International Journal of Contemporary Research in Computer Science and Technology (IJCRCST) ISSN: 2395-5325
Volume 3, Special Issue 3 (September '2017)
Proceeding of - International Conference on Recent Trends in Computational Life Science and Information Technology,
Conference held at Tiruppur Kumaran College for Women,Tirupur, Tamilnadu,India.

CLASSIFICATION OF DERMOSCOPIC SKIN CANCER IMAGES AND DIAGNOSIS USING BAYES NET & BEST FIRST SEARCH & EXPERT SYSTEM

K.Devipriya,
Assistant Professor,
Department of Computer Science
Tiruppur Kumaran College for Women,
Tirupur, Tamilnadu, India.

L.Subathra Devi,
Assistant Professor,
Department of Computer Science
Tiruppur Kumaran College for Women,
Tirupur, Tamilnadu,India.

Abstract: Early detection of skin cancer has the potential to reduce mortality and morbidity. In this world near about 1/7th of total world population suffer from some sort of skin disorder. This paper presents two hybrid techniques for the classification of the skin images to predict it if exists. The proposed hybrid techniques consist 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. There are six different categories of skin diseases which shares somewhat same features. So for the classification of these diseases Bayes net a Bayesian technique along with feature selection has been used in this study. 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. Using Expert System it exhibits the diagnosis of skin disease identification accuracy of 85% for Eczema, 95% for Impetigo and 85% for Melanoma.

Keywords: Association rules, Clustering, Classification Cancer Survey , Data Mining , Medical Image Processing, Bayes Net, Expert System.

I. INTRODUCTION

Skin Cancer:

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). Non melanoma skin cancer (MMSC) is the 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. 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, heart diseases, brain tumors etc. Recent work 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 can be used to classify the normal/abnormal images. Therefore image processing becomes 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., Feedforward-Backpropagation Neural Network (FPBANN) used as a classifier. Also unsupervised classification techniques such as k-NN algorithm is 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.

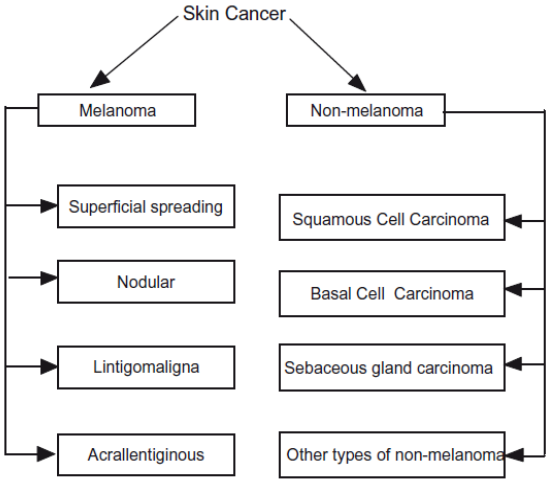


Figure1.1: Types of skin cancer.

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

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

Luminance = 0:299*Red+0:587*Green+0:114*Blue

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

ii) 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.

Image processing unit

Image processing is the main part of the design process in our expert system. Initially it is required to identify the region of skin infected with the disease and in order to do that the image processing part of this process should be carried out. The image of the skin disease should be captured and saved to eliminate noises in the form of hairs, bubbles etc. with the help of average filtering and Gaussian smoothing process. Algorithms are used in the image segmentation process to remove the background of the image. It means we separate the region of the disease. Then we will be done with the image segmentation of the image.

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. and 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.

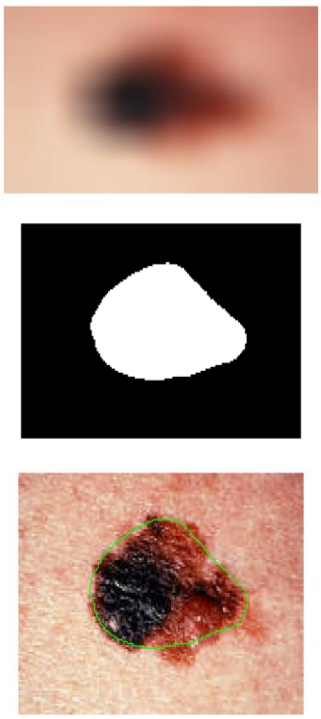


Figure1.2: Superficial spreading image segmentation

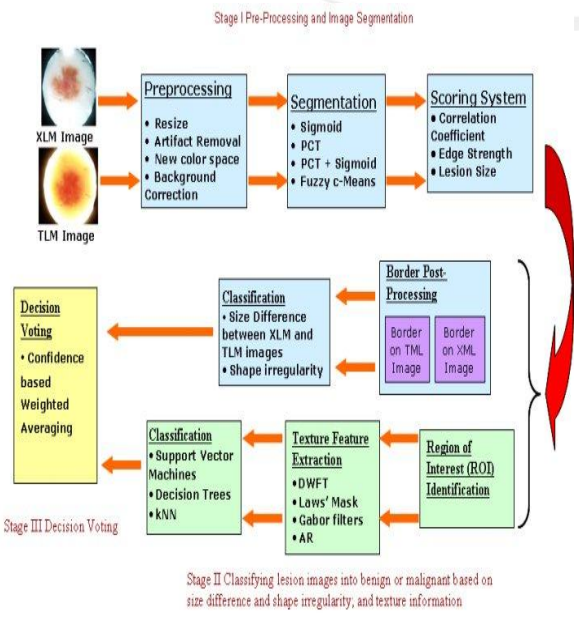


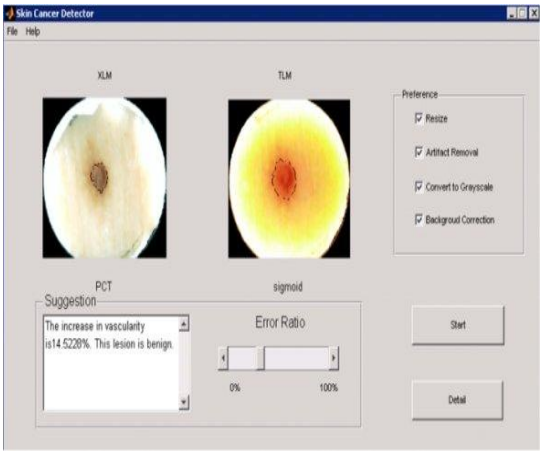
Figure1.3: Pre processing and image segmentation

Skin Lesion Texture Analysis

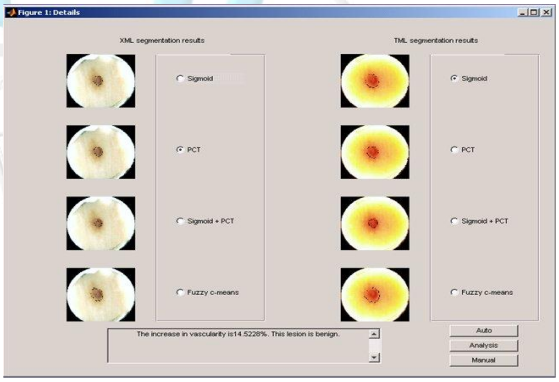
GUI Design

Following figures show GUI design for pre-processing and segmentation steps of the system. The system will be used by clinicians and we expect the GUI will change to fit their taste.

Following Figure shows available pre-processing options and segmentation results.



The figure below also provide user flexibility to choose Different segmentation algorithms manually.



More detailed information, such as result analysis for texture feature extraction and selection.

Data mining unit

The features obtained from the image processing unit will be examined and the disease will be identified and displayed by this unit. Then some questions such as age, gender, affected areas in the body, how long it was infected etc. will be asked from the user. It is called questionnaire inputs of the system and will be really helpful to get details about the diseases. Then based on those images extracted features and input for the questionnaire used for creating classification model. Here as the classification model, we have used five different data mining classification algorithms (AdaBoost, BayesNet, J48, MLP and NaiveBayes)

and based on the result best classifier is selected. Finally using this data mining classification model, our expert system will the diagnosed skin disease and predicted skin disease and medical treatment suggestion will be displayed for the user interface used to create the classification model for predict the skin disease.

i) 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, details of both will be given below.

A. k-Nearest neighbors (K-NN):

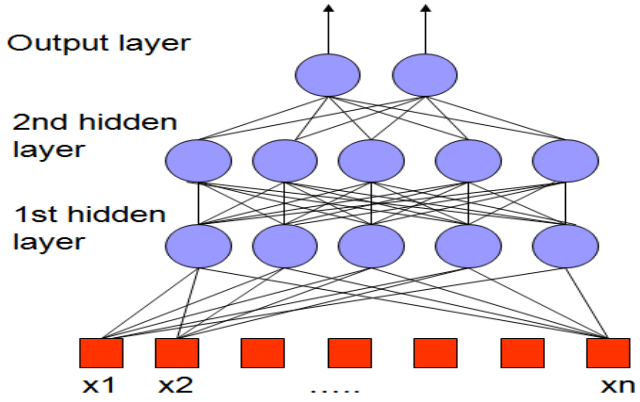
One of the most straightforward instance-based learning algorithms is the nearest neighbor algorithm's K-NN is based on the principle that the instances within a dataset will generally exist in close proximity to other instances that have similar properties. If the instances are tagged with a classification 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 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:

- Compute the distance between t and each instance in I.
- Sort the distances in increasing numerical order and pick
- the first k elements.
- 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.

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. The feed forward neural network was the first and arguably simplest type of artificial neural network devised. Classifier required in this study had three layers, 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. The most frequently used training algorithm in classification problems is the back-propagation (BP) algorithm, which is used in this paper too.

Fig.2.1 Feed forward neural network.



The details of back-propagation (BP) algorithm are well documented in the literature . 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. The performance is measured by mean square error(MSE):

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

here yt is the target value, yo is the actual output, and m is the number of training data, shows ANN Back- propagation algorithm.

III.CONCLUSION

When considering about this application's accuracy, it can over more than 85% accuracy for user. We used various classifiers to calculate and evaluate the accuracy level of our system. Multi-Layer Perceptron (MLP) and J48 are main classifiers we used. And reason to so many users would like to work with our application is, these are normally conclude our output which considering both result which are obtains from image processing method and questionnaire method. It means, our system will consist with the several questions which prepared by system from the user. And system get help answer which given for above mentioned questions to diagnosis skin disease. In this system while administrator managing the information of skin disease, symptoms, medical treatment suggestions and it prepared statement to view the summery about Skin Disease. And also this has high reliability and high performance. Both image based technique and questionnaire technique help to increase reliability and performance of the system. 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.

IV.REFERENCES

[1]. Munirah M. Yusof, Ruhaya A. Aziz, and Chew S. Fei - The Development of Online Children Skin Diseases Diagnosis System.
[2]. Florence Tushabe, Ernest Mwebaze and Fred N. Kiwanuka - An image based diagnosis of virus and bacterial skin infections.
[3]. Anal Kumar MittraandDr. Ranjan Parekh - Automated Detection of Skin Diseases Using Texture Features.
[4]. https://play.google.com/store/apps/details?id=com.rubytribe.skinvision
[5]. P. K. Buxton (1991) - ABC of Dermatology - Jolly & Barber Ltd, Great Britain, pp. 1-30.
[6]. K. Darlington (2000) - The Essence of Expert Systems, Prentice Hall.
[7]. A. Rubin (2007) - ―Design of an expert system and its application to dermatopathology‖ - Histopathology, vol. 21, no. 3, pp. 269–274.
[8]. Al. Abadi , N. K.; Dahir , N. S.; Alkareem, Z. A. (2008): Skin texture recognition using neural network, in Proceedings of the International Arab Conference on Information Technology, Tunisia, December 16-18, pp.1-4.
[9]. Rubegni, P. et al. (2002): Automated Diagnosis on Pigmented Skin Lesions - International Journal on Cancer, 101, pp. 576-580.
[10]. Image recognition with neural networks how to - http://neuroph.sourceforge.net/image_recognition.html

