Diabetes Mellitus Detection Based on Facial Texture Feature Using the GLCM

Pavana S[1], Dr Shailaja K[2]

[1] *MTech Student , Biomedical Signal Processing and Instrumentation , Dept. of IT, SJCE, Mysuru, Karnataka, India*
[2]*Associate Professor, Dept. of IT, SJCE, Mysuru, Karnataka, India*

---***---

Abstract - *Researchers have recently discovered that Diabetes Mellitus can be detected through non-invasive computerized method. However, the focus has been on facial texture features. In this paper, we extensively study the effects of texture features extracted from facial regions at detecting Diabetes Mellitus(DM) using Gray Level Co-occurrence Matrix(GLCM). Texture feature extracted here are contrast, correlation, energy, homogeneity and Haralick' s features. Afterwards, Support Vector Machine (SVM) classifier is applied for classification. Experimental results on a data set of 40 facial images where 20 are DM facial images and rest 20 are of Healthy facial images are used to show that the proposed method can distinguish between Diabetes Mellitus and Healthy samples. The performance of the classifier is evaluated with an accuracy , a sensitivity , and a specificity , using a combination of facial images.*

***Key Words*: Diabetes mellitus(DM), Gray level co-occurence matrix(GLCM), Facial texture features, Haralick's features, Support Vector Machine(SVM) Performance evaluation.**

1. INTRODUCTION

Diabetes is a complex group of diseases with a variety of causes. People with diabetes have high blood glucose, also called high blood sugar or hyperglycemia. Diabetes is a disease in which the body is unable to properly use and store glucose. Diabetes Mellitus (DM) can be detected in a non-invasive manner through the analysis of human facial textural features. Non-invasive means that the body is not invaded or cut open as during surgical investigations or therapeutic surgery[9].Until the last several decades, exploratory surgery was routinely performed when a patient was critically ill and the source of illness was not known. In dire cases, the patient's thorax, for instance, was surgically opened and examined to try to determine the source of illness. Diagnostic imaging was first performed in 1895 with the discovery of the x-ray. For the first time, physicians could see inside the body without having to perform exploratory surgery. Thus diagnostic imaging is a "non-invasive" way to look at internal organs and structures. The traditional way for diagnosis of DM is through a Fasting Plasma Glucose (FPG) which takes a small blood sample from each patient. Before this test, the patient cannot eat anything for at least 12 hours. Therefore, this test is considered to be painful, invasive, time consuming, and as well as bringing discomfort [5].

In this proposed work, to implement a new non-invasive system to detect the health status (Healthy or Diseased) of an individual based on facial texture features extracted using the GLCM. Here we use Gray Level Co-occurrence Matrix (GLCM) to extract statistical texture features and Haralick's features. Afterwards, Support Vector Machine (SVM) classifier is applied for classification. The performance of the classifier is evaluated with an accuracy , a sensitivity , and a specificity , using a combination of facial images.

2. METHODOLOGY

This work presents the diabetes mellitus detection and classification using facial images. Important statistical features based on image features and facial texture features are extracted and classified to distinguish between diabetic and healthy person. In order to carry out this process initially diabetic and healthy facial images are collected from the AIIMS Hospital, New Delhi. Fig .1 shows the block diagram representation of diabetes mellitus detection and classification.

From the captured digital face images the statistical textural features are extracted by using gray level co- occurrence matrix(GLCM)and Haralick's features. The extracted features are then given to the classifier i.e support vector machine (SVM) which distinguishes between diabetic and healthy person.

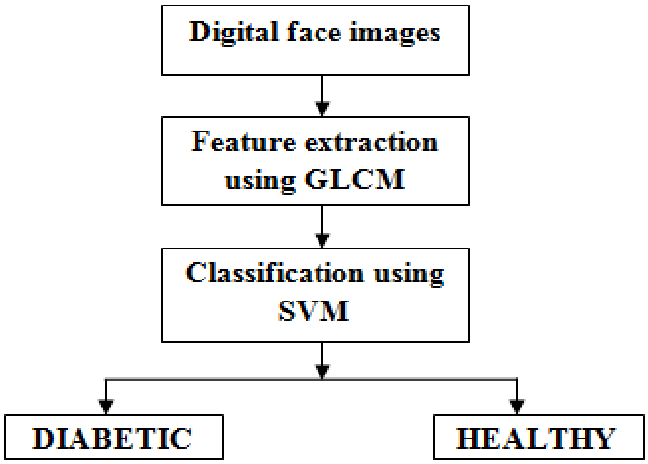

Fig .1 Block diagram representation of diabetes mellitus detection and classification

2.1 FEATURE EXTRACTION

Feature extraction is a method of capturing visual content of images for indexing and retrieval. Feature extraction is used to denote a piece of information which is relevant for solving the computational task related to a certain application There are two types of texture features measure. They are first order and second order. In the first order, texture measures are statistics calculated from an individual pixel and do not consider pixel neighbor relationships. In the second order, measures consider the relationship between neighbor relationships. Here, statistical features are extracted from GLCM.

2.1.1 Gray Level Co-occurrence Matrix

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. The number of gray levels in the image determines the size of the GLCM. The gray-level co-occurrence matrix can reveal certain properties about the spatial distribution of the gray levels in the texture image. For example, if most of the entries in the GLCM are concentrated along the diagonal, the texture is coarse with respect to the specified offset[13,14]. The graycomatrix function creates a gray-level co-occurrence matrix (GLCM) by calculating how often a pixel with the intensity (gray-level) value i occurs in a specific spatial relationship to a pixel with the value j. By default, the spatial relationship is defined as the pixel of interest and the pixel to its immediate right (horizontally adjacent).

The second order statistics of an image can be obtained from GLCM which accounts for the spatial inter-dependency or co-occurrence of two pixels at specific relative positions. Co-occurrence matrices are calculated for the directions of θ (0^0, 45^0, 90^0, 135^0).

For each matrix, the thirteen features like, Contrast, Correlation, Energy, Homogeneity, Mean, Standard deviation, Entropy, Variance, Smoothness, Kurtosis, Skweness, Inverse Difference Moment are obtained for the synthesized facial image.
The features extracted are:

Contrast: Contrast returns a measure of the intensity contrast between a pixel and its neighborhood. The contrast feature is a measure of the amount of local variations present in an image.

$$\text{Contrast} = \sum_{i,j} f(i,j)(i-j)^2 \quad (1)$$

Correlation: Correlation measures how correlated a pixel is to its neighborhood. It could also be described as a measure of linear dependencies among neighboring pixels in an image.

$$\text{Correlation}=\sum_{i,j} f(i,j)(i-\mu_i)(j-\mu_j)/\sigma_i \ \sigma_j \quad (2)$$

Energy: Energy can be defined as the quantifiable amount of the extent of pixel pair repetitions. Energy is a parameter to measure the similarity of an image. It is also referred to as angular second moment, and it is defined as.

$$\text{Energy}=\sum_{i,j} f(i,j)^2 \quad (3)$$

Homogeneity: Homogeneity measures the similarity of image pixels. It is inversely proportional to the contrast.

$$\text{Homogeneity}=\sum_{i,j} f(i,j)/(1+|i-j|) \quad (4)$$

Mean: The mean of an image is calculated by adding all the pixel values of an image divided by the total number of pixels in an image.

$$\text{Mean}(\mu) = \sum_{i,j} f(i,j)/N \quad (5)$$

Standard Deviation: The standard deviation is the second central moment describing probability distribution of an observed population and can serve as a measure of in homogeneity. A higher value indicates better intensity level and high contrast of edges of an image.

$$\text{SD}(\sigma) = \sum_{i,j}\left(\frac{(f(i,j)-\mu)^2}{N}\right)^{1/2} \quad (6)$$

Entropy: Entropy is a measure of randomness of intensity image.

$$\text{Entropy}=-\sum_{i,j} f(i,j)\log\left(f(i,j)\right) \quad (7)$$

Variance: Variance is a measure that is similar to the first order statistical variables called standard deviation. Variance is the square of standard deviation

$$\text{Variance} = (\sigma)^2 \qquad (8)$$

Kurtosis: The shape of a random variable's probability distribution is described by the parameter called Kurtosis. For the random variable X, the Kurtosis is defined as,

$$\text{Kurtosis}=\sum_{i,j}\left((i-\mu_i)+(j-\mu_j)\right)^4 f(i,j) \qquad (9)$$

Skewness: Skewness is a measure of the asymmetry of the data around the sample mean.

$$\text{Skewness}=\sum_{i,j}\left((i-\mu_i)+(j-\mu_j)\right)^3 f(i,j) \qquad (10)$$

Inverse Different Moment: Inverse Difference Moment (IDM) is a measure of image texture.

$$\text{IDM} = \sum_{i,j} f(i,j)/|i-j|^k \qquad (11)$$

2.2 CLASSIFICATION

It is well known that classifiers are machine learning tools which are able to take data items and place them into different classes. After the feature extraction process, the extracted features are directly applied to the classifiers, for classification into two different classes i.e diabetic and healthy.

2.2.1 Support Vector Machine Classifier

Support vector machine is a machine learning method that is widely used for data analyzing and pattern recognizing. A support vector machine constructs a hyper plane or set of hyper planes in a high- or infinite-dimensional space, which can be used for classification, regression, or other tasks. Intuitively, a good separation is achieved by the hyper plane that has the largest distance to the nearest training-data point of any class (so-called functional margin), since in general the larger the margin the lower the generalization error of the classifier[1,3].

SVM is not optimal for hyper plane construction in the input space but rather in high dimensional so called feature space Z. Using a kernel function to substitute the dot product of data points can construct an optimal separating hyper plane in a higher dimensional space. So, the decision function is formulated as

$$f(x)=\text{sign}\left(\sum_{i=1}^{m} \propto_i y_i\left(x_i,x_j\right)+b\right), \qquad (12)$$

where k is kernel function.

3. PERFORMANCE EVALUATION

The performance of classifier depends on various factors namely accuracy, sensitivity and specificity. Here accuracy is calculated by considering the following equations where TP and TN are the number of samples which are correctly identified as diabetic and healthy by the classifier in the test set , respectively and FN and FP represent the numbers of samples corresponding to those cases as they are mistakenly classified as diabetic and healthy respectively.
Various performance measures like sensitivity, specificity and accuracy are calculated using the matrix shown in fig 2.

Actual Vs. Predicted	Positive	Negative
Positive	TP	FN
Negative	FP	TN

Fig 2 Confusion matrix

Accuracy=(TP+TN)/(TP+FP+TN+FN) (13)
Sensitivity =TP/ (TP+FN) (14)
Specificity =TN/ (TN +FP) (15)

4. RESULTS

Diabetes detection process is carried out using facial images of diabetic people and healthy people. The accuracy of the classifier is then evaluated using performance evaluation .

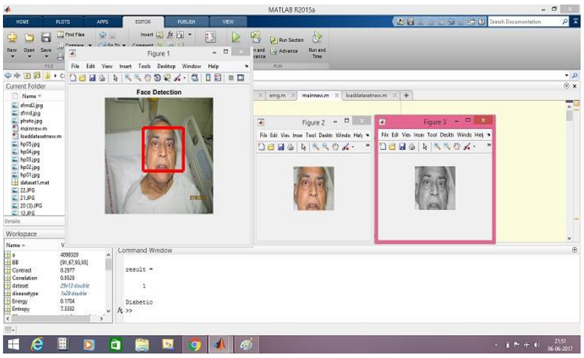

Fig .3 Result of a diabetic patient

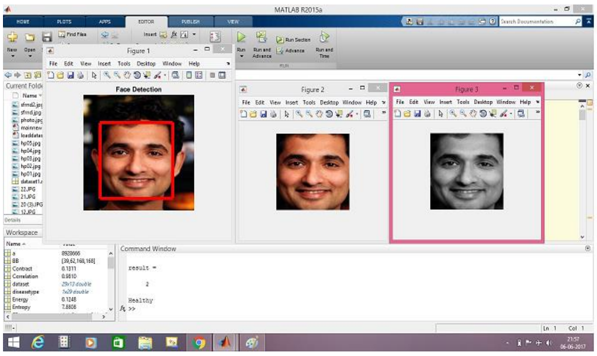

Fig.4 Result of a healthy person

In this work 70:30, ratio is taken for training and testing.The details of the diabetic and healthy dataset according to the ratio mentioned above is shown in the table 1. The facial image dataset includes 20 diabetic images out of which 14 images are given for training and rest 6 for testing and 20 healthy images are included in the image dataset, where 14 are given for training and 6 are given for testing. And output snapshots of diabetic and healthy facial images are as shown in Fig 3 and Fig 4.

Table.1 Ratio of images used for training and testing

Class	Diabetic	Healthy	Total
Training Set (70%)	14	14	28
Testing Set (30%)	6	6	12
Total	20	20	40

Table.2 Confusion matrix for diabetic and healthy people

TERMS	SVM CLASSIFIER
TRUE POSITIVE	6
TRUE NEGATIVE	5
FALSE POSITIVE	1
FALSE NEGATIVE	0

Table .3 Performance of the classifier

METRICS	SVM CLASSIFIER
ACCURACY	**91.67%**
SENSITIVITY	**100%**
SPECIFICITY	**83.33%**

5. CONCLUSION

Diabetes mellitus detection is based on facial textural characteristics of facial images where different statistical features are extracted based on GLCM. Features extracted were fed as input to SVM classifiers which helps to distinguish the facial images as diabetic and healthy. Classification result obtained from SVM classifier are 91.67% of Accuracy, Sensitivity of 100% and Specificity of 83.33% in classifying the test images.

Future work can include vast dataset of diabetic and healthy person images which can increase the performance of the classifier. To further improve overall performance of this algorithm, some more discriminate features and some advanced machine learning methods can be used .

REFERENCES

[1] Carlos M. Travieso, Jestis B. Alonso. Miguel A. Ferrer “Facial Identification Using Transformed Domain By SVM” IEEE 2004.

[2] Ahmad Fadzil M.H. and Abu Bakar H “Human face recognition using neural networks ” IEEE 1994.

[3] Minh Hoai Nguyen, Joan Perez and Fernando De la Torre “Facial Feature Detection with Optimal Pixel Reduction SVM” IEEE 2008.

[4] Jianke Li, Baojun Zhao, Hui Zhang and Jichao Jiao “Face Recognition System Using SVM Classifier and Feature Extraction by PCA and LDA Combination”, IEEE trans, 2009.

[5] Nichols, G. A., Hillier, T. A., and Brown, J. B., Normal fasting plasma glucose and risk of type 2 diabetes diagnosis. Am J Med, 2008.

[6] John Wright, Allen Y. Yang, Arvind Ganesh, S. Shankar Sastry, and Yi Ma., “Robust Face Recognition via Sparse Representation”2010.

[7] Bob Zhang, B. V. K. Vijaya Kumar, and David Zhang, “Detecting Diabetes Mellitus and Nonproliferative Diabetic Retinopathy Using Tongue Color, Texture, and Geometry Features”, IEEE transactions on biomedical engineering, 2014.

[8] Bob Zhang, B. V. K. Vijaya kumar, and David Zhang,, “Noninvasive Diabetes Mellitus Detection Using Facial Block Color With a Sparse Representation Classifier”,IEEE transactions on biomedical engineering, 2014.

[9] Ting Shu,Bob Zhang ‘Non-invasive Health Status Detection System Using Gabor Filters Based on Facial Block Texture Features’ Journal of Medical Systems, 2015.

[10] T. Shu, B. Zhang, and Yuan. Y. Tang, “Simplified and improved patch ordering for diabetes mellitus detection”, Cybernetics (CYBCONF), 2015 IEEE 2nd International Conference,2015.

[11] Anuradha S. Deshmukh, Kiran P. Lokhande“ Human Skin Texture Analysis Using GLCM Technique” IJEEE, 2016.

[12] Ting Shu and Bob Zhang “Facial Color Analysis of Overweight Obesity and its Relationship to Healthy and Diabetes Mellitus Using Statistical Pattern Recognition”IEEE 2015.

[13] Robert M. Haralick, Shangmugam K. and Dinstein L., “Textural Features for Image Classification” IEEE trans, 1973.

[14] P.Mohanaiah, P.Sathyanarayana,L .GuruKumar ”Image texture feature extraction using GLCM approach” International Journal of Scientific and Research Publications, 2013.