms core research area within engineering and computer science disciplines. Image processing basically includes the following three steps- Importing the image with optical scanner or by digital photography, analyzing and manipulating the image which includes data compression and image enhancement and spotting patterns that are not to human eyes like satellite photographs, output is the last stage in which result can be altered image or report that is based on image analysis.

Digital image processing allows the use of much more complex algorithms, and hence, can offer sophisticated performance at simple tasks, and implementation of methods which would be impossible by analog means. Digital image processing uses computer algorithm to perform image processing on digital images. It allows a much wider range of algorithms to be applied to the input data and can avoid problems such as the build-up of noise and signal distortion during processing. Since images are defined over two dimensions digital image processing may be modeled in the form of multidimensional systems.

In order to overcome the problems in prediction of skin cancer stages, fuzzy logic segmentation

is used then the features can be extracted using co-occurrence matrix algorithm and the classification can be done.

2 PROPOSED SYSTEM

A skin lesion is a part of the skin that has an abnormal growth or appearance compared to the skin around it. A skin lesion is an abnormal lump, bump, ulcer, sore on the skin.

The proposed method for skin lesion segmentation is shown in Figure 3.1

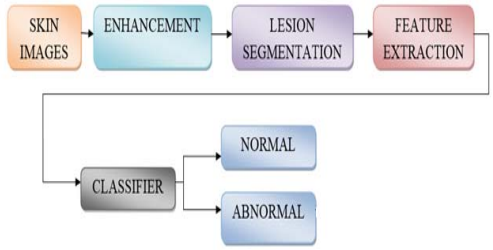

Fig. 1 Block diagram of proposed System

Classification algorithms are based on the assumption that the image in question depicts one or more features and that each of these features belongs to one of several distinct and exclusive classes.

The skin image is given as an input. For lesion Segmentation, the image enhancement is much more important technique. Then after the enhancement the segmentation can be done. From the segmented image, different features of the image can be calculated. By using classifier, the image can be classified from their calculated features.

DESCRIPTION

Image segmentation is the process of partitioning a digital image into multiple segments (sets of pixels, also known as super-pixels). The goal of segmentation is to simplify and change the representation of an image into something that is more meaningful and easier to analyze. Image segmentation is typically used to locate objects and boundaries in images. More precisely, image segmentation is the process of assigning a label to every pixel in an image such that pixels with the same label share certain characteristics. After the segmentation process, feature extraction is done for classification.

Feature extraction starts from an initial set of measured data and builds derived values intended to be informative and non-redundant, facilitating the subsequent learning and generalization steps, and in some cases leading to better human interpretations. Feature extraction is related to dimensionality reduction. When the input data to an algorithm is too large to be processed and it is suspected to be redundant then it can be transformed into a reduced set of features. Determining a subset of the initial features is called Feature Extraction. The selected features are expected to contain the relevant information from the input data, so that the desired task can be performed by using this reduced representation instead of the complete initial data. Finally the classification is done.

Image classification analyzes the numerical properties of various image features and organize data into categories. Classification algorithms typically employ two phases of processing: training and testing. In the initial training phase, characteristic properties of typical image features are isolated and, based on these, a unique description of each classification category, is created. In the subsequent testing phase, these feature-space partitions are used to classify image features. All

3 IMPLEMENTATION METHODOLOGY

Image Enhancement

Digital images are prone to various types of noise. Noise is the result of errors in the image acquisition process that result in pixel values that do not reflect the true intensities of the real scene. There are several ways that noise can be introduced into an image, depending on how the image is created. For example:

If the image is scanned from a photograph made on film, the film grain is a source of noise. Noise can also be the result of damage to the film, or be introduced by the scanner itself. If the image is acquired directly in a digital format, the mechanism for gathering the data (such as a CCD detector) can introduce noise. Electronic transmission of image data can introduce noise.

Median Filter

The median filter is a nonlinear digital filtering technique, often used to remove noise from an image or signal. Such noise reduction is a typical pre-processing step to improve the results of later processing. Median filtering is very widely used

in digital image processing because, under certain conditions, it preserves edges while removing noise, also having applications in signal processing.

The noise of the input skin image can be reduced by enhancing the image using median filter technique. The technique can be implemented using medfilt function in matlab. This function can performs median filtering process in the images given as input.

Lesion Segmentation

Automatic skin lesion segmentation in dermoscopic images is a challenging task due to the low contrast between lesion and the surrounding skin, the irregular and fuzzy lesion borders, the existence of various artifacts, and various imaging acquisition conditions. Through this Segmentation technique, the skin lesion area can be easily classified based on the color, texture and shape of the lesion. Various Segmentation process are carried out for the effective way of classification of skin lesion segmentation. The accuracy of skin cancer detection from images is directly proportional to the accuracy of the skin lesion segmentation. Then the results can be identified using feature extraction.

K-means clustering Segmentation

Many researches have been done in the area of image segmentation using clustering. There are different methods and one of the most popular methods is *k*-means clustering algorithm. *K* -means clustering algorithm is an unsupervised algorithm and it is used to segment the interest area from the background. But before applying *K* -means algorithm, first pre-processing of images can be done. Pre-processing can be done by the following steps:

1. Load the input image in RGB format
2. Convert the input image from RGB color space to L*a*b color space, by converting the image to XYZ format. RGB to XYZ conversion takes place as

$$X=0.4124*Ir+0.3576*Ig+0.1805*Ib$$

$$Y=0.2126*Ir+0.7152*Ig+0.0722*Ib$$

$$Z=0.0193*Ir+0.1192*Ig+0.9505*Ib$$

XYZ to L*a*b conversion takes place as

$$L=(116*Yr)-16$$
$$A=500*(Xr-Yr)$$
$$B=200*(Yr-Zr)$$

Where,

$$Xr=X/95.047$$
$$Yr=Y/100$$
$$Zr=Z/108.883$$

3. Form a new grayscale image by using $Ig=a+b-L$
4. Apply filter to remove the effects of hair and skin scale during segmentation
5. Improve the contrast of the smoothed image to distinguish the boundaries of the lesion.

The preprocessed image obtained is subjected to a clustering technique for lesion segmentation. Each pixel present in the image is initially assigned to a cluster such that cluster sum of squares is minimized and finally cluster pixels are updated on the basis of new mean obtained for each cluster.

The output of K-means clustering algorithm result an image which has pixels divided into different clusters.

Feature Extraction using Co-occurrence matrix

A co-occurrence matrix or co-occurrence distribution is a matrix that is defined over an image to be the distribution of co-occurring pixel values (grayscale values, or colors) at a given offset.

The offset, is a position operator that can be applied to any pixel in the image. An image with different pixel values will produce a co-occurrence matrix, for the given offset. The value of the co-occurrence matrix gives the number of times in the image that pixel values occur in the relation given by the offset. For an image with different pixel values, the co-occurrence matrix **C** is defined over an image **I**, parameterized by an offset. Co-occurrence matrix is also parameterized in terms of distance, angle and an offset. Any matrix or pair of matrices can be used to generate a co-occurrence matrix, though their most common application has been in measuring texture in images. Co-occurrence matrix can be used to extract the features like Contrast, Correlation, Energy,

Homogeneity, Mean, Standard Deviation, Entropy, RMS, Variance, Smoothness, Kurtosis, Skewness and IDM.

Classification

Classification can be done by round off the values extracted from the Co-occurrence matrix. If the values reaches to 1, then it gives the output as stage 1. If the round off value is 2, then it produces the output as 2. If the value becomes 3 then it generates output as 3.

```
V1=xlsread('features.xlsx')
```

The above code will helps to read the feature values stored in the excel sheet.

```
P1 = V1';

T = [1 1 1 2 2 2 3 3 3 3];

net = newrb(P1,T);

Y = sim(net,P1);

Y=round(Y)
```

This code will help to round off the extracted values from the feature extraction. Then by using if else condition the classification can be done.

4 RESULT AND DISCUSSION

The skin image is given as input. Fuzzy k-means clustering algorithm is applied on the image. The segmentation helps in extracting the lesion region, the output of fuzzy k-means clustering is shown in Fig. 2

Fig. 2 Input image for Segmentation

The output after segmentation process is shown in Fig. 3

Fig. 3 Segmentation result using k-means

The image after segmentation can be processed for feature extraction using co-occurrence matrix. The following figure shows the features extracted from various images as input.

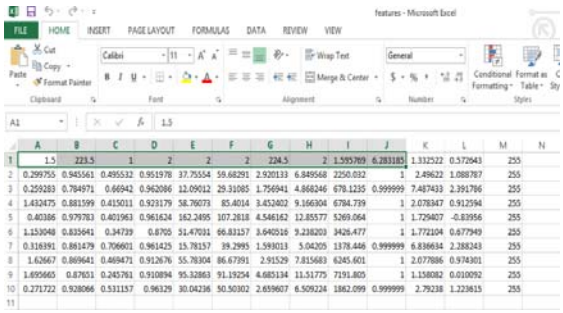

Fig. 4 feature extracted from co-occurrence matrix

Then by using classifier, the image can be classified based on the features extracted from co-occurrence matrix. The following figure shows the results of classification.

Fig. 5 classification of input image

The different stages shown in the above image represents the different stages of input skin images. If we give an input skin image, then the corresponding output is also displayed in the same window.

5 CONCLUSION AND FUTURE WORK

From the overall proposed solution, the Fuzzy k-means algorithm is much easier for image Segmentation. Many features can be extracted using co-occurrence matrix and can classified easily. This Segmentation process not only helps to identify the skin disease but also used to predict various types of diseases like thyroid, Melanoma, Bullae, Seborrheic keratosis, Shingles, squamous cell, chickenpox etc.

The detection of various skin diseases by analyzing the size, color and structure can help doctors to have a second option for predicting the stages of skin disease.

From this idea the affected area of the skin is analyzed and can predict the stages of diseases.

In order to enhance the project, the segmentation technique can be changed. This helps to determine the accurate results in segmentation.

6 ACKNOWLEDGEMENT

We place on record and warmly acknowledge the continuous encouragement, invaluable supervision, timely suggestions and inspired guidance offered by our guide Dr.M.Nirmala, Assistant Professor, Department of Electronics and Communication Engineering, at K.S.Rangasamy College of Technology, Tiruchengode in bringing this project to a successful completion.

REFERENCE

[1] Ashnil Kumar, Jinman Kim, David Lyndon,Michael Fulham,Dagan Feng (2016) 'An Ensemble of Fine-Tuned Convolutional Neural Networks for Medical Image Classsification',journal of Biomedical and Health Informatics,Vol.26 No.6, pp.2168-2194.

[2] Assaf Hoogi, Arjun Subramaniam, Rishi Veerapaneni, Daniel L.Rubin (2016) 'Adaptive Estimation of Active Contour Parameters Using Convolutional Neural Networks and Texture Analysis', Transactions on Medical Imaging,Vol.13 No.1, pp.0278-0062.

[3] Gui-Song Xia, Gang Liu, Xiang Bai, Liangpei Zhang (2017) 'Texture Characterization Using Shape Co-occurrence Patterns',Transactions on Image Processing,Vol.27 No.8,pp.1057-7149.

[4] Haifeng Sima, Ping Guo, Youfeng Zou, Zhiheng Wang, mingliang Xu (2017) .

'Bottom-Up Merging Segmentation for Color Images With Complex Areas',Transactions on Systems, man and Cybernetics, Vol.35 No.2, pp.2168-2216.

[5] Jiangye Yuan, DeLiang Wang, Anil M.Cheriyadat (2015) 'Factorization-Based Texture Segmentation', Transaction on Image Processing, Vol.18 No.8, pp.1057-7149.

[6] Jie Chen, Guoying Zhao, Mikko Salo, Esa Rahtu, Matti Pietikainen (2012) 'Automatic Dynamic Texture Segmentation Using Local Descriptors and Optical Flow', Transactions on Image Proceesing, Vol.22 No.1, pp.1057-7149.

[7] Khan, Hayat, Bennamoun,Togneri,Sohel (2016) 'A Discriminative Representation of Convolutional Features for Indoor Scene Recognition', Transactions on Image Processing, Vol.30 No.4, pp.1057-7149.

[8] Lie Bi, Jinman Kim, Euijoon Ahn, Ashnil Kumar, Michael Fulham (2016) 'Dermoscopic Image Segmentation via Multi-Stage Fully Convolutional Networks',Transactions on Biomedical Engineering, Vol.11 No.3, pp.0018-9294.

[9] Saddam Hussain, Syed Muhammad Anwar, Muhammad Majid (2017) 'Brain Tumor Segmentation using Cascaded Deep Convolutional Neural Network', Vol.27 No.8, pp.1998-2001.

[10] Xiaofang Wang, Yuxing Tang, Simon Masnou, Liming Chen (2015) 'A Global/Local Affinity Graph for Image Segmentation',Transaction on Image Processing,Vol.34 No.12, pp.1057-7149