Indian J.Sci.Res. 16 (2): 63-71, 2017 ISSN: 0976-2876 (Print)
ISSN: 2250-0138(Online)

SKIN DISEASE CLASSIFICATION USING GLCM AND FCM CLUSTERING

PRADEEP MULLANGI[1a],Y. SRINIVASA RAO[b] AND PUSHPA KOTIPALLI[c]

[a]Department of ECE, Shri Vishnu Engineering College for Women, Andhra Pradesh, India

[bc]Department of Instrument Technology Department, A.U. College of Engineering, Andhra Pradesh, India

ABSTRACT

Classification of skin diseases based on texture analysis is presented in this paper. Main focus is on the extraction of statistical features inclusive of contrast, correlation, homogeneity, and energy for detection of the skin disease. Spatial relationship of pixels is determined using a statistical method gray-level co-occurrence matrix (GLCM) and then statistical features are measured. GLCM features are applied on the skin disease dataset images. The dataset is divided into training & testing dataset. The feature values are calculated along four GLCM directions 0°, 45°, 90° & 135° directions. These values are then averaged to get final feature vectors. Chi square distance is measured between feature vectors of the reference and test images. As hundreds of images are present as train and test images, time consumed by GLCM based image classification is impracticable, we propose Fuzzy C- means (FCM) Clustering along with GLCM. The proposed method reduces the time taken for skin disease classification. It is shown through simulation results that the proposed method is more efficient than GLCM alone method.

Keywords: Dermatology, FCM, Fuzzy logic, Fuzzy C-means clustering, GLCM, Gray level co-occurrence matrix, Confusion matrix, Image classification, Image texture analysis, and Feature extraction

Skin disorders are among the most common diseases in both developing and industrialized countries. People living with skin disease experience stressful life as skin disease affects their confidence and self-esteem in so many different ways. In 2013, with prevalence rate of 10 percent, the population affected across India from skin disease is estimated at nearly 151 millions. It is estimated that at a compound annual growth rate (CAGR) of 12 percent, about 188 million people is likely to suffer from skin disease by 2015. The severity of growing skin diseases in India is further emphasized by the fact that the World Health Organization (WHO) has included skin disease under the most common non-communicable diseases in India. In addition, there is a lack of facilities that provide comprehensive skin related treatments under one roof. The situation is further worsened by the low availability of dermatologists in India. At present, there are about 6,000 dermatologists catering to a population of over 1.21 billion. This means that for every 100,000 people, only 0.49 dermatologists are available in India as compared to 3.2 in many states of the US [1].

In this scenario, medical imaging plays an important role for quick decision making in skin disease identification. Medical imaging is used for revealing internal structures hidden by the skin as well as to diagnose and treat disease. This paper discusses detection of skin diseases using texture based image classification.

Problem Statement and Objectives

Few sample images of skin diseases[2] are shown in Figure 1.

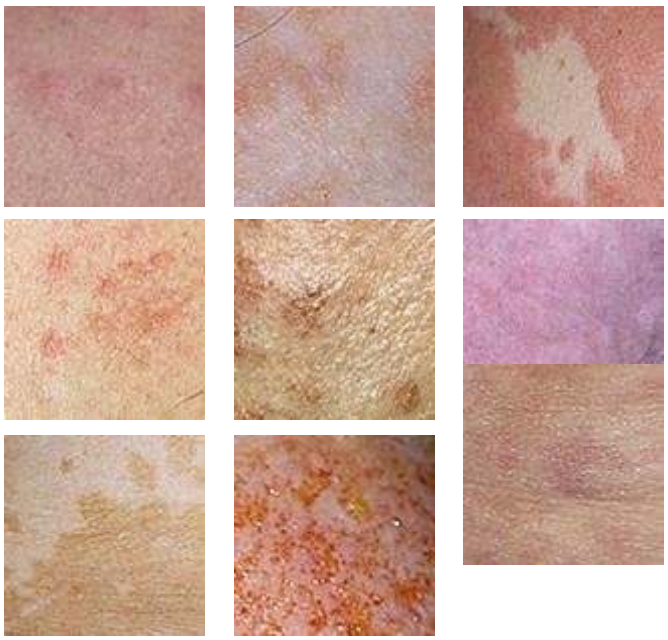


Figure 1: Skin disease images of different classes

All the above images appear to be similar but in real they belong to three different skin diseases Angioedema, Actinic Keratosis and Eczema. First column images belong to Eczema.

1Corresponding Author

Similarly, second and third column images belong to Actinic Keratosis and Angeoedema respectively. Hence, it is not possible for a human eye to identify the skin disease on observation. Dermatoscopy can be used but it takes more time for the evaluation of the results and sometimes it may give errors, therefore image processing methods are more suitable in application to skin disease classification.

Image Texture plays an important role in describing and assessing object surfaces. Texture of an object surface depends on spatial relation between primitive texture elements. Image texture features can be extracted using several methods: using statistical, structural, model-based and transform information. In recent years, researchers have developed methods based on Texture analysis to diagnose skin diseases.

The main objective of this paper is to classify given test image of skin disease using texture based image processing method in combination with clustering technique. Another objective is to show that the efficiency of texture based method in combination with clustering technique is more than the efficiency of texture based method in skin disease classification.

Literature Survey

Damilola [3] proposed automatic diagnosis of skin cancer using well-defined segmentation and classification technique. Arivazhagan [4] developed texture analysis based method for recognizing human skin diseases. In this method they classified skin diseases by extracting Independent components. Sparavigna and Marazzato [5] proposed a texture based method in which differences in color and coarseness of skin are quantitatively evaluated by using statistical approach to the pattern recognition. Sheha et al. [6] used Grey-Level Co- occurrence Matrix (GLCM) and Multilayer perceptron classifier (MLP) for automatic Detection of Melanoma Skin Cancer using Texture Analysis. Anand and Ashwin [7] proposed a method for plant disease detection. They used Gabor filter for noise removal and segmentation purpose. Texture based and color based features are extracted from the result of segmentation and artificial neural networks are applied for classification. This method yields better efficiency when number of features are 168 for an image, number of hidden neurons are 50. Extracting 168 features for each image in classification is a very time consuming process and practically not suitable in real-time applications. Authors of [8] used Runlength Matrix in extracting texture based features like and recognized the fungal disease affected fruits by using Artificial Neural Networks. Abadi et al.[9], achieved 96% accuracy for 300 skin texture images but used 9 color features in addition to 4 texture feature, entropy, energy, contrast and homogeneity. Dhandra et al.[10] extracted features by considering a total of 121 features and the Support Vector Machine (SVM) with RBF kernel. The system was tested on a set of 180 medical images and the experimental result confirms the efficiency of the proposed method as 92.2% accuracy. Oberti et al [11] proposed the design of an automatic sensitivity (unhealthy tissue) detection system. They performed segmentation by fitting Gaussian mixture to a histogram of NIR (Near Infrared) pixel intensities by the implementation of an 'expectation maximization' algorithm. Then, rule-based classification is applied to compare non- infected and disease affected tissues. The authors stated that delectability of sensitivity could be improved under specific conditions and laboratory setup. Authors also suggested doing imaging an angle within the range of about 40–60 degrees. These conditions somehow make the task more complex and time-consuming. Anil Kumar [12] proposed GLCM based skin disease classification. They considered 60 images per class, out of which 30 images are chosen as the training set. Only two features are used in classification. They stated that several GLCM based features should be combined together for better modeling possibilities. Authors of [13] retrieved images based on texture features using GLCM and image sub-blocks and Manish Pawar [14] tried skin disease classification using three features of GLCM and neural networks, and achieved 66.7% efficiency.

Selection of Clustering Method

Clustering is the process of grouping samples so that the samples are similar within each group. Clustering is used in statistical data analysis, data mining, pattern recognition, image analysis etc.

There are three types of non-hierarchical clustering techniques mostly used in image classification. They are

1. Artificial Neural Networks
2. K-means clustering method, and
3. Fuzzy C-means clustering method

Selection of clustering method depends on the criteria that the percentage of observations that are correctly classified and also, on the accuracy of the clustering algorithm which is assessed by internal dispersion of the groups in the partition.

Artificial Neural network builds models that are more reflective of the structure of the data in significantly less time. It operates well with modest computer hardware and can continue without any problem even if an element of neural network fails. It is good at adapting to changing situations. Key limitation of neural network is its inability to explain how the network has been build. Neural network gets better answer but have hard time explaining how they got there. Extraction of rules from neural network is difficult. Also, it is a time consuming process of training the neural network from complex data set [15].

The accuracy of the K-means procedure is very dependent upon the choice of the initial seeds of clustering, one for each cluster. To obtain better performance the initial seeds should be very different among themselves. As the K-means algorithm the desired number of clusters C has to be pre-defined and c initial seeds of clustering are required to perform the Fuzzy C- means. Contrary to the K-means method the Fuzzy C-means is more flexible because it shows those objects that have some interface with more than one cluster in the partition [16].

In this paper, GLCM is used for measuring texture features such as correlation, contrast, energy and homogeneity.

The remainder of this paper is organized as follows. Proposed method is presented in Section 2. GLCM calculation for a given image, texture measures from GLCM are discussed in Section 3. Section 4 presents Fuzzy C-means clustering. In Section 5, performance measures of GLCM and GLCM with FCM classification methods are compared.

PROPOSED METHOD

This method is developed for classification of skin diseases by analyzing textures obtained from a collection of images using features based on GLCM. This method has two phases:

(a) Training phase and (b) Classification phase.

We considered extraction of four texture features of an image: Correlation, Contrast, Homogeneity and Energy. In training phase, the texture features are extracted from a given set of training images. In classification phase, the given test image is segmented and then the above mentioned texture features are extracted. Then, Fuzzy C-means clustering is applied on the texture features obtained from training and test images. Then Chi. Square distances are calculated for skin disease classification. The flow chart shown in Figure 2 explains the methodology of our proposed method.

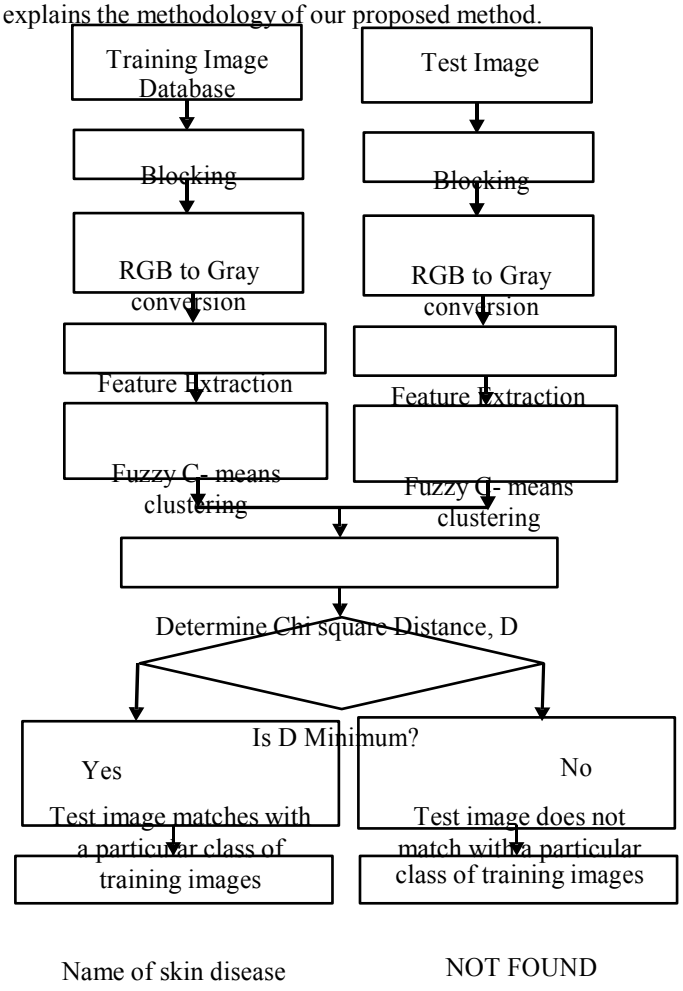


Figure 2: Flow chart of proposed method

Then Fuzzy C-means clustering technique is used with texture features extracted by GLCM.

Computing GLCM of an image and extracting texture features of the image using GLCM are discussed in the section that follows.

GRAY LEVEL CO-OCCURRENCE MATRIX

Given an image composed of pixels each with an intensity (specific gray level), the GLCM is a tabulation of how often different combinations of gray levels co-occur in an image or image section. Texture feature calculations use the contents of the GLCM to give a measure of the variation in intensity at the pixel of interest [17]. GLCM of an image is computed using two parameters; they are Offset and Distance d between pixels. Here, offset represents the direction of pairing pixels. For example, with 0 offset, GLCM is computed by pairing pixels horizontally. Distance d represents spacing between pixels selected for pairing. For example, d = 1 represents two adjacent pixels.

GLCM of an Image

From image data matrix we compute GLCM by computing the frequency of each pair. GLCM with the parameters 0 offset and distance d = 1 is formed by computing the frequency of two adjacent pair of pixels with 0-offset.for

example, in the image data matrix, (0,0) pixel pair may occur once, (0,1) pixel pair may occur zero times, (1,1) pixel pair may occur twice and so on.

GLCM of the image with 45 offset and distance d = 1 is computed by considering the right diagonal elements. In the similar way, GLCM of the image with 90^0 and 135 offsets at distance d = 1 are computed considering the two adjacent elements in the vertically upward direction and the left diagonal direction respectively.

Let G_{0^o}, G_{45^o}, G_{90^o} and G_{135^o} represent GLCMs with 0 offset, 45 offset, 90 offset, and 135 offset respectively. To get the symmetrical GLCM, original GLCM is added to transposed GLCM [18,19]. That is, symmetrical GLCM is expressed as

$$S \; G \; G^T \qquad (1)$$

The symmetrical GLCM in (1) is normalized to make the maximum limit of its elements to Unity. Normalized symmetrical GLCM is obtained by dividing it with sum its elements. Let normalized symmetrical GLCMs are denoted by N_{0^o}, N_{45^o}, N_{90^o} and N_{135^o} for the offsets 0, 45, 90 and 135 respectively. Normalized symmetrical GLCMs are computed and the results are as given below.

$$N_{0^\circ} = \frac{1}{22} \begin{bmatrix} 2 & 0 & 0 & 1 \\ 0 & 4 & 3 & 0 \\ 0 & 3 & 4 & 0 \\ 1 & 0 & 0 & 4 \end{bmatrix} \qquad N_{45^\circ} = \frac{1}{18} \begin{bmatrix} 0 & 2 & 1 & 0 \\ 2 & 1 & 2 & 1 \\ 1 & 2 & 1 & 2 \\ 0 & 1 & 2 & 0 \end{bmatrix}$$

$$N_{90^\circ} = \frac{1}{24}\begin{bmatrix} 0 & 3 & 1 & 0 \\ 3 & 2 & 2 & 2 \\ 1 & 2 & 2 & 0 \\ 0 & 2 & 0 & 4 \end{bmatrix} \qquad N_{135^\circ} = \frac{1}{20}\begin{bmatrix} 0 & 2 & 0 & 1 \\ 2 & 0 & 2 & 1 \\ 0 & 2 & 2 & 2 \\ 1 & 1 & 2 & 2 \end{bmatrix}$$

Then, texture features like Correlation, Contrast, Energy and Homogeneity of the given image are extracted from normalized symmetrical GLCMs.

Texture Measures from GLCM

Most of the texture calculations are weighted averages of the normalized symmetrical GLCM cell contents [20].

Contrast (Con)

Contrast is a measure of intensity or gray-level variations between the reference pixel and its neighbor. In the visual perception of the real world, contrast is determined by the difference in the color and brightness of the object and other objects within the same field of view.

Contrast is expressed mathematically as

$$Con \sum_{i,j=0}^{K} p_{i,j}(i-j)^2 \qquad (2)$$

where $p_{i,j}$ represents element (i, j) of the normalized symmetrical GLCM, K represents number of gray levels in the image as specified by number of levels in under Quantization on the GLCM texture and $(i,- j)^2$ represents the weight. If i and j are equal, the cell is on the diagonal and $i - j = 0$. These values represent pixels entirely similar to their neighbor, so they are given a weight of 0. The weights continue to increase exponentially as $(i - j)$ increases. A low contrast image presents GLCM high around the principal diagonal and features low spatial frequencies.

Correlation (Cor)

Correlation feature shows the linear dependency of gray level values in the co-occurrence matrix. It presents how a reference pixel is related to its neighbor. It has values from -1 to 1.-1 refers to uncorrelated, 0 is uncorrelated, 1 is perfectly correlated.

It simply returns a measure of how correlated a pixel is to its neighbor over the whole image.

$$Cor \sum_{i,j=0}^{K} \frac{(i-\mu_i)(j-\mu_j)p_{i,j}}{\sigma_i \sigma_j} \qquad (3)$$

where $\mu_{i,j}$ and $\sigma_{i,j}$ represent the means and standard deviations of $p_{i,j}$ respectively. is the GLCM mean (being an estimate of the intensity of all pixels in the relationships that contributed to the GLCM) and is given by

$$\sum_{i,j=0}^{N-1} i\, p_{i,j} \qquad (4)$$

and σ^2 is the variance of the intensities of all reference pixels in the relationships that contributed to the GLCM, calculated as

$$\sigma^2 \sum_{i,j=0}^{N-1} p_{i,j}(i-\mu)^2 \qquad (5)$$

3.2.3 Homogeneity (HOM)

Dissimilarity and Contrast result in larger numbers for more contrast windows. If weights decrease away from the diagonal, the result will be larger for windows with low contrast.

As shown in (6), homogeneity weight is the inverse of the Contrast weight. Therefore, with homogeneity, weights decrease exponentially away from the diagonal and return a value that measures the closeness of the distribution of elements in the GLCM to the GLCM diagonal.

$$H \sum_{i,j=1}^{k} \frac{p_{i,j}}{1+|i-j|} \qquad (6)$$

3.2.4 Energy

Energy, also known as Uniformity, measures the textural uniformity that is pixel pair repetitions. It detects disorders in textures. Energy uses each $p_{i,j}$ as a weight for itself. It simply returns the sum of squared elements in the GLCM.

$$E \sum_{i,j=1}^{K} p_{i,j}^2 \qquad (7)$$

High values of Energy occur when the window is very orderly. Energy reaches a maximum value equal to one. High energy values occur when the gray level distribution has a constant or periodic form. Energy has a normalized range.

GLCM Based Classification

Let N training images be denoted by T_1 , T_2 , T_3 , … , T_N . Normalized symmetrical GLCMs for four offsets 0, 45, 90, and 135 respectively are computed for each image. Let $N_{0^\circ}, N_{45^\circ}, N_{90^\circ}, N_{135^\circ}{}_{T_n}$ represent set of normalized symmetrical GLCMs for the training image T_n where n ranges from 1 to N. Then, for each set of normalized symmetrical GLCMs, four texture features Correlation (Cor), Contrast (Con), Homogeneity (H) and Energy (E) using equations (2) to (7). Let these features for image T_n be denoted as below.

$$Cor_{0^\circ}, Cor_{45^\circ}, Cor_{90^\circ}, Cor_{135^\circ}{}_{T_n};$$

$$Con_{0^\circ}, Con_{45^\circ}, Con_{90^\circ}, Con_{135^\circ}{}_{T_n};$$

$$H_{0^\circ}, H_{45^\circ}, H_{90^\circ}, H_{135^\circ}{}_{T_n};$$

$$E_{0^\circ}, E_{45^\circ}, E_{90^\circ}, E_{135^\circ}{}_{T_n};$$

Final values of texture features are obtained by taking average of individual feature set of each image [18]. Let Cor, Con, H, E_{T_n} represent final values of texture features of training image T_n; $n = 1, 2,\ldots,N$ and are given by

$$Cor = \frac{Cor_0 + Cor_{45} + Cor_{90} + Cor_{135}}{4};$$

$$Con = \frac{Con_0 + Con_{45} + Con_{90} + Con_{135}}{4};$$

$$H = \frac{H_0 + H_{45} + H_{90} + H_{135}}{4};$$

$$E = \frac{E_0 + E_{45} + E_{90} + E_{135}}{4};$$

In similar way average texture feature values of Test image are computed. Then the similarity between train and test images is checked by measuring Chi square distance and skin disease is classified.

Experimentation on Clustering Using GLCM

Skin images are collected from the DERMNET [2]. These image collections are used for experimentations. The selected image data set consists of 210 images and these images belong to three classes each with 70 images. The disease classes considered for classification are as follows:

1. Class 1: Angioedema - swelling under the skin.
2. Class 2: Actinic Keratosis - caused by the damage from UV rays.
3. Class 3: Eczema - an inflammation of the epidermis with symptoms like itching, dryness, blistering.

The images are scaled to standard dimensions of 100*100 and stored in JPEG format. The disadvantage of GLCM based image classification is that we must compare the test image with each and every training image of every class. Hence, it takes longer time in recognizing the skin disease.

Now let us consider Fuzzy Logic. In Fuzzy Logic, we will group the images of each class, which is called "Clustering". Using Fuzzy clustering technique, we get final texture features of each class. Therefore, the texture features of the test image are compared with the final texture features of each training class. Time consumed here is equal to the time taken for comparing with three training images (as there are three training classes). In this way we can overcome the above mentioned disadvantage linked with GLCM based image classification.

Introduction and implementation of Fuzzy Logic is discussed in the section that follows.

FUZZY CLUSTERING

Many clustering algorithms have been introduced in the literature. Since clusters can formally be seen as subsets of the data set, one possible classification of clustering methods can be based on the subsets whether they are fuzzy (soft) or crisp (hard).

Hard clustering methods are based on classical set theory, and require that an object either does or does not belong to a cluster. Hard clustering means partitioning the data into a specified number of mutually exclusive subsets. Fuzzy clustering (also referred to as soft clustering) methods, However, allow the objects to belong to several clusters Simultaneously, with different degrees of membership. In many situations, fuzzy clustering is more natural than hard clustering. Objects on the boundaries between several classes are not forced to fully belong to one of the classes, but rather are assigned membership degrees between 0 and 1 indicating their partial membership.

In fuzzy clustering, data elements can belong to more than one cluster, and associated with each element is a set of membership levels. These indicate the strength of the association between that data element and a particular cluster. Fuzzy clustering is a process of assigning these membership levels, and then using them to assign data elements to one or more clusters.

In the present skin disease identification problem, image classification is to be carried on the test image to identify to which one of the three classes it belongs. As hundreds of images are present as train and test images and as time consumed by GLCM based image classification is impracticable, we considered Fuzzy C-Means clustering along with GLCM. The details of Fuzzy C-Means clustering are discussed the section that follows.

Fuzzy C-Means Clustering

Fuzzy C-Means (FCM) clustering is an iterative process. First, the initial fuzzy partition matrix is generated and the initial fuzzy cluster centers are calculated. In each step of the iteration, the cluster centers and the membership grade point are updated and the objective function is minimized to find the best location for the clusters. The process stops when the maximum number of iterations is reached, or when the objective function improvement between two consecutive iterations is less than the minimum amount of improvement specified.

Before using the FCM algorithm, the following parameters must be specified: the number of clusters *c*, the 'fuzziness' exponent *m*, the termination tolerance , and the norm- inducing matrix A. Moreover, the fuzzy partition matrix, U, must be initialized. The choices for these parameters are described one by one in [18].

FCM exhibits several advantages. It yields regions more homogeneous than those of other methods. It reduces the spurious blobs and it is less sensitive to noise than other techniques. This technique is a powerful method for noisy image segmentation and works for both single and multiple-feature data with spatial information [19].

Experimentation on Clustering using FCM

The dataset samples of skin diseases belonging to Angioedema, Actinic Keratosis, and Eczema are taken. For each of the species, observations for contrast, correlation, homogeneity and energy are recorded. The dataset is partitioned into three groups named Angioedema, keratosis, and eczema. The data to be clustered is 4-dimensional data and represents contrast, correlation, homogeneity and energy. From each of the three groups (Angiodema, keratosis and

eczema), two characteristics (for example, contrast vs. homogeneity) of the skin diseases are plotted in a 2-dimensional plot. The parameters required for Fuzzy C-Means clustering such as number of clusters, exponent for maximum number of iterations and minimum improvement are defined and set.

The algorithms and simulation results of GLCM and FCM in application to skin disease identification are discussed in the next section

SIMULATION RESULTS

Images were obtained from www.dermnet.com. Dermnet is the largest independent photo dermatology source dedicated to online medical education though articles, photos and video. Dermnet provides information on a wide variety of skin conditions through innovative media. From the wide availability, only three categories were considered for the skin disease classification system.

The three categories considered are Angiodema, Keratosis and Eczema. These images are divided into training and testing set, where 70 images from each group are used to train the system and the remaining 60 images in each group serve as the testing set. The parameters considered in this project include Contrast, Correlation, Homogeneity and Energy.

Skin disease detection using GLCM

The skin disease detection using GLCM involves the following steps:

1. RGB to Gray scale conversion
2. Evaluation of GLCM
3. Feature Extraction
4. Identification of class of skin disease

Step 1: RGB to Gray scale conversion

Here the test image is colour image; it is converted into a gray scale image because the gray level co-occurrence matrix uses a gray scale image.

Step2: Evaluation of GLCM

After the conversion, GLCM is calculated from the gray scale images of Angiodema, Keratosis and Eczema. GLCM for the image of Angiodema is shown in Figure 3.

Step 3: Feature Extraction

After obtaining the GLCM of an image, the features (contrast, correlation, homogeneity and energy) are calculated using equations (2) to (7). As shown in Table 1, features of GLCM for three skin diseases under consideration are computed for four different offsets.

The average of four offset values of each parameter of GLCM comes out to be as shown in Table 2.

Step 4: Detection of the class of the Skin Disease

Now a single test image is considered and its values are compared with the all the training images in three classes using the formula of Chi Square distance,

$$d \sum_{n=1}^{N} \frac{(S_n - M_n)^2}{S_n + M_n} \quad (8)$$

where S_n, M_n and N represent Training Image, Test Image and number of test images respectively. After obtaining the distances see for the minimum value. The comparison with the class whose distance comes out to be minimum, is said to be the class the test image belongs to. Simulation results are tabulated in Table 3.

	1	2	3	4	5	6	7	8
1	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0
3	0	0	2	6	2	0	0	0
4	0	0	7	1551	462	3	0	0
5	0	0	1	463	3702	632	2	0
6	0	0	1	3	634	2327	27	0
7	0	0	0	0	1	29	3	0
8	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**

Figure 3. Test Image (Angiodema) and its GLCM

Table 1: Features of images with four different offsets

Skin disease	Prop-erties	Offsets			
		0	**45**	**90**	**135**
Angio-dema	CON	0.2207	0.2520	0.1982	0.2622
	COR	0.7853	0.7549	0.8081	0.7452
	H	0.8903	0.8753	0.9014	0.8717
	E	0.2400	0.2262	0.2496	0.2234
Keratosis	CON	0.5734	0.7014	0.5581	0.6614
	COR	0.5495	0.4460	0.5598	0.4476
	H	0.7626	0.7349	0.7669	0.7428
	E	0.1558	0.1467	0.1578	0.1493
Eczema	CON	0.1607	0.2124	0.1737	0.2235
	COR	0.8064	0.7455	0.7904	0.7311
	H	0.9196	0.8938	0.9131	0.8882
	E	0.3033	0.2737	0.2952	0.2686

Table 2.:Average values of properties of different classes

Property / Class	**CON**	**COR**	**H**	**E**
Angiodema	0.2333	0.7734	0.8834	0.2348
Keratosis	0.6236	0.5082	0.7518	0.1524
Eczema	0.1926	0.7681	0.9037	0.2852

Table 3. Identifying the skin disease using Chi square distance

Class of Test Image	Chi square distance, D from			Skin disease with D_{min}
	Angio-dema	Keratosis	Eczema	
Angio-dema	0.25	14.85	0.26	Angio-dema
Keratosis	0.55	0.49	0.73	Keratosis
Eczema	0.35	0.25	0.09	Eczema

All the above mentioned steps (1 to 4) are applied for each and every image whether its training image or test image but using step 5 the class of the test image is identified taking the values of the training image as reference.

Fuzzy Logic and GLCM based Skin Disease Detection

The skin disease identification using GLCM and fuzzy logic involves the following steps:

1. RGB to Gray scale conversion
2. Evaluation of GLCM
3. Feature Extraction
4. Implementation of fuzzy logic
5. Identifying the skin disease

The above mentioned step 1 to step 3 are implemented on all the training images, from which we get all the values of the properties of each and every training image. Then the fuzzy logic (C means clustering) is implemented on those values of the training image set.

Fuzzy C-Means clustering is an iterative process. First, the initial fuzzy partition matrix is generated and the initial fuzzy cluster centers are calculated. In each step of the iteration, the
cluster centers and the membership grade point are updated and the objective function is minimized to find the best location for the clusters. The process stops when the maximum number of iterations is reached, or when the objective function improvement between two consecutive iterations is less than the minimum amount of improvement specified.

Implementation of fuzzy logic

Fuzzy c-means was implemented using a degree of fuzziness k = 2.The maximum number of iterations was set to 100. Now consider the training set of angiodema and fuzzy logic is implemented on it which gives the result as follows: In the below figure, 1(bolded) represents that the clustering is done for angiodema.

From the figure 4, it can be observed that we get a single value for each property. Expanded view of a part of Figure 4 is shown in Figure 5.

From Figure 5(a), values of Contrast and Correlation are observed as Contrast = 0.35 and Correlation = 0.46. From Figure 5(b), values of Homogeneity and Energy are observed as Homogeneity = 0.84 and Energy = 0.3. In the similar way, the single value for each property of Keratosis and Eczema are obtained and these values are tabulated below.

Identification of Skin Disease

After the clustering technique, these single values of the training image obtained are considered as the reference and the Chi square distance is calculated between the reference and each of the test images and the reference image to which test image has minimum Chi square distance is declared as the class of test image.

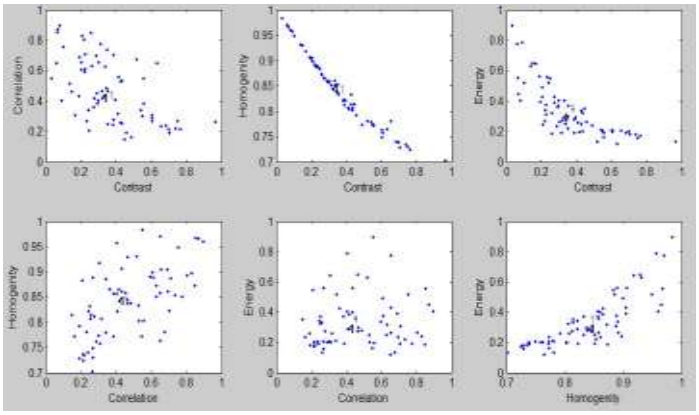
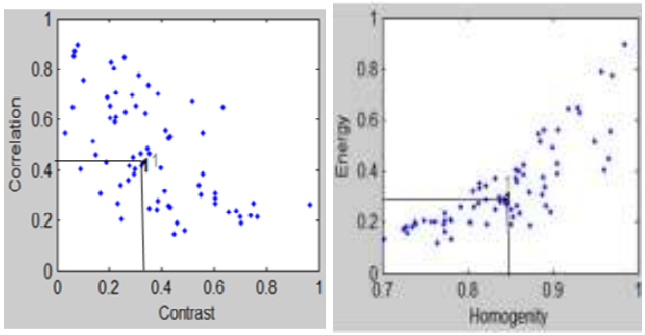


Figure 4: Fuzzy C means clustering on Angiodema



(a) (b)

Figure 5: Expanded view of (a) Correlation versus Contrast and (b) Energy versus Homogeneity parts of FCM of angiodema

Table 4: Properties of Keratosis and Eczema after clustering

Skin disease	Contrast	Correlation	Homogeneity	Energy
Keratosis	0.4	0.58	0.85	0.34
Eczema	0.25	0.57	0.86	0.36

Performance Measures

A confusion matrix [21] is used for computing performance measures of classification such as accuracy, sensitivity, specificity, positive predictive value and negative predictive value. As shown in Table 5, a confusion matrix contains information about actual and predicted classifications done by a classification system.

Table 5: Confusion matrix

	Predicted	
Actual	Positive	Negative
Positive	True Positive (TP)	False Negative (FN)
Negative	False Positive (FP)	True Negative (TN)

Accuracy: It indicates the total success of both positive and negative cases.

$$\text{Accuracy (\%)} \frac{TP\ TN}{TP\ FP\ FN\ TN} 100 \quad (9)$$

Sensitivity: It is the ability of the test to correctly identify the disease which belongs to the one under test (true positive rate).

$$\text{Sensitivity (\%)} \frac{TP}{TP\ FN} 100 \quad (10)$$

Specificity: It is the ability of the test to correctly identify those diseases which do not belong to the one under test (true negative rate).

$$\text{Specificity (\%)} \frac{TN}{FP\ TN} 100 \quad (11)$$

Positive Predictive Value (PPV): It is the proportion of positive test results that are true positives.

$$\text{PPV (\%)} \frac{TP}{TP\ FP} 100 \quad (12)$$

Negative Predictive Value (NPV): It is defined as the proportion of diseases with a negative test result which are correctly classified.

$$\text{NPV (\%)} \frac{TN}{FN\ TN} 100 \quad (13)$$

Performance measures of GLCM and GLCM with FCM

Performance of GLCM and GLCM with FCM classification methods is compared in application to skin disease identification. 70 images from each category of the skin diseases Angiodema, Keratosis and Eczema are used for training and 50 images from each category are used for testing. From the simulation results of classification methods GLCM and GLCM with FCM, the confusion matrices are developed and are shown Table 6. TP, FN, FP and TN for all the four parameters Contrast, Homogeneity, Correlation of GLCM are computed and the average values are placed in the Table 6.

Correct identification of skin disease of images that belong to particular skin disease category (Angiodema, Keratosis or Eczema) and also correct identification of skin diseases that do not belong to that particular skin disease category is given by Accuracy and it is 86.23% for classification method GLCM with FCM as compared to GLCM method which has accuracy of 66.13%. All the performance measures of GLCM with FCM method are more than 84.74%; it represents its superior

Table 6: Confusion Matrix for skin disease diagnosis

Confusion Matrix	GLCM	GLCM with FCM
True Positive (%)	64.33	84.08
False Negative (%)	35.67	15.92
False Positive (%)	25.13	11.61
True Negative (%)	74.87	88.38

Best recognition capability of GLCM with FCM classification method is 84.08% as compared to 64.33% of GLCM classification method. GLCM with FCM classification method has higher TP and TN values as compared to GLCM based classification method i.e., success rate of GLCM with FCM method in correct diagnosis of skin disease is high. The performance measures accuracy, sensitivity, specificity, PPV and NPV of both classification methods are computed using equations (9) to (13) respectively and results are tabulated in Table 7.

Table 7: Performance Measures

Performance measures	GLCM	GLCM with FCM
Accuracy (%)	66.13	86.23
Sensitivity (%)	51.00	84.08
Specificity (%)	81.25	88.38
PPV (%)	73.12	87.86

CONCLUSION

In this paper, classification methods GLCM and FCM clustering are used for skin disease classification. We used GLCM for feature extraction and fuzzy C means clustering is an efficient cluster analysis in feature space. Three skin disease classes Angioedema, Actinic Keratosis and Eczema are considered for skin disease classification. Texture features are extracted using normalized symmetrical GLCMs. FCM clustering is applied on the texture features of three classes of training data sets. Then Chi. Square distance is used for image classification. Through simulation results, it is shown that GLCM with FCM clustering is more efficient than GLCM method in skin disease classification. Proposed method reduces the number of computations drastically as compared to GLCM method in image classification. Therefore fast Skin disease classification is possible with GLCM with FCM clustering method.

Accuracy of the classification using GLCM and neural networks reported in [12] using Manhattan distance is 82.22%. The results of FCM clustering and pattern recognition method of image classification reported in [21] revealed that the accuracy of the system is 68.04% and its PPV is 80.88%. In comparison to these methods, the proposed method exhibits superior performance with an accuracy of 86.23% and PPV of
87.86%.

ACKNOWLEDGMENT

Authors express their deep sense of gratitude to the management of Shri Vishnu Engineering College for Women for encouraging them in research work.

REFERENCES

Ankit Ankit Goyal,2014. “Around 19 Crore Indians Likely to Suffer from Skin Diseases by 2015 – notes Frost & Sullivan”, article.

www.dermnet.com.

Damilola et al.,2013. "Automating skin disease diagnosis using Image classification", WCECS,23-25.

S. Arivazhagan, R. Newlin Shebiah, K. Divya, and M.P. Subadevi,2012."Skin Disease Classification by Etracting Independent Components", J. of Emerging Trends in Comp. and Info. Sciences,**3:**10.

A.Sparavigna, R.Marazzato,2010. "An image processing Analysis of skin textures", Skin research and technology,**16-2**:161-167.

Sheha, M.A., Mabrouk, M.S., and Sharawy, A.,2012. "Automatic Detection of Melanoma Skin Cancer using Texture Analysis", Intl. J. of Computer Applications (0975 – 8887),**42-20**

Anand. H. Kulkarni and Ashwin Patil R. K.,2012. "Applying image processing technique to detect plant diseases", Intl. J. of Modern Engg. Research,**2-5:**3661-3664.

Jagadeesh Devdas Pujari, Rajesh Yakkundimath and Abdulmunaf Syedhusain Byadgi,2013 "Grading and Classification of Anthracnose Fungal Disease of Fruits based on Statistical Texture Features", Intl. J. of Advanced Science and Tech., Vol. 52.

Al. Abadi, N. K., Dahir, N. S., and Alkareem, Z. A.,2-2013. "Skin texture recognition using neural network", Proc. of the Intl. Arab Conf. on Info. Tech., Tunisia, December 16-18:1-4.

B.V.Dhandra, Shridevi Soma, Shweta Reddy and Gururaj Mukarambi, 2013."Color Histogram Approach for Analysis of Psoriasis Skin Disease", Asso. of Comp. Electronics and Electrical Engineers.

Oberti R, Marchi M, Tirelli P, Calcante A, Iriti M, Borghese A.N.,014 "Automatic detection of powdery mildew on grapevine leaves by image analysis: Optimal view-angle range to increase the sensitivity", Comput Electron Agric **104:**1-8.

Anal Kumar Mittra and Ranjan Parekh,2011 "Automated Detection of Skin Diseases Using Texture Features", Intl. J. of Engg. Science and Technology, Vol. 3.

Metty Mustikasari and Sarifuddin Madenda,2015. "Texture based Image Retrieval Using GLCM and Image Sub-Block", Intl. J. of Advanced Research in Computer Science and Software Engg, **5:3**

Manish Pawar, Dipesh Kumar Sharma and R.N. Giri, 2014."Multiclass Skin Disease Classification Using Neural Network", Intl. J. of Comp. Science and Info. Tech. Research, **2-4:**189-193.

Sukhjinder Kaur,2015 "Comparative Study Review on Lung Cancer Detection Using Neural Network and Clustering Algorithm", Intl. J. of Advanced Research in Electronics and Commun. Engg., **4:2.**

Sueli A. Mingoti and Joab O. Lima,2006. "Comparing SOM neural network with Fuzzy c-means, K-means and traditional hierarchical clustering algorithms", European Journal of Operational Research **174**, 1742–1759..

R.M. Haralick, ''Statistical and Structural Approaches to Texture," Proc. IEEE, vol. 67, no. 5, 1979, pp. 786_804.

Jing Zhang, Gui-li Li, Seok-wum He, 2008. "Texture-Based image Retrieval. By Edge Detection Matching GLCM", in Proc. of 10th Int. conference on High Performance Computing and Communications, Sept.

Ranjan Parekh, "Using Texture Analysis for Medical Diagnosis", Media in medicine, IEEE Computer Society,

Pushpa Kotipalli, Pradeep Mullangi, and Y. Srinivasa Rao, 2015."Texture and Clustering-based Skin Disease Classification", CSI..

Indira Rustempasic and Mehmet Can,2013 " Diagnosis of Parkinson's Disease using Fuzzy C-Means Clustering and Pattern Recognition", Southeast Europe J. of Soft Computing.