k-NN CLASSIFIER FOR SKIN CANCER CLASSIFICATION

Dr. M. Vimaladevi[1]
Department of Computer Science and Engineering
Hindusthan Institute of Technology, Coimbatore
Email : vimal.vcet@gmail.com

B.Kiruba[2]
Department of Computer Science and Engineering
Hindusthan Institute of Technology, Coimbatore
Email :kirubamugesh@gmail.com

A.Nivethitha[3]
Department of Computer Science and Engineering
Hindusthan Institute of Technology, Coimbatore
Email : nivethi92@gmail.com

Abstract: Skin cancer is now becoming a challenging issue to identify the exact location of affection on the skin tone. A novel hierarchical k-Nearest Neighbors (k-NN) classifier is more useful to find the affected level of skin and the type of cancer disease. The k-NN classifier is comparatively simple, quick and effective. The structure which is of hierarchical level mainly decomposes classification into a set of easy issues. At the first level of the classification, feature selection is done. The most relevant feature subsets such as color and texture features are extracted from skin lesions and passed to each node of the classification level. The efficiency of the proposed scheme is typically larger in discriminating cancer and pre-malignant cells from the benign cells and it reaches overall classification accuracy over a few common classes of skin lesions that include two non-melanoma cancer classifications. An automatic skin cancer classification system is developed and the relationships of skin image across different type of training networks are studied with different types of image preprocessing. To enhance the classification results, the image properties of the normal skin is eliminated from the skin affected area and the cancer cell is presented in the image..

Keywords: Skin Cancer, k-NN Classifier, Lesions classification, Hue, Saturation

I. INTRODUCTION

Skin cancer may appear as tumors in the form of malignant or benign. Benign Melanoma is simply appearance of some pores and holes on skin such as mole, etc. Malignant melanoma is the look of sore like structures that create bleeding or the flow of blood. Malignant is the deadliest form of all skin affecting diseases. It gets arose from growth of cells in pigmented skin tone estimation and named after the cell from which it arises. Melanoma diagnosis is tedious and needs sampling tests. Melanoma can be spread out to all organs of our human body through the total system. The main issue to be considered covering with melanoma is that, the first trouble of the disease can to the future ones. Sampling often leads to the inflammation or even drastic surpass of lesion. Computer based diagnosis can enrich the speed of skin cancer detection which works according to the disease symptoms.

The similarities among lesions cause the diagnosis of malignant cells a tedious one. But, there are some similar symptoms of skin cancer, such as: Border irregularity, Color variation and Diameter. Asymmetry is part of the tumor that shouldn't match the other part. Color intensity change in the region is irregular. It also intimates about the Computer based Skin cancer identification. Automatic early detection system is a classification system which distinguishes Malignant Melanoma from other skin diseases. Artificial Intelligence is used for classification purpose.

Skin images for the classification process will be converted to the digital format. The image may contain noises and that could be eliminated by pre-processing techniques. To preserve the edges, few steps are too carried out at last. To separate the cancerous region from non-cancerous areas in skin, segmentation is done. There are similar features for the cancerous images. The features are extracted using Two Dimensional Wavelet Transformation method which is implemented using MATLAB software.

II. LITERATURE REVIEW

Gordon .A.D., (1987) summarized the relationships within a set of objects by a set of hierarchically-nested classes of similar objects, representable by a rooted tree diagram. Methods for obtaining tree diagrams, comments on the selection of appropriate methods of analysis and the validation of classifications, distributions of different types of tree, and consensus trees were discussed.

Jain, A.K et al.,(2000) discussed various pattern recognition applications with pattern recognition models. The patterns from the applications will be extracted by feature extraction methods such as Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA). The classifier like k-NN, Bayes and Decision tree is used to classify the cancer data.

Aslandogan, Y et.al. (2004), proposed a method that combines three methods for classification of breast and skin lesions and the classification results are used to find the solution to the uncertainty problems. The experimental results proved that the evidence combination method outperforms well than the individual methods.

In the method proposed by Ho Tak Lau , et. al (2009), image of the skin affected by the cancer were feed into the neural networks in the digital form. The image is enhanced by using various image processing techniques. The classification process is done on the training data and tested by the testing data.

Maglogiannis (2009) defines the visual features used for skin lesion classification, and the methods for defining them. The features are extracted through digital image processing methods such as, segmentation, border detection, and color and texture processing.

The features are categorized into documents by consulting the relevant training documents and Duwairi (2011) et.al. and by applying feature selection methods, the important terms only remains for classification. Modified k-NN classifier was used for classification.

Salah et.al. (2011) proposed image processing techniques, a neural network system (NN) and a fuzzy inference system for detection of different types of skin cancer. The features of human skin such as pigmentation or color variation was used in the study. Color and texture features are extracted from the images and hierarchical structure based on the k-NN classification method.

The features are extracted by decomposing images into different frequency sub-bands using wavelet transform. The output of Discrete Wavelet Transform becomes input to the Classification System which classify whether the input image is cancerous or noncancerous, Yogendra Kumar Jain et. Al (2012). The classification system is based on the application of Probabilistic Neural Network and Clustering Classifier.

Mahmoud Elgamal , (2013) defined hybrid techniques to classify the images affected by the skin cancer. He used discrete wavelet transformation to obtain the features related with images and principle component analysis to reduce the dimensionality of the images. The images were classified using feed forward back propagation neural networks and k-NN classifiers.

Idris Nayaz Ahmed, (2014) proposed a work comprised of preprocessing, segmentation, feature extraction and classification. Weiner Filter was used for preprocessing, Distance Regularized Level Set method was used for segmentation and Support Vector Machine was used for classification. The classification efficiency was measured by accuracy, sensitivity and specificity.

Deepti Sharma et. al.(2016) used Statistical region merging algorithm to enhance the properties of skin cancer images and proposed two neural networks for classification skin cancer images namely Back-propagation neural network and Auto-associative neural network.

Ebtihal Almansour et. al(2016) proposed an algorithm for classification of Dermoscopic Skin Cancer Images. In their work, they used two types of texture features for classification namely, Local Binary Pattern on different scales and Gray Level Co-Occurrence Matrix. The extracted features were classified using Support Vector Machines.

III. METHODOLOGY

A.EXISTING SYSTEM

PARALLEL HIERARCHICAL:

The structure is fixed based on priority by grouping our image classes into two main classifications. The initialized group contains various lesion classes. The second group, Group 2 contains some other lesions classes. The two groups were constructed by clustering each class contains images which were similar at this first split. AK is treated as a pre-malignant condition that can give rise to disease and sometimes can be visually alike to early trivial cancerous categories. In the second group both benign forms of skin lesions having a tremendous appearance. The class grouping leads to the cascading structure. This construction makes a common interval between categories at the upper degree while finer decisions are made at a lower one. As a result, this scheme decomposes the original problem into 3 sub-problems.

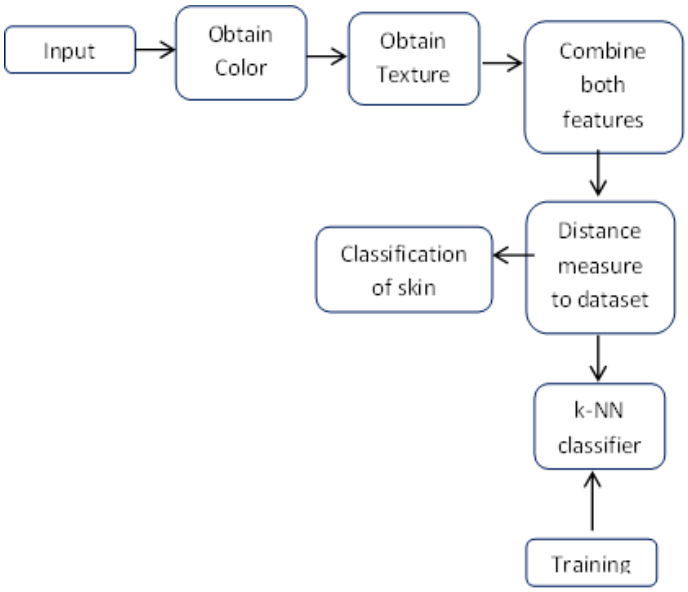

Fig 1: Overall Block Diagram of the proposed technique

HUMAN IDENTIFICATION (MANUAL):

The manual method is implemented over here to diagnose and to track the exact location of the cancerous cell that is available on the skin. Skin may be affected by any degree level of hazardous values. This can be implemented to track the exact location of the cancerous cells. The cancerous cells can be divided off cells.

B. ***PROPOSED SYSTEM***

The proposed method is found with the color features extraction. The primary color or RGB color space is taken for this method because primary colors can only give better result in 5 color spacing factors such as HSV, CIE LAB, YCbCr and etc. Then finding the texture features that were extracted from generalized co-occurrence matrix (GCM) is done. Here extracting more texture features such as energy, contrast and entropy. To combine color and texture feature to calculate distance between input image and database image using distance calculation factors. Implementations employ K-NN classifier for classification purpose. Here we implement both the input sets and training set. The input image is testing phase and remaining all image looking at as a training images. At last we classify the skin lesion value as three types.

Pre-processing is to remove the noise in the image. For smoothing an image, median filter is used. Median filtering is a type of filter in image processing. Median filtering is used for reduce the influence of small structures and isolated pixels like small air bubbles. Post-processing is done to develop the shape and edges of image. In addition, contrast enhancement can sharpen the image border and improve the accuracy for segmentation.

HSV:

Hue, Saturation and Value deal with the identification of the changes in the color region and generation. Hue value shows the sharpness of the image whereas Saturation value tells us about the change in appearance of the color imparters in the image. Hue, Saturation and the Identification features identify the various saturated values output the specified image as a signal. These values provide the drastic change between the normal input image and the change in the saturated image.

YCbCr generation from RGB:

Y, Cb, Cr are the generation versions of the primary colors Red, Green and Blue. These colors form the primary and the ultimate generation of all the colors as shown below.

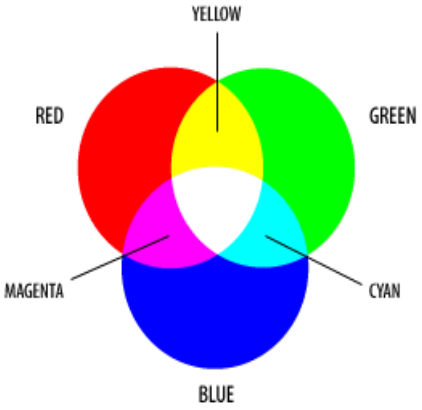

Fig 2: Classification of RGB Analysis

K-NN Classifier:

NN is nothing but the form of Nearest Neighbors. Skin cancer classification can be done with the identification of his algorithm such that the test samples and the training samples are loaded in the databases. Samples are classified by calculating the nearest distance to the training case. Its part then determines the classification of the sample. K-NN classifier extends this idea by taking the k nearest points and declaring the sign of the majority. It is unique to select k values.

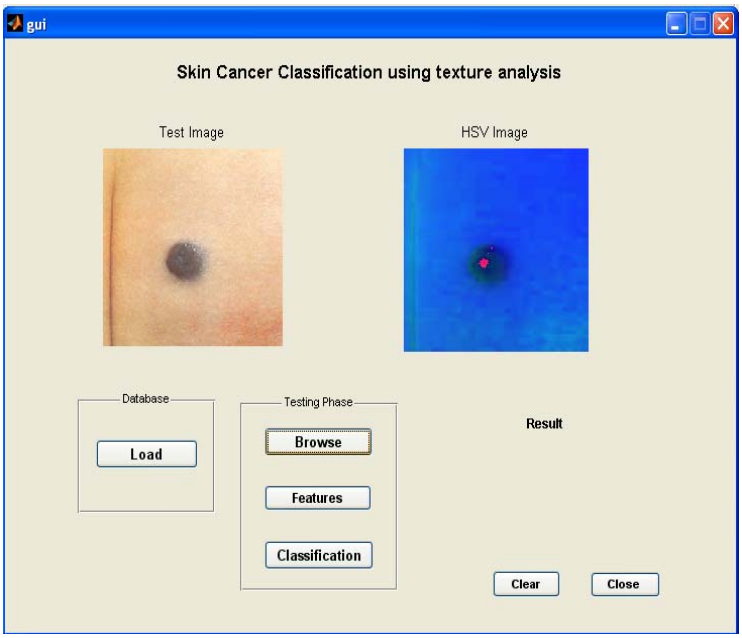

Fig 3: Skin Feature Extraction Process

Values which are greater in the value of k help in reducing the effects of noisy levels in the pixels rate within the training data set and opt the value of k is often performed through cross-validation. There are several techniques available for developing the accuracy and speed of a nearest neighbor classification. Identification for this problem is to select a subset of the training data such that classification by the 1-NN rule using the values of many subsets thus approximates the Bayes classifier which is having the nearest neighbor values.

TEXTURE CLASSIFICATION:

An image texture is a composite of metrics and the values are calculated in image processing designed to quantify the perceived texture of an image. Image Texture gives us information about the spatial arrangement of color or intensities in an image or selected region of an image. Image textures can be artificially created or found in natural scenes captured in an image. Image textures are one way that can be used to help in Segmentation or classification of images.

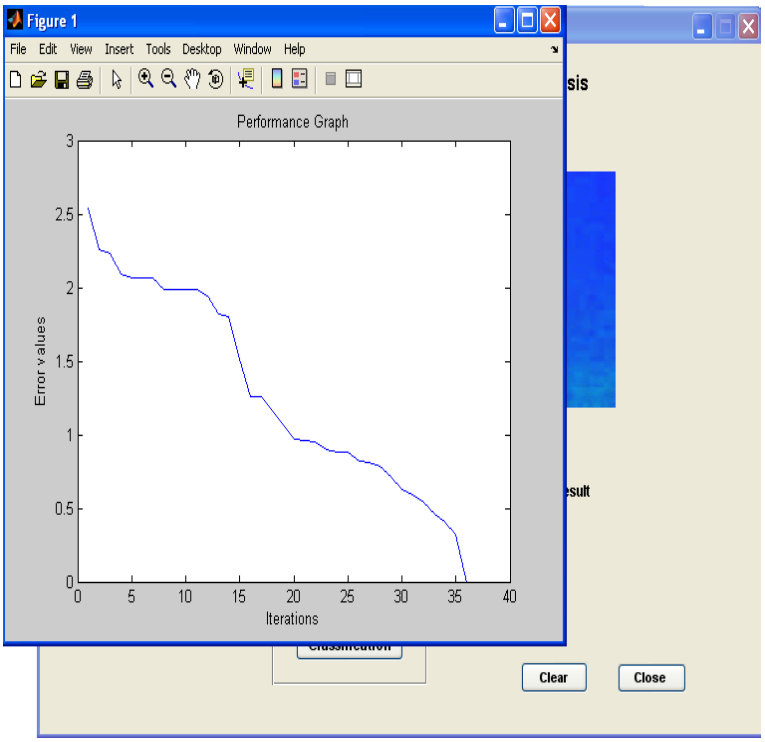

Fig 4: Texture Feature Extraction Performance Validation

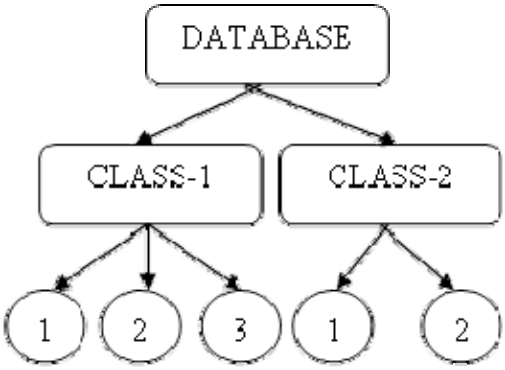

Fig 5: Flowchart of Class Separation

ADVANTAGES OF THE PROPOSED SYSTEM:

1. Overall setup and this workflow tells us about the identification of the level of skin cancer level and this idea tells us about the classified name also
2. Overlapping of image pixels are not occurred here
3. Highly robust in nature
4. Accuracy in finding out the classification is greater.

IV. RESULTS

Overall results show that the classification of cancerous cells using K-NN classifier achieves a greater level of accuracy in identification. HSV conversion paves the way in identifying the color contrast level and HSV identifies the exact location of cancerous level cells. Next to that, the level of color separation occurs. Color separation gives the output the given input image and the classification of the cancer level paves way in converting the color based on the combination of Y, Cb, Cr conversions.

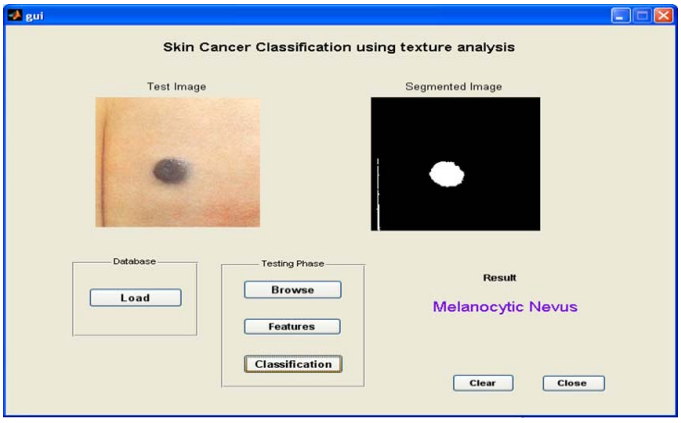

Fig 6: Overall Output of our Proposed System

V. CONCLUSIONS

The algorithm based on a hierarchical K-NN classifier tree for the classification of a few classes of skin cancer cells are performed here. The overall performance is not yet at more than the level for differential diagnosis of moles which is opponent to the skin cancer cells. Somehow, the task makes 2 key contributions:

1) The results are based only on normal color inputs, not similar to the other method, which requires well-equipped components.
2) Overall multi-class classification technique reaches the majority of skin cancer types. This deviates from most other techniques that understand only two or three class examples of the problem.

REFERENCES

[1] Maglogiannis, I & Doukas, C.N, "Overview of advanced computer vision systems for skin lesions characterization", IEEE Transactions on Information Technology in Biomedicine, 13(5) : 721–733, 2009.
[2] Jain, AK, Duin, RPW, & Mao, J, "Statistical pattern recognition: A review", IEEE Trans. on PAMI, 22(1) : 4–37, 2000..
[3] Gordon, AD ,"A review of hierarchical classification", Journal of the Royal Statistical Society. Series A (General), 150(2) : 119–137, 1987.
[4] Duwairi, R & Al-Zubaidi, R ,"A hierarchical K-NN classifier for textual data", The International Arab Journal of Information Technology, 8(3) : 251–259, 2011.
[5] Salah, B, Alshraideh, M, Beidas, R & Hayajneh, F, "Skin cancer recognition by using a neuro-fuzzy system", Cancer Informatics, 10: 1–11, 2011.
[6] Aslandogan, YA, & Mahajani, GA , "Evidence combination in medical data mining", ITCC 2: 465 – 469, 2004.
[7] Mahmoud Elgamal, "Automatic Skin Cancer Images Classification", International Journal of Advanced Computer Science and Applications, 4(3), 2013.
[8] Ho Tak Lau , Al-Jumaily , "Automatically Early Detection of Skin Cancer: Study Based on Nueral Netwok Classification" , Soft Computing and Pattern Recognition , 2009.
[9] Yogendra Kumar Jain & Megha Jain , "Comparison between Different Classification Methods with Application to Skin Cancer", International Journal of Computer Applications, 53(11), 2012.
[10] Idris Nayaz Ahmed & Chaya P, "Segmentation and Classification of Skin Cancer Images", International Journal of Advanced Research in Computer Science and Software Engineering, vol. 4, Issue 5, 2014, pp. 1349-1353, 2014.
[11] Deepti Sharma, Swati Srivastava , "Automatically Detection of Skin Cancer by Classification of Neural Network", International Journal of Engineering and Technical Research, 4(1) : 15-19, 2016.
[12] Ebtihal Almansour and M. Arfan Jaffar, "Classification of Dermoscopic Skin Cancer Images Using Color and Hybrid Texture Features", International Journal of Computer Science and Network Security, 16(4) :135-139, 2016.

BIOGRAPHIES

M. Vimaladevi received her Bachelor's degree in Computer Science from Madras University in the year 1998 and Master's Degree Computer Science and Engineering from Anna University in the year 2009 and Ph.D in the field of Data Mining at Anna University 2016. She is having around 10 Years of experience and presently working as an Assistant Professor at Hindusthan Institute of Technology, Coimbatore. She is a member of various technical committees and active reviewer in various journals. Her area of interest includes Data Mining, Knowledge Engineering, Web Mining, Image Processing and e-learning.

B.Kiruba received her Bachelor's degree in Computer Science and Engineering from Anna University in the year 2014 and Master's Degree in Computer Science and Engineering from Anna University in the year 2016. She is having around 1 year of experience and presently working as an Assistant Professor at Hindusthan Institute of Technology, Coimbatore. Her area of interest includes Digital Image Processing , Web mining and. software engineering with special interesting driving and monitoring program executions to find bugs, using various software testing techniques.

A.Nivethitha obtained her B.E degree in CSE from Anna University in the year 2014 and PG degree in M.E CSE from Anna University in the year 2016. She is having around 1 year of experience and presently working as an Assistant Professor at Hindusthan Institute of Technology, Coimbatore. she has presented and published 4 technical papers in national conferences and one technical paper in international journal. Her area of interest includes Wireless Sensor Networks, Cloud computing and Database security