ALLERGIC DISEASE CLASSIFICATION

A
PROJECT REPORT
Submitted in partial fulfillment of the requirements for the award of the degree of

BACHELOR OF TECHNOLOGY
IN
COMPUTER SCIENCE & ENGINEERING

Under the supervision
of
Prateek Thakral
(Assistant Professor)

by
Rohan Thakur
(141286)

to

JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY
WAKNAGHAT, SOLAN – 173234
HIMACHAL PRADESH, INDIA
MAY 2019

STUDENT'S DECLARATION

I hereby declare that the work presented in the Project report entitled "**Allergic Disease Classification** " submitted for partial fulfillment of the requirements for the degree of Bachelor of Technology in Computer Science & Engineering at **Jaypee University of Information Technology, Waknaghat** is an authentic record of my work carried out under the supervision of **Prateek Thakral**. This work has not been submitted elsewhere for the reward of any other degree/diploma. I am fully responsible for the contents of my project report.

Signature of Student

Name – Rohan Thakur

Roll No. 141286

Department of Computer Science & Engineering

Jaypee University of Information Technology, Waknaghat, India

Date-

CERTIFICATE

This is to certify that the work which is being presented in the project report titled "**Allergic Disease Classification**" in partial fulfillment of the requirements for the award of the degree of Bachelor of Technology in Computer Science & Engineering submitted to the Department of Computer Science & Engineering, **Jaypee University of Information Technology, Waknaghat** is an authentic record of work carried out by **Rohan Thakur (141286**) during a period from January, 2019 to May, 2019 under the supervision of **Prateek Thakral** Department of Computer Science & Engineering, Jaypee University of Information Technology, Waknaghat. The above statement made is correct to the best of our knowledge**.**

Date-

Signature of Supervisor	Signature of HOD	Signature of External
Name	Name	Examiner
Designation	Professor and Head	
Department of …...	Department of ……...	
JUIT,Waknaghat		

ACKNOWLEDGEMENT

I take this opportunity to express our gratitude to our supervisor **Prateek Thakral**, for his insightful advice, motivating suggestion, invaluable guidance, help and support in successful working of this project and also for his constant encouragement and advice throughout our Bachelors programme.

The in-house facilities provided by the department throughout the course are also acknowledgeable .I would like to covey our thanks to teaching and non teaching staff of Department of Computer Science & Engineering for their support.

Rohan Thakur
141286

TABLE OF CONTENTS

LIST OF FIGURES

ABSTRACT

In the poposed work we use image processing techniques to find the skin disease . we use the k mean classifiaction to seprate the region of the affected area to the normal skin. It uses the concept that pixel belonging to similar cluster have same values to pixel belonging to the different cluster.Then we Extract data from images and find the simliarity
of the calculated data to pre defined data and show which disease it is and give results.

1.INTRODUCTION

Skin is the biggest gift within the body . Its functions embody regulation of vital sign still as manifestaion of any change inside the body with the difference in color . The composition of it contains of associate degree outmost layer known as stratum. the highest layer of stratum contains a tricky, fibrous supermolecule known as scleroprotein whereas rock bottom layer contains animal pigment that may be a dark pigment that protects the body from the harmful rays of sun.

According to AAAAI,Worldwide, adverse drug reactions might have an effect on up to ten percent of the world's population and have an effect on up to twenty of all hospitalized patients

Findings from a 2009 to 2010 study of thirty eight,480 youngsters (infant to 18) indicated,8% have a allergic reaction.In 2010, black youngsters within the U.S. were a lot of seemingly to own had skin allergies (17%) than white (12%)or Asian (10%) youngsters.Worldwide, rash happens with lifespan prevalence higher than 2 hundredth . In 2012, 12.0% or 8.8 million youngsters reportable skin allergies within the past twelve months

A patient will pass though consequences of skin diseases if it's identified and corrected within the earliest stages and that couldbring home the cure ratios of over ninety five percent . Early identification relies on patient attention and correct identification by a caregiver.,thanks to the prices of skin doctor to watch each patient, there's a requirement for a processed or organised method to guage patient's extent of skin disorder mistreatment pictures of their skin changes captured employing a commonplace photographic technique.

The normal identification method consist of a collection of what our eye will see employing a photographic technique ,this concept look at rising the standardof already diagnosed systems by implementing advanced feature extraction

The most diseases that touching the skin originates beneath the skin, these changes are valid reason within the classification of a range of diseases. There's some truth within the belief that “the skin mirrors a person’s internal health“. The observability and accessibility of skin makes it as the primary organ of our body to point out common signs of underlying unwellness. Changes of the skin often recommend metabolic, malignant, and organ diseases.

The projected technique require collected pictures consists of eczema, benign carcinoma and malignant carcinoma image except for traditional pictures. The collected pictures undergo through various techniques of the image pre processing like changing the size, rgb to black image conversion and filtraion and changing the contrast.

After the image is done through pre processing the image undergo through the segmentation process using various methods such as c-means algorithm and watershed algorithm which find the affected region of the images . After finding the affected region in the image the features of the image are extracted using the technique of feature extraction such as Grey Level Co Occurence Matrix and Image Quality Assessment which generates the data from the segmented images than this data is collected and the new image undergo the same process as mentioned above and both data are matched to find the similarity thus the disease.

2.LITERATURE REVIEW

In 2012, R. Subash Chandra Boss , suggested that x ray image segmentation using the k-means clustering algorithm. The filter is used for pre filtration of image. Varius different feature upto 14 features are extracted using the image techinques of feature extraction

In 2009, Sookpotharom Supot, used the concept of finding the border or edges of the affected region in the given input subject by using the fuzzy c means and thresholding method . the first thing he did was analysing and pre process image and filtering. Next thing he did was to use fuzzy c-means thresholding method i to segment and identify the affected region.

Huiyu Zhou , He also used the same method of c means but the additional thing he did was to use anisotropic mean shift on top of the c means algorithm.Segmentation is the most useful feature in finding the boundry of affected region . He used a new mean shift based fuzzy c-means algorithm which required a low computational work and time than other propsed methods.

In 2009, Margarida Silverira , used the differnces between the segmentation techniques for skin cancer. Different methods are used thresholding , adaptive snake , EM level set, split-and-merge algorithm rule. Good results were obtained by the use of the adaptive and EM methods. The best automatic method was c means , with results only slightly worse.

In 2012, Lucia Ballerini ,used non-melanoma skin affected region differenciation using sample data in hierarichal based testing. This had aan accuracy of above 90 percent cancer and pre-malignant area from benign area, and has a accuracy of over 70 percent in five common classes.

Y.P Gowaramma ,used the color based watershed method which require to know the color of the affected region and he used the curvelet filter along with the use of k-n classifier .Then he used the GlCM techinques for the data extraction and classified the images based on the fuzzy classification.This proposed system has a higher accuracy than the existing methods.His colleagues used novel texture based on skin affected segmentation techniques to differenciate stages of the cancer by using PNN.They also used Joint Statistical Texture Distinctiveness to classify various melanoma cancer .

3.METHODOLOGY

Skin condition images were taken as input and their data features were extracted and they are matched with the collected images data set .

These images are subjected to the change in color ,rgb to black,changing the constract,brightness Techniques of pre -image processing were applied to these images. Then these image are required to go through segmentation algorithmic rule such as C-means and Watershed algorithmic rule to get the affected region on the image which can be clearly differenciated from the normal skin .

After the above process the image segmented image are required to go through data/features extraction and these were calculated using G.L.C.M method and I.Q.A features .After Feature extraction the data is used to identify the disease .

3.1Image Pre-processing

Pre processing is a often used term for functions performed on the image at the lowest level of use . The main use of pre processing is to improvement of image data that supresses distortion or enhances or changes the important features of the images which can be used for further enhancement of extraction .

Some steps of pre processing

1.Resizing of image,Sometimes we need to resize the image because the image we have and image required are different in size

2.Gaussian blur is used to smooth the image and remove the unwanted noise

3.Changing the contrast and brightness of the image

4.Seperating the RGB plane from the colored image.

3.2Segmentation Methods

Image segmentation is a process in which the image is divided into different segment . It is used to find some specific part in image or the bou ndry of region in an image .The goal of this is to divide image into cohrent region .there are various techniques of image segmentation but for the proposed work following two methods are used because these methods provide the best result in indentifying the affected region of the skin in the image provided.

- K-Means clustering
- Watershed Algorithm

3.2.1 K-Means Clustering Algorithm

The most common used algorithm is K mean clustering algorithm. This algorithm is very simple and the computation required for it is less than hierarchical . This method can also work for large number of data.

It produces different result for the different clusters,so the main thing required for it identify the no of clusters .Images which are used here are represented as an array in the system, and the no depicting the values of applicable quantities which define contrast between different part .Improves from of fuzzy c means is used which require otsu thresholding and using fcm and it has tendency to find threshold that is of large in ocurrence .this is performed at the edeges or boundry of clusters.

1.Image is read

2.Convert Image from rgb to L*A*B (the lab color space enables us to differenciate between different colors)

3.Then the colors in A*B space are classified using clustering(clustering differenciate the different group of colors, kmeans thinks that object in region of cluster are close to each other and far from other cluster.)

4.Now make image that segment the H&E image by color(we will get three images using pixel_label)

5.find the nuclei of the region

6.get the affected region

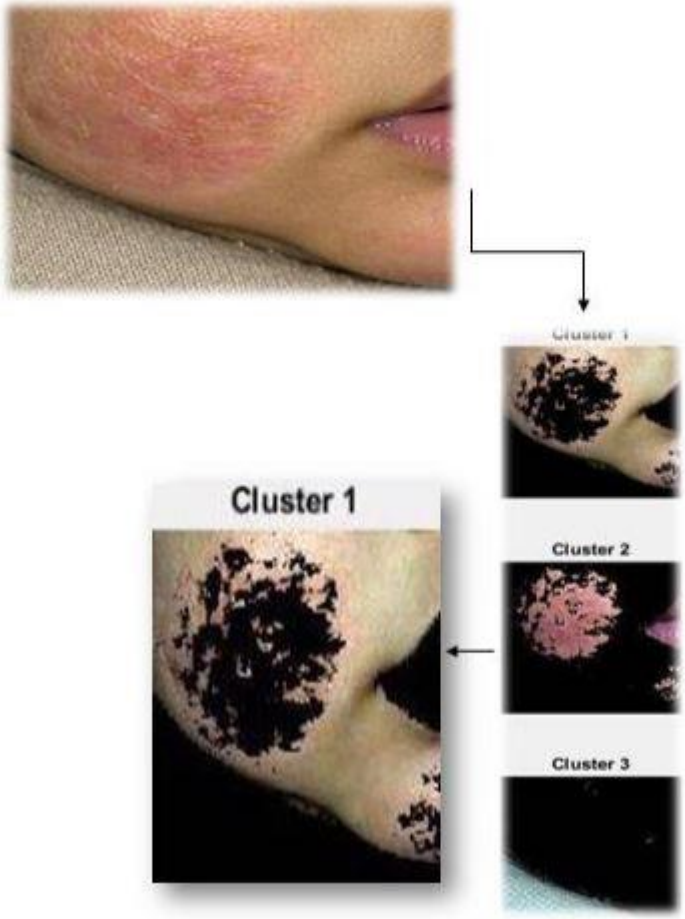


Fig 3.1

Above figure shows the clusters formed by k means clustring

3.2.2 Watershed Algorithm

Watershed is an method which is based on the area based method .In this method the sample image is seen as a real landscape with river and valleys. The gray values are defined as elevation values of the landscape

Using such 3D representation given algorithm interprets the image as an catchment basin. It seprates the basin from each other ,this method transform the image totally and assigns the each pixel to either a region and watershed.

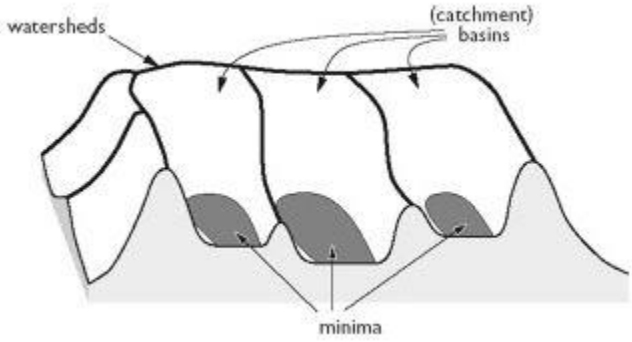


Fig 3.2

- Gradient magnitude of input image
- Erosion
- Dilation
- Reconstruction
- Complement

Gradient Displays the value of each pixel which are high and all pixel and intensity display all pixels position. Which is required to find out all of the pixel edges/boundaries.

Binary Dilation and Erosion the black pixels and white pixels represent an outline of image that is of binary nature. Now we consider only the pixel that are black and treat all the white pixels as the background . The operation which can be used are dilation .erosoion and difficult morphological operations such as gap opening, closing and shape decomposition can be established

Erosion this function is used to remove the affect region in the images which are very small and the image only contains affected region which is large and image is remade to get the lost data.

Dilation and Complement:It can be described as the change that changes all the pixels in the back by using the isotropic structring.it can be described as an increasing function.

It is employed to fill little holes and slim gulfs in objects. It will increase the size of the sample if first size must be saved and the above mention concept is used with the erosion.

Steps of this algorithm

1. Compute the segmentation function. This is an image whose dark regions we trying to segment.

2. Now we will compute foreground markers.

3. Now we will calculate the background markers. These are pixels that are not part of any object.

4. Change the segmentation function so that it only shows the minima function at the foreground and background marker locations.

5.Now we use watershed method for modified resutls

3.3Feature Extraction

Feature extraction means that the we tranform the input image into the group of features,those are the function of input pattern.
An image consists of pixels. Considering each pixel can have an 8bit value, even a 640x480 image will have 640x480x8 bits of information.. Too much for a computer to make head or tail out of it directly. So in feature extraction we figure out what parts of an image are distinctive , like lines, corners, special patches that can uniquely describe the image.

In simple words these are nothing but the distinctive features of the given image or special properties that shows an image.
Features are extracted in order to differentiate between the images.
Features extraction are used in almost all machine vision algorithms.

Two methods that are used as:

- GLCM

- IQA

3.3.1 GLCM:

GLCM is defined as grey level co occurence matrix .It is a mathematical method for examining the images that uses the relationship of pixels in glcm.This function charaterize the image texture by finding out how many a times a pair of the pixel is occur in the image that has some specified value and then measuring from the matrix

Statistic	Description
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.

Contrast feature

$$Contrast = \sum_{i,j=0}^{N-1} P_{ij} (i-j)^2$$

Energy feature

$$Energy = \sum_{i,j=0}^{N-1} \left(P_{ij}\right)^2$$

Correlation feature

$$Correlation = \sum_{i,j=0}^{N-1} P_{ij} \frac{(i-\mu)(j-\mu)}{\sigma^2}$$

Correlation feature

$$Correlation = \sum_{i,j=0}^{N-1} P_{ij} \frac{(i-\mu)(j-\mu)}{\sigma^2}$$

3.3.2 **IQA:**

The IQA is reffered as image quality Assessment .It focuses on the how close an input image is to the sample image.The mmodel of iqa can also be defined as full reference image assessment .It uses the inetensity function of 2 images and distorted version to assess the image.Iqa can also de divided into reduced refrence ,full refrence ,no refernce.

NR iqa uses the algorithm in which we can only use the distorted image and we have no data about the sample image.

Also ssim index and other structure oriented methods are also used.

immse	Mean-squared error
psnr	Peak Signal-to-Noise Ratio (PSNR)
ssim	Structural Similarity Index (SSIM) for measuring image quality

$$PSNR=10\cdot\log\sum Dh-Du22552$$

$$SSIM=lfgcfgsfg \qquad (2)$$

$$lfg=2\mu f\mu g+C1\mu f2+\mu g2+C1,cfg=2\sigma f\sigma g+C2\sigma f2+\sigma g2+C2,sfg=\sigma fg+C3\sigma f+\sigma g+C3$$

brisque	Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) no-reference image quality score
fitbrisque	Fit custom model for BRISQUE image quality score
brisqueModel	Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) model
niqe	Naturalness Image Quality Evaluator (NIQE) no-reference image quality score
fitniqe	Fit custom model for NIQE image quality score
niqeModel	Naturalness Image Quality Evaluator (NIQE) model
piqe	Perception based Image Quality Evaluator (PIQE) no-reference image quality score

1. $Sharpneesdegree = 1MN\sum i=0M-1\sum j=0N-1G2xy$

where

$Gxy = Dxy - Dx-1,y2 + Dxy - Dx,y-12$

2.BlurMetric=SumofalledgewidthsNumberofedges

3.$BIQI = \sum i=15pi \cdot qi$

where qi {i = 1,2,…,5} represents the quality scores from each of the five quality assessment algorithm

4 .ALGORITHMS

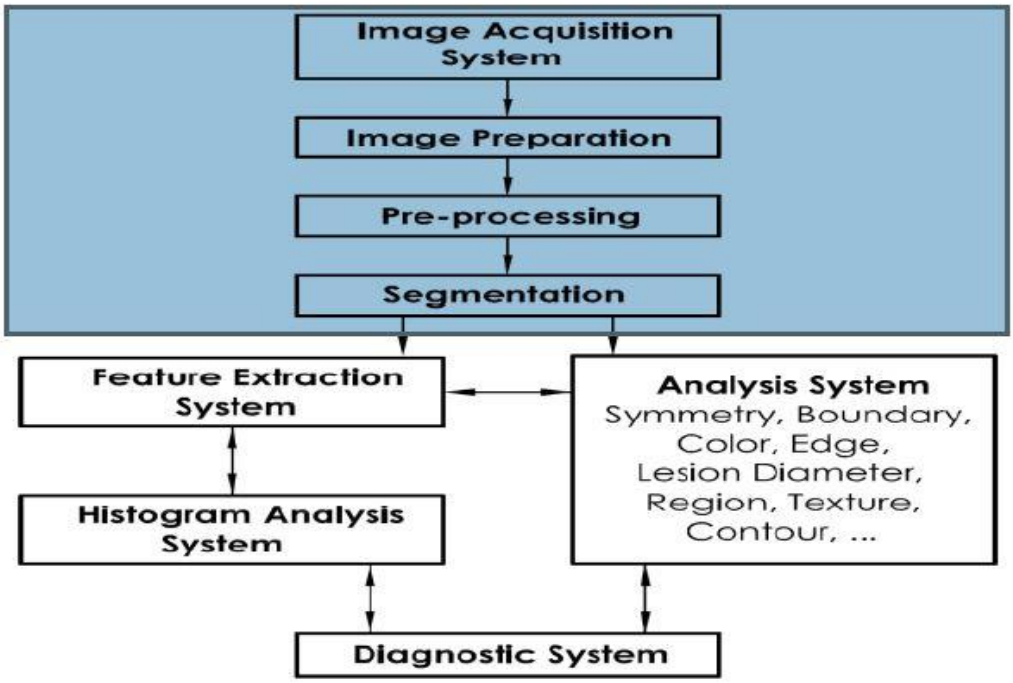


fig 4 .1

First step is to pre process the image

1. input image is uploaded and read.

2. Differnciate the R,G and B plane from color image.

3. Each of the three planes are applied with median filter.

4. All the three different rgb images are combined to get the next step

Segmentation skin image

1. The given image is changed into a vector.

2. Now the k mean algorithm is used on this image to genrate two clusters. One of the region shows to background skin and other shows to affected region.

3if the cluster number one has more pixel than cluster number two than cluster number two is affected region

4.image obtained above might contain some unwanted region which may not point to to the actual affected region to remove it all isolated region are calculated

5. The result we get in above step is a black and white image it shows affected region as white and rest as black

For Feature extraction

1. First we get the matrix of the image above

2. Find the centroid of the image usin above matrix.

3.find the border or edegs of the affected region using canny edge method

.

4.Find standard deviation of contour signature

4.1Contour Signature

Shape of the image can be considered as disatnce of all the meaningful points to some refernce point. Its additionally referred a contour signature . This reference point is centroid of image.Centre of this region is centre of all distance of all points of affected region . For other shape it will be different. In the given case points or edeges of the region of the image are considerd which are separated by angle of ten degrees.

4.2Co-Occurrence Matrices

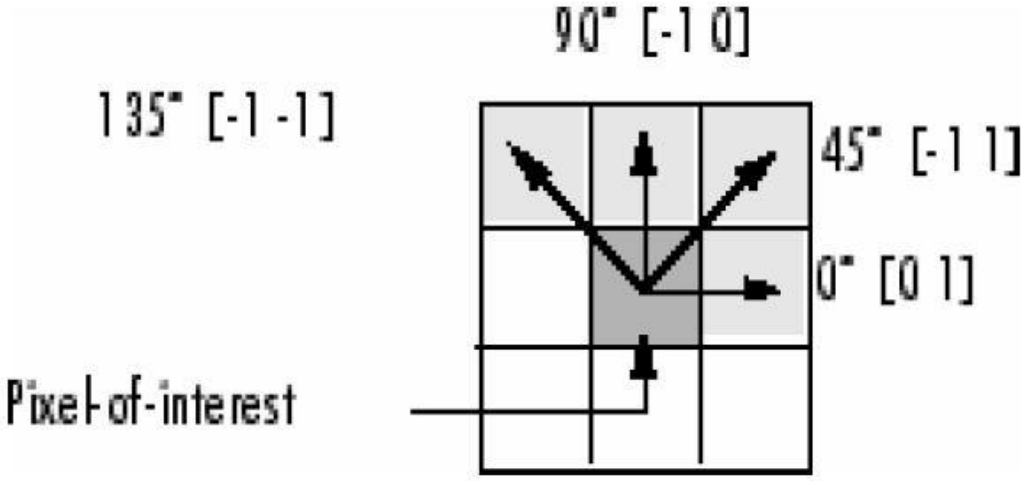
This matrix is a matrix which contains the occurence of a certain pair of pixels in an sample image. In order to find the co-occurrence matrix it is

We must know the following values:Number of grey levels: a greyscale image contains 256 grey levels, which means a high computational cost because all possible pixel pairs must be taken in account. The method is to generate the matrix reducing the number of greyscales, and so the number of possible pixel combinations.Above matrix is always a square . This value is often set to eight.Angle. Similarly to the distance it is necessary to define the direction of the pair of pixels. The most common directions are 0, 45º, 90º, 135º, and its symmetric equivalents.

.

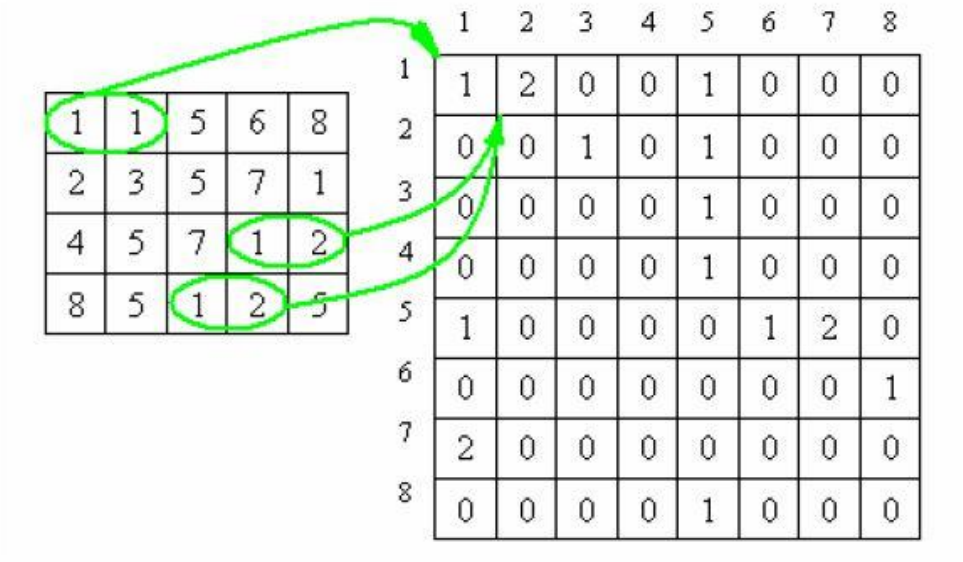
Statistical properties of the co occurence matrix

Property	Formula
'Contrast'	$\sum_{i,j} \lvert i-j \rvert^2 p(i,j)$
'Correlation'	$\sum_{i,j} \frac{(i-\mu i)(j-\mu j)p(i,j)}{\sigma_i \sigma_j}$
'Energy'	$\sum_{i,j} p(i,j)^2$
'Homogeneity'	$\sum_{i,j} \frac{p(i,j)}{1+\lvert i-j \rvert}$



Co occurance matrix fig 4.2

This matrix finds the Glcm from the scaled image.Initially I is a binary image this scales the image into 2 levels if I is the intrnsity image glcm scales the image into 8 gray level as depicted in image below .



Example of co-occurrence matrix process fig 4.3

Above figure shows how this matrix calculates values in the glcm of 4 by 5 image

5. RESULT

Input image

fig 5.1

2. Applying fcm

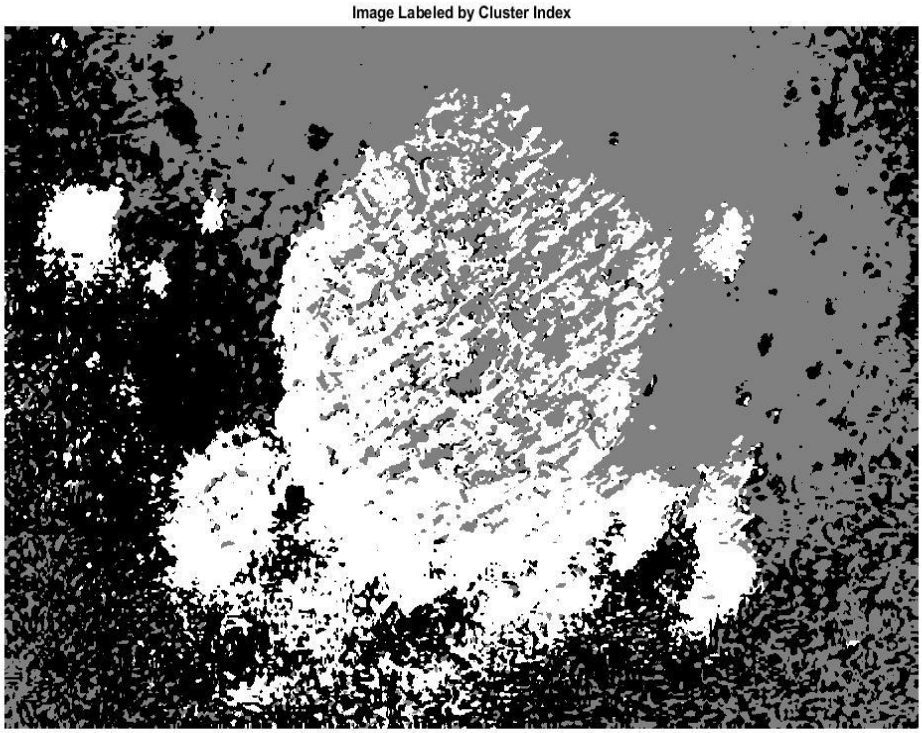


figure5.2

Segmented image

figure 5.3

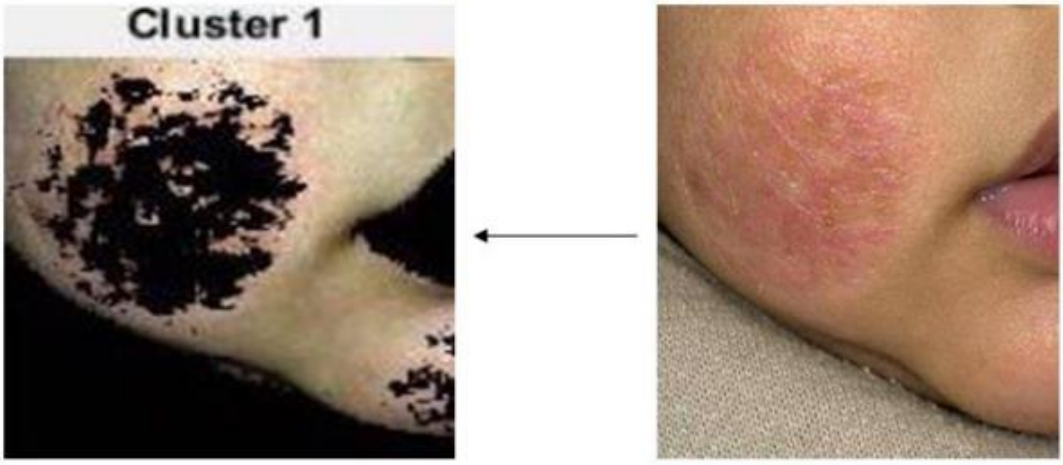


fig 5.4

parameters	fig 5.4	fig 5.3	average data	
Affected area	46.954	50.4558	54.9714	50.7937
contrast	0.9816	0.7996	0.9942	0.92513
correlation	0.8596	0.9280	0.9085	0.8987
energy	0.2846	0.2514	0.1999	0.2453
homogeneity	0.9288	0.9194	0.8977	0.9153
mean	56.0075	75.2324	83.7614	71.667
std	67.3785	80.9849	80.6263	76.3299
entropy	4.5385	4.8004	5.3139	4.2842
variance	4.1019e+03	5.0718e+03	5.5423e+03	4.905e+03
kurtosis	2.1574	1.6703	1.4004	1.7427
skewness	0.7461	0.4431	0.3697	0.5196

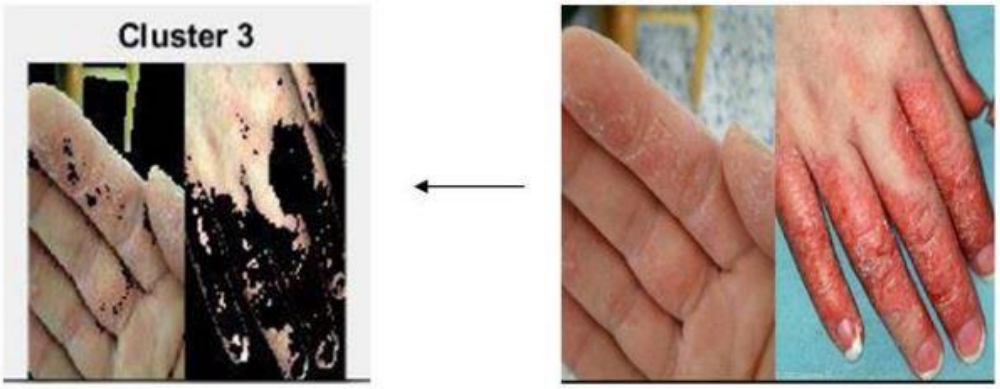


fig 5.5

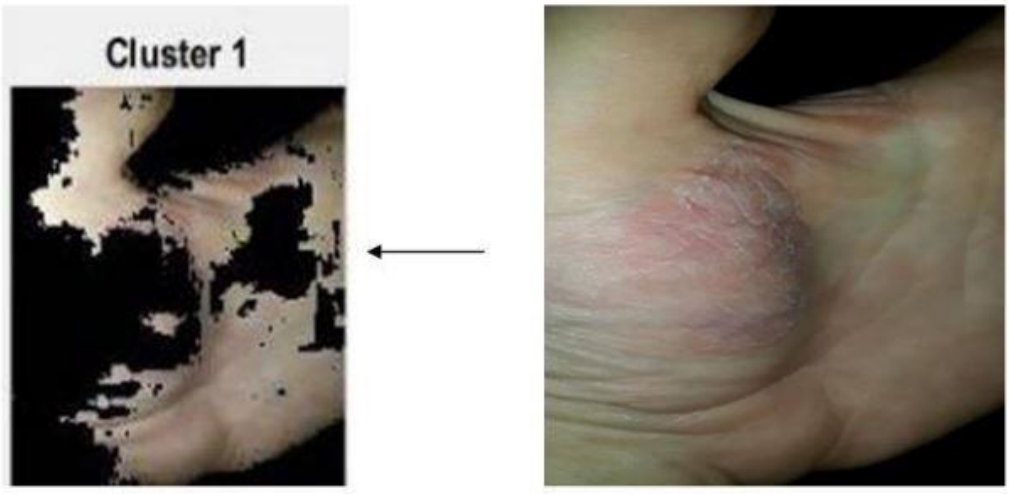


fig 5.6

parameters	fig 5.5	fig 5.6	average values	
Affected area	64.698	17.5537	43.6859	41.9792
contrast	1.5935	0.8989	0.9932	1.1618
correlation	0.8697	0.8337	0.9235	0.8756
energy	0.1686	0.4307	0.3280	0.3091
homogeniaty	0.8724	0.9316	0.9464	0.9168
mean	11.7704	37.7889	80.8234	43.4609
std	90.3369	61.1987	92.5032	81.346
entropy	5.5609	3.4140	4.3994	4.4581
variance	5.4504e+03	3.2824e+03	7.9760e+03	5.5696e+03
kurtosis	1.3996	3.5058	1.4943	2.133
skewness	-0.1632	1.3592	0.4463	0.5474

fig 5.7

parameter	fig 5.7	sample 1	average data	
Affected area	15.2881	14.653	15.0201	14.987
contrast	0.8342	0.5024	0.6409	0.6591
correlation	0.9412	0.9298	0.9182	0.9297
energy	0.2918	0.4031	0.5972	0.4307
homogeniaty	0.9546	0.9686	0.9550	0.9594
mean	83.3903	44.0921	37.2528	54.9117
std	91.3877	68.5514	73.6490	77.8627
entropy	4.5785	3.8074	2.3735	3.5864
variance	6.7652e+03	4.4708e+03	4.8297e+03	5.3552e+03
kurtosis	1.3870	3.2072	3.9967	2.8637
skewness	0.3603	1.2805	1.6356	1.0921

6.CONCLUSION

The above project work has assigned a classification sytem for the different common sin disease . the data was non monogenic for digital skin pictures that ar pre-processed exploitation image method techniques. The image segmentation is applied through the k-Means formula and Watershed formula on a private basis. An approach towards the advanced c- means that where Otsu thresholding is performed that helped in tracing the boundaries of the origin of the affected region. The watershed formula may be a slightly better than other methodology for region segmentation. Some functions of the watershed algorithm are used as Gradient magnitude, Erosion, dilution, reconstruction of abrasion for better reult.Based on this the images were differentiated .Different diseases were classified and their accuracy displayed further because the total space affected during a specific image is displayed within the graphical computer program (GUI) as output.

This proposed work can be used in the field of medicine as within the telemedicine for automatic diagnosing of various condition of eczema . This work also conjointly helps dermatologists to convey higher treatment to the patients by correct diagnosing of the many malady conditions. The system planned during this project is wont to give a bad value and economical resolution for automatic recognition of skin diseases. The multi-class classification will function a good tool in distinguishing skin diseases.

The long term work are reaching to be supported making algo to identify varied different skin , to extend efficiency and together to a lot of cut back the method time. On one hand this may well be helpful for dermatologists to cut back diagnostic errors, whereas on the alternative it'll operate the initial geographic point for patients in rural areas where there is a scarceness of fine medical professionals. a much bigger informationassortment and clarification of technique, along with side testing on comprehensive variety of photos,

7.Refrences

[1]R. Subash Chandra Boss, K. Thangavel, and D. Arul Pon Daniel,"Mammogram Image segmentation Using Fuzzy Clustering" *IEEE* Conference on Pattern Recognition, Informatics and Medical engineering, pp no 290-295, 2012.

[2] Sookpotharom Supot, "Border Detection of Skin Lession Images Based on Fuzzy CMeans Thresholding," in Third International Conference on Genetic and Evolutionary Computing, pp. 777-780 , 2009

[3] Huiyu Zhou, Gerald Schaefer, Abdul H. Sadka, and M.Emre, "Anisotropic Mean shift Based Fuzzy C-Means Segmentation of Dermoscopy Image," IEEE Journal of Selectd Topics in Signal Processing, vol.3. no. 1, pp 26-34, February 2009.

[4] Margarida Silveria, Jacinto C. Nascimento, Jorge S. Marques, Andre R.S.Marcal, Teresa Mendonca, Syogo Yamauchi, Junji Maeda, and Jorge Rozeira, "Comparison of Segmentation Methods For Melanoma Diagnosis in Dermoscopy Images," IEEE Journal of Selected Topic in Signal Processing, vol.3, no.1, pp. 35-45.

[5]Lucia Ballerini , Robert B. Fisher ,Ben Aldridge , and Jonathan Rees, "NonMelanoma Skin Lesion Classification Using Colour Image Data In A Hierarchical K-NN Classifier", IEEE Conference in Biomedical Imaging, pp 358-361, 2012

[6]Shang Keke, Zhou Peng, and Li Guohui, "Study on Skin Color Image Segmentation used by Fuzzy C-Means Arithmetic," in Seventh International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2010).vol. 2, pp. 612-615,2010.

[7] www.ijarcsse.com/docs/papers/volume-3/7- july2013/v317-0279.pdf IEEE TRANSACTIONS ON IMAGE PROCESSING, IQA Fake Biometric Detection, February 2014.

[8]http://www.datadictionary.nhs.uk/data_dictionary/messages/clinical_data_sets/data_sets/cancer_outcomes_and_services_data_set/cancer_outcomes_and_services_data_set_-_skin_fr.asp?shownav=1

[9]http://www.datadictionary.nhs.uk/data_dictionary/data_field_notes/l/le/lesion_diameter_greater_than_20mm_indication_code_de.asp?shownav=0 http://www.datadictionary.nhs.uk/data_dictionary/attributes/l/len/lesion_diameter_greater_than_20mm_indication_code_de.asp?shownav=0

[10]http://www.datadictionary.nhs.uk/data_dictionary/messages/clinical_data_sets/data_sets/cancer_outcomes_and_services_data_set/cancer_outcomes_and_services_data_set_-_skin_fr.asp?shownav=1

[11]https://archive.ics.uci.edu/ml/machine-learning databases/wisconsin/wdbc.data

[12] http://www.labbookpages.co.uk/software/imgProc/otsuThresh old.html