Pragmatic System For Detection Of Skin Lesion Using Feature Extraction Techniques

[1]Suruchi Padhy, [2]Prof Sandeep Kumar

[1]Student, Mtech in CSE, [2]Associate Professor

[1]Computer Science and Engineering,

[1]Cambridge Institute of Technology, Bangalore, India.

Abstract: This paper describes the design and implementation of pragmatic system for detection of skin lesion using feature extraction techniques. The skin properties like skin dryness, fungus and allergic symptoms of skin layer may lead to starting symptoms of malignant melanoma skin cancer. The correct identification of skin spots based on certain features is the key steps in detecting the skin cancer disease in advance. The affected skin texture profile correlation with malignant melanoma skin cancer is discussed in the proposed work. Detection of skin cancer in earlier stages can be a lifesaving process. The detection of skin lesion includes four important stages namely Pre-processing, Segmentation, Feature Extraction and Classification. Detection can help in curing the cancer and hence detection plays a very vital role. In this paper, this system proposes two methods of preprocessing and features were extracted using GLCM. The KNN classifier, applied on extraction features which will classify image as melanoma and benign.

Index Terms **–** Skin lesion, Melanoma, Benign ,Gaussian filter, RGB channel, preprocessing, segmentation, thresholding, edge detection, feature extraction, K-Nearest Neighbor (KNN); Gray Level Co-occurrence Matrix (GLCM).

I. Introduction

The body's outer covering, which protects against heat and light, injury, and infection is called skin. It regulates body temperature and stores water, fat, and vitamin D. Skin has three layers. The epidermis, the outermost layer of skin, provides a waterproof barrier and creates our skin tone. The dermis, beneath the epidermis, contains tough connective tissue, hair follicles, and sweat glands. The deeper subcutaneous tissue (hypodermis) is made of fat and connective tissue. The skin's color is created by special cells called melanocytes, which produce the pigment melanin. Melanocytes are located in the epidermis. A skin lesion is an abnormal lump, bump, and ulcer, sore or colored area on the skin. Common skin lesion include moles and actinic keratosis. Benign skin lesions are non-cancerous skin growths that may be pointed out by the patient or discovered during routine skin examinations. Most skin lesions are benign though some, such as actinic keratosis and certain moles, can be precursor to skin cancer or already a skin cancer.

Cancer is a disease affecting the skin. Skin cancer may appear as malignant or benign form. Malignant melanoma is the appearance of sores that cause bleeding. Melanoma is the deadliest form of all skin cancers. Melanoma, also known as malignant melanoma. It is named after the cell from which it presumably arises, the melanocyte. If diagnosed at the right time, this disease is curable [11]. Melanoma diagnosis is difficult and needs sampling and laboratory tests. Melanoma can spread out to all parts of the body through lymphatic system or blood. In women, they most commonly occur on the legs, while in men they are most common on the back. Sometimes they develop from a mole with changes such as an increase in size, irregular edges, and change in color, itchiness, or skin breakdowns. The primary cause of melanoma is (UV) exposure in those with low levels of skin pigment.

Moles are small skin marks caused by pigment producing cells in the skin. They can be flat or raised, smooth or rough. They may occur anywhere on the body. Most are dark brown or black, but some flesh color red or yellow. They can change in appearance over time. Some can develop into cancer [12]. Benign moles are generally smaller than the size of a pencil eraser. Actinic keratosis occurs when the skin has sustained too much sun exposure and damage, they appear as crusty bumps. Actinic keratosis may be flesh-colored, brown, pink or red. Affected areas may inflamed, itch or bleed. Actinic keratosis can develop into skin cancer. For that reason, most dermatologists want to treat these lesions. The key features of skin lesions are, the type of lesion, secondary changes to the surface of the lesion, and the arrangement and distribution of the lesions. These are based on diameter of the lesion, relationship of the lesion to the surface of the skin.

If melanoma is recognized and treated early, it is almost always curable, but it is not, the cancer can advance and spread to other parts of the body, where it becomes hard to treat and can be fatal. While it is not the most common of the skin cancers, it causes the most deaths. An estimated 7,230 people will die of melanoma in 2019. Of those, 4,740 will be men and 2,490 will be women. However, early detection of skin cancer is an expensive affair. As skin lesions look quite similar to each other, it is difficult to determine whether a lesion is benign or malignant. Extensive analysis needs to be performed to identify the category of the lesion. Traditionally, an image using a special device, known as a dermatoscope, is taken to study the lesion closely. Unfortunately, dermatoscopes are expensive and not widely available with dermatologists [13].

In this paper, we present an approach to pre-process and segment lesions, extract features from the segmented lesions and using KNN which would then classify the lesions into their respective categories. We consider three variants each of benign and malignant lesions. The benign lesion category comprises of Melanocytic Nevi, Seborrheic Keratoses and Acrochordon, whereas the malignant lesion category comprises of Melanoma, Basal Cell Carcinoma (BCC) and Squamous Cell Carcinoma (SCC).

There are some unique symptoms of skin cancer, such as: Asymmetry, Border irregularity, Color variation and Diameter. Those are popularly known as ABCD parameters. ABCD parameters. Asymmetry, Border irregularity, Color, Diameter. Asymmetry is one half of the tumor does not match the other half. Border irregularity is the unevenness of images. Color intensity change in the lesion region is irregular. Malignant melanoma is having a diameter greater than 6 mm [14]. 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.

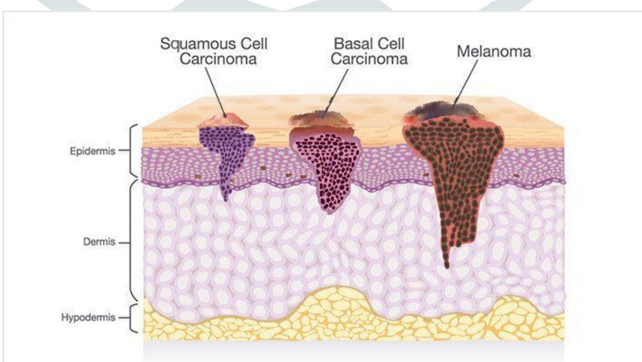


Fig 1.Types of skin cancer

(a) (b) (c) (d)

Fig 2 a) melanoma b) Basal cell carcinoma c) Squamous cell carcinoma b) Benign

II. Related work

This paper introduces a new computer-aided method for classifying both melanocytic skin lesion and non-melanocyte skin lesion. This method distinguishes among melanoma, nevi, BCC, and SKs and calculate four texture features. Here introduces two classification method SVM and KNN. This method is tested on 42 dermoscopic images: 16 melanoma, 9 nevi,8 BCC and 9 SK. The SVM classifier outperformed the KNN classifier, achieving an accuracy rate of 75.95% [1].This paper tend to two distinct frameworks for identification of skin growth in dermoscopy pictures. The primary frame utilizes worldwide strategies and the second frame work utilizes neighborhood highlights and the classifier. RBF neural network derives classification. This paper proposed new learning method using K mean clustering [2].In this paper, a novel approach for automatic segmentation and classification of skin lesions is proposed. For segmentation region growing method is applied by automatic initialization of seed points. The extracted lesion areas are represented by color and texture features. The performance of the system is tested on 726 samples from 141 images consisting of 5 different classes of diseases. The result are very promising with 61% of F-measures using SVM and KNN [3].In this paper multilevel 2-D wavelet decomposition is used for feature extraction technique. These features are given to neural network, which categories the images into cancerous or non-cancerous [4].In this paper the lesion image analysis tools for the various melanoma parameter like ABCD by texture, size and shape analysis for image segmentation and features extraction [5].The system consist of two phases whether the pigmented skin lesion is malignant or benign, the second phases recognizes melanoma skin cancer types [6]. In this system Otsu thresholding, median filter to remove noise has been used. Various classifier such as SVM, KNN, Decision tree (DT) and Boosted tree (BT). Comparison of the classification is done. The algorithm shows the accuracy of classification rate of KNN 92.70%, SVM is 93.70%, DT is 89.5% and finally the boosted tree(BT) is 84.30% [7].An automatic skin cancer classification system is developed and the relationship of skin images across type of training networks are studied. 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 [8].In this paper the author propose a new skin melanoma CAD system using texture analysis methods. The proposed CAD system consist of four steps, hair removal, filtering, feature extraction and classification. In the feature extraction step, they evaluate the performance of five widely used texture analysis methods, Grey level co-occurrence matrix, Gabor filter, Histogram of oriented gradients, Local binary patterns and Local directional number. The experimental result shows that extracting HOG features after hair removal yields the best classification results. HOG gives in AUC of .9783 with melanoma or common nevi classification and an AUC of .9439 with melanoma or dysplastic nevi classification [10].

III. Proposed system

Four different stages in proposed system. They includes preprocessing, segmentation, feature extraction and classification. Two system developed with different preprocessing and segmentation technique. Knn classifier had been used for classification. The performance of both system computed, rely on true report of parameter computed using confusion matrix. Fig3 shows corresponding steps in the system and architecture of system are shown in fig 4. The two system designed and implemented using matlab. Apart from that color pixel measurement of lesion region were also done.

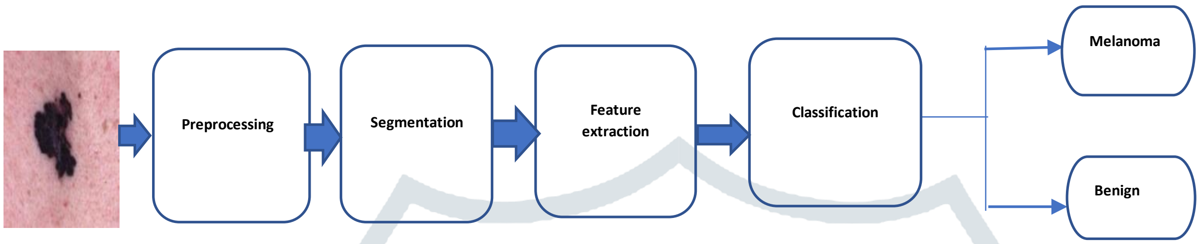


Fig 3 Proposed Skin Lesion Detection System

A) Preprocessing

Before the images can be segmented, they need to be preprocessed and cleaned so as to improve the performance of the segmentation algorithm. The preprocessing of first system include: Input the rgb image, convert it into greyscale using principle component analysis image using equation 1. After remove the noise by applying median filter. Perform complementation and morphological close operation. While in second preprocessing technique convert rgb to gray conversion. Gaussian filter used for noise removal.

$$.21R+.72G+.07B \qquad (1)$$

Where R –red, G –green and B -blue

B) Segmentation

The preprocessed image is then segmented using Otsu's thresholding algorithm, a global thresholding algorithm .For first system Otsu's algorithm differentiates between the background and the foreground of the image by thresholding the image above a computed grey scale value. The decomposition of image perform using 2 level discrete wavelet transform, then apply calculated threshold of four wavelet output. The segmented image is then masked upon the original image and a segmented colored lesion is obtained. The reconstruction of image formed using inverse wavelet transform. In the second system active contour was used for segmentation.

C) Feature Extraction

After segmentation of image, the lesion region identified, then feature extracted of lesion region using Grey-level co-occurrence matrix. Here four glcm features extracted from lesion region, which are listed in the table 1 and other color feature such as mean, standard deviation also extracted from the segmented region. Then features as given as input to classifier for classification. Below given the description of working of glcm feature extraction techniques.

GLCM

A statistical method of examining texture that considers the spatial relationship of pixels is the gray-level co-occurrence matrix (GLCM), also known as the gray-level spatial dependence matrix. The GLCM functions characterize the texture of an image by calculating how often pairs of pixel with specific values and in a specified spatial relationship occur in an image, creating a GLCM, and then extracting statistical measures from this matrix. After you create the GLCMs, you can derive several statistics from them using the graycoprops function. These statistics provide information about the texture of an image [16].

- Contrast: Measures the local variations in the gray-level co-occurrence matrix.
- Correlation: Measures the joint probability occurrence of the specified pixel pairs.
- Energy: Provides the sum of squared elements in the GLCM. Also known as uniformity or the angular second moment.
- Homogeneity: Measures the closeness of the distribution of elements in the GLCM to the GLCM diagonal.

Feature	Formula
Contrast(Cont.)	$\sum\sum (i,j)^2 C(i,j)$
Correlation(Corr.)	$\sum\sum \frac{(i-\mu_i)(j-\mu_j)C(i,j)}{\sigma_i\sigma_j}$
Homogeneity (Homo.)	$\sum\sum \frac{C(i,j)}{1+(i,j)^2}$
Energy (Enrg.)	$\sum\sum C(i,j)^2$

Table 1 GLCM features

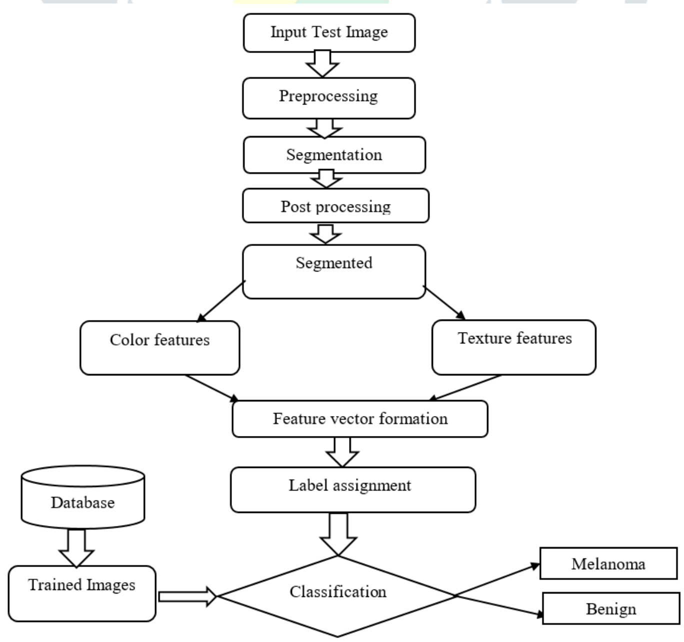


Fig 4 Architecture of proposed system

D) Classification

Classification of image as melanoma or benign were done using knn classifier. KNN is a classification or supervised learning algorithm. For training 200 images were taken among that 140 melanoma and 110 were benign. We would be looking at the K nearest training data points and take the most frequently occurring classes and assign that class to the test data. Therefore, K represents the number of training data points lying in proximity to the test data point which we are going to use to find the class. It classifies the rows of the data matrix Sample into groups, based on the grouping of the rows of Training. Sample and Training must be matrices with the same number of columns. Group is a vector whose distinct values define the grouping of the rows in Training. Each row of Training belongs to the group whose value is the corresponding entry of Group. Knnclassify assigns each row of Sample to the group for the closest row of Training. Group can be a numeric vector, a string array, or a cell array of strings. Training and Group must have the same number of rows.

Syntax

Class = knnclassify (*Sample*, *Training*, *Group*)

E) Analysis of system using Confusion matrix

It is a table that is often used to describe the performance of a classification model on a set of test data for which the true vales are known. The confusion matrix displays the total number of observations in each cell. The rows of the confusion matrix correspond to the true class, and the columns correspond to the predicted class. Diagonal and off-diagonal cells correspond to correctly and incorrectly classified observations [15].

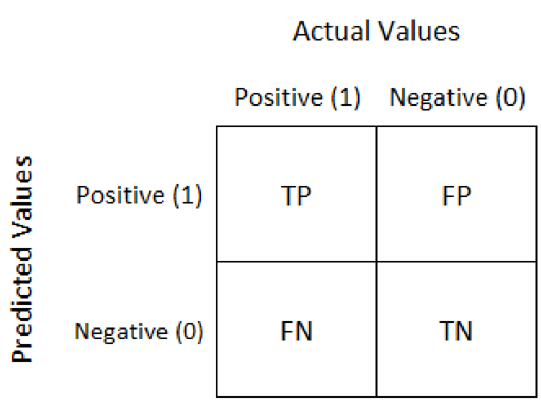


Fig 5: Confusion matrix

F) Algorithm for measuring color pixel and knn classifier

Algorithm 1:

Step 1: Read rgb image
Step 2: detect the skin lesion region using segmentation technique using first system preprocessing technique
Step 3: The image contains the lesion extracted from the surrounding skin then calculate the color pixel.
Step 4: Output six color pixels black, white, red, light brown, dark brown, and blue gray in the lesion area detected.

Algorithm 2:

- Load the training and test data
- Choose the value of K. For getting the predicted class, iterate from 1 to total number of training data points,
  For each point in test data
  - i) Calculate the distance between test data and each row of training data. That is find the Euclidean distance to all training data points.
  - ii) Store the Euclidean distances in a list and sort it.
  - iii) Assign a class to the test point based on the majority of classes present in the chosen points.
- Return the predicted class.

G) System implementation using MAT LAB: It is basically a high level language used in technical computing. It provides an integrated programming environment along with sophisticated tool/functions for computations and visualization. MATLAB supports wide range of data structures, built-in tools for editing and debugging and visualization, along with the extended support for some the object-oriented features. Thus MATLAB is considered to be a best tool in the field of education and research.

IV RESULT

The proposed method is implemented with MATLAB platform. The images are processed and their features evaluated using the Image Processing Toolbox, A set of 200 images were acquired from different web sources. This dataset is available for research and educational purposes belonging to melanoma and benign. This dataset of images were divided into 140 with colors of white, red, light-brown, dark-brown, blue-gray and black and 110 as benign Training taken place on those set of images categorized into melanoma and benign as the way images are tested, these method usually consist of the three steps:-border detection of skin lesion region, feature extraction and classification. Here using guide gui are designed.

A) Implemented system result

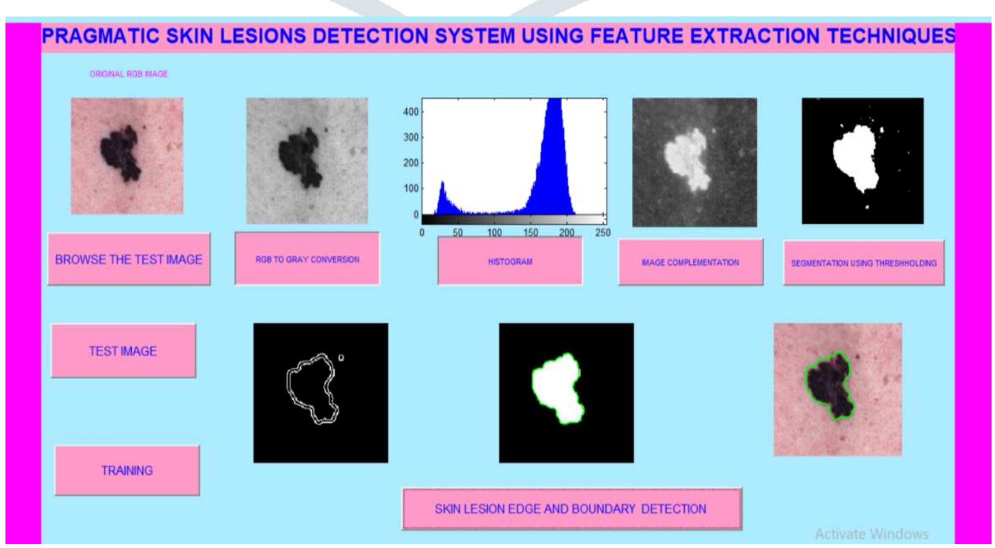


Fig 6 Skin Lesion detection system1

Training image	Mean	Standard_Deviation	Entropy	Variance	Smoothness	IDM
TH 1	0.008538	0.092008634	0.0709403	0.008253	0.9995498	0.017313
TH 2	0.087569	0.282667946	0.4283038	0.062636	0.9999561	0.149169
TH 3	0.019719	0.139034014	0.1398612	0.016255	0.999805	0.041621
TH 4	0.131481	0.337926051	0.5614853	0.075047	0.9999707	118.9203
TH 5	0.021592	0.145348413	0.1502899	0.018454	0.9998219	0.039998
TH 6	0.004612	0.067751678	0.0424257	0.004479	0.9991667	0.010652
TH 7	0.033442	0.179788888	0.2113714	0.031295	0.999885	47.28812
TH 8	0.017554	0.13132317	0.1274765	0.016752	0.9997809	0.052537
TH 9	0.098373	0.297818998	0.4638171	0.085728	0.9999609	152.8349
TH 10	0.0702	0.255484268	0.3666696	0.061623	0.9999452	52.13114
TH 11	0.099269	0.299023736	0.4666748	0.086242	0.9999613	16.29309
TH 12	0.080769	0.272480913	0.4048838	0.068992	0.9999524	104.189
TH 13	0.036085	0.186501072	0.2240428	0.034478	0.9998934	0.050984
TH 14	0.002027	0.044975798	0.0210551	0.002025	0.9981061	0.000737

Table II Color Feature

Training image	Contrast	Correlation	Energy	Homogeneity
TH 1	0.01128	0.3362	0.97185	0.994359073
TH 2	0.00262	0.98368	0.83703	0.99869112
TH 3	0.00164	0.95762	0.95955	0.999177606
TH 4	0.00395	0.98275	0.76693	0.998023166
TH 5	0.00132	0.96887	0.95627	0.999339768
TH 6	0.00706	0.23415	0.98378	0.996471042
TH 7	0.04032	0.37856	0.89641	0.979837838
TH 8	0.01856	0.46382	0.94716	0.990718147
TH 9	0.10171	0.42862	0.73064	0.949146718
TH 10	0.07527	0.42543	0.79938	0.962362934
TH 11	0.09548	0.4679	0.73419	0.952258687
TH 12	0.06888	0.53776	0.78685	0.965559846
TH 13	0.03214	0.53971	0.89907	0.983930502
TH 14	0.00011	0.97338	0.99583	0.999945946

Table III GLCM features

Bar chart of feature vector of image 30 detected as Melanoma skin lesion

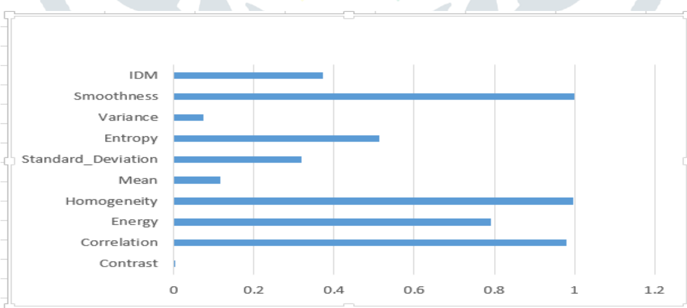
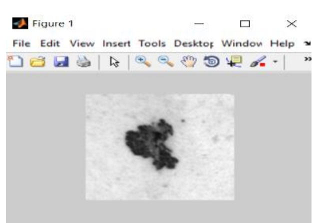


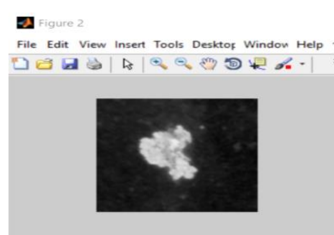
Fig 7 Bar chart

Below are the final output images stage by stage of First lesion detection algorithm and classification system

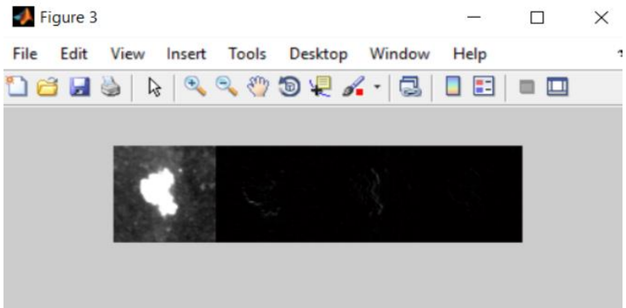

(a)



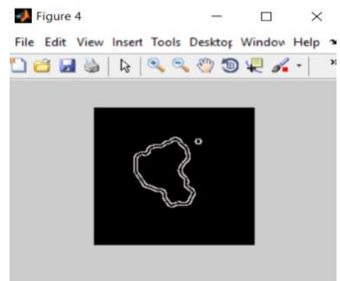
(b)



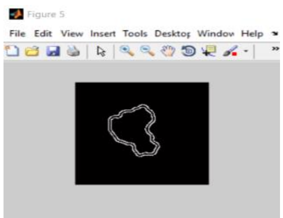
(c)



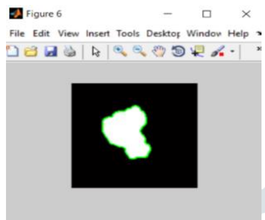
(d)



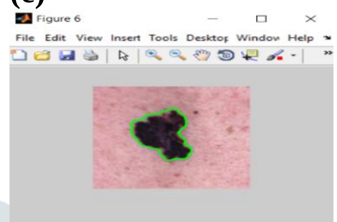
(e)



(f)



(g)



(h)

Resulted Images Fig 8 a) RGB Image b) RGB to Gray conversion c) Complemented image

d) Four Wavelet output e) Edge Detection f) Region Elimination g) Filling lesion region

h) RGB image boundary detection

B). System Evaluation result

KNN classifier result had been judged by confusion matrix. Sensitivity indicates to the percentage of the correct detection of malignant melanoma (the percentage rate of true positive detection). Specificity indicates to the percentage of the correct detection of the non-melanoma skin lesion i.e. benign (the percentage rate of false positive detection). Both of sensitivity and specificity could be computed by using the following concepts: True positive (TP): positive cases classified as positive. True negative (TN): negative cases classified as negative. False positive (FP): negative cases classified as positive. False negative (FN): positive cases classified as negative. Here a, Precision, Recall and F-measure to evaluate the performance of our CAD system. Precision could be defined as positive prediction value, Recall is the same as sensitivity, F-measure is the measure of classifier effectiveness which depends on both precision and recall as shown in the F-measure equation. The following equations equated using confusion.

$$total = TP + TN + FP + FN \quad (1)$$

$$accuracy = \frac{TP+TN}{total} \quad (2)$$

$$recall = \frac{TP}{TP+FN} \quad (3)$$

$$specificity = \frac{TN}{TN+FP} \quad (4)$$

$$precision = \frac{TP}{TP+FP} \quad (5)$$

$$F_{measure} = \frac{2*recall*precision}{recall+precision} \quad (6)$$

$$misclassification\ rate = \frac{FP+FN}{total} \quad (7)$$

Two system developed and tested using KNN. The different parameter as show in equation 1 to 7 evaluated. The computed parameter are list in table3 for comparisons. The result is obvious by this, the first system is better than second regarding accuracy, sensitivity, specificity. From the experiment result shown that KNN classify more accurately using first system than second system. The KNN classifier applied two system done prediction of 185 images. In which actually 120 images are melanoma and 65 are benign images.

System	classifier	Accuracy	Sensitivity (TP rate)	Specificity (TN rate)	Precision	F-measure	Misclassification rate
SYS1	KNN	91%	95%	85%	92%	93%	8%
SYS2	KNN	78%	79%	76%	86%	82%	22%

Table III parameters for measuring performance

C) Advantages of knn

- Knn is simple to understand and implement.
- Knn executes quickly for small training data sets.
- No retraining is required if the new training pattern is added to the existing training set.
- Don't need any prior knowledge about the structure of data in the training set

V Conclusion

At successfully execution of project and resulted segmented lesion image using two different preprocessing technique and two different segmentation technique, detected skin lesion region and color pixel measurement technique used which accurately count the number of six different color pixel in the skin lesion region. Gray level co-occurrence matrix based feature extraction to obtain energy, entropy, contrast, correlation. These texture features are served as the input to classify the image accurately. From the experiment result yield first system has misclassification rate 8% compare to second system with 22%.Effective use of multiple features of the image and the selection of a suitable classification method are especially significant for improving classification accuracy. Here KNN classification techniques used for improving accuracy. In future work different texture extraction technique such as LBP, HOG would be used and compare against GLCM using different types of classifier such SVM and Neural Network.

Acknowledgement

I, **Suruchi Padhy** would like to place my deep sense of gratitude to our college chairman **Shri. D. K. Mohan** for providing excellent Infrastructure and Academic Environment at CITECH without which this work would not have been possible. I am extremely thankful to our principal Dr**. Suresh L** for providing me the academic ambience and everlasting motivation to carry out this work and shaping our careers. I express my sincere gratitude to **Dr. Shashi Kumar,** HOD, Dept. of Computer Science and Engineering, CITtech, Bangalore, for his stimulating guidance and support. I extend my thanks to my co-author **Prof Sandeep Kumar** for his extended help to complete this work. I also like to extend my thanks to all my friends and family who were a constant source of encouragement throughout this work.

References

[1] Fathima Nizar, G.S Santhosh Kumar ,Classification of Various Skin Lesions using SVM and KNN Classifiers, International Journal of Innovative Research in Computer and Communication Engineering ,Vol. 4, Issue 8, August 2016

[2] Research Scholar, Dr.N.G.P Arts and Science College, Bharathiar University, Coimbatore, India, Vol-6, Special Issue-8, Oct 2018 , International Journal of Computer Sciences and Engineering, Vol-6, Special Issue-8, Oct 2018

[3] Sumithra Ra, Mahamad Suhilb, D.S.Guru, Segmentation and Classification of Skin Lesions for Disease Diagnosis.

[4] Skin Cancer Detection using Arificial Neural Network, Ekta Singhal, Volume 6, No. 1, Jan-Feb 2015

[5] Shivangi Jaina, Vandana jagtapb, Nitin Pise ,Computer aided Melanoma skin cancer detection using Image Processing, Procedia Computer Science 48 (2015) 735 – 740

[6] Dr.Abbas Hanon. Alasadi , Baidaa M.ALsafy ,Early Detection and Classification of Melanoma Skin Cancer, I.J. Information Technology and Computer Science, 2015, 12, 67-74 Published Online November 2015 in MECS.

[7] Automatic Detection and Classification of Skin Cancer ,Akila Victor Muhammad Rukunuddin Ghalib,International journal Of Intelligent Engineering Systems

[8] Dr. M. Vimaladevi , B.Kiruba, k-nn classifier for skin cancer classification, International Journal on Computer Science and Engineering (IJCSE)

[9] Tanuja Sarode2, Uzma Bano Ansari,Skin Cancer Detection Using Image Processing , International Research Journal of Engineering and Technology (IRJET) Volume: 04 Issue: 04 | Apr -2017

[10] Habeba Mahmoud, Mohamed Abdel-Nasser, and Osama A.Omer,Computer Aided Diagnosis System for Skin Lesions Detection using Texture Analysis Methods, 2018 International Conference on Innovative Trends in Computer Engineering (ITCE 2018), Aswan University, Egypt.

[11] http://www.asds.net/skin-experts/skin-conditions/skin-lesions

[12] https://www.asds.net/skin-experts/skin-conditions/skin-lesions

[13] https://accessmedicine.mhmedical.com

[14] https://www.skincancer.org/skin-cancer-information/melanoma

[15] http://www.dataschool.io/simple-guide-to-confusion-matrix-terminology

[16] https://www.mathworks.com/help/images/texture-analysis-using-the-gray-level-co-occurrence-matrix-glcm.html