**NON-INVASIVE TECHNIQUE FOR DIABETES MELLITUS AND HIGH CHOLESTEROL DETECTION USING IRIS IMAGE.**

**D. Prashanth Kumar**

Assistant Professor

Computer Science and Engineering

Kamala Institute of Technology & Science Singapuram, Huzurabad, Karimnagar. prashanth.cse@kitssingapuram.ac.in

## Abstract:

**Iridology is a technique in science used to analyze color, patterns and various other properties of the iris to assess an individual’s general health[**1**]. Diabetes Mellitus is oneof the diseases which is hard to detect at early stages. It has no effective medication which cures it completely, so that the patient has to maintain a proper diet throughout his entire life. One of the main reasons for diabetes is lackof production of insulin by the pancreas. It also causes illness, urinary tract disease, immune deficiency, blindness. As the technology is advancing in the area of Biomedical and machine learning, the combination of both emerging areas is helpful for diabetes diagnosis wherein Iris scanning is one of the best non-invasivetechniques for the diagnosis of diabetes at preliminary stages. In this paper Iridology is used for prediction of diabetes with an image construction model using machine learning. This system consists of eye image acquisition instruments and image processing algorithms. The 3D- GLCM method is used for the feature extraction process to obtain texture characteristics in the image. Classification is carried out by training and testing on Dataset I and Dataset II. The method effectiveness is compared to the number of gray levels, namely 16, 32, 64, 128, and 256. Based on the five levels of gray, the best value is shown from the number of gray levels of 256, with the value obtained. These values are sensitivity value(0.9375), specificity value (0.0208), and accuracy (0.9844). The results showed that the higher the gray level of the image database used, the higher the sensitivity and accuracy values, while the lower the gray level indicates the specificity value.**

## Introduction

Lifestyle refers to the characteristics of the inhabitants of a region at a certain time and place. It includes dailybehavior and individual functioning in work, activity, pleasure, and diet. According to the WHO, lifestyle is 60% correlated with individual health and quality of life. Millions of people have practiced an unhealthy lifestyle. Thus, they suffer from a disease, disability, and even death. It causes metabolic disease, cardiovascular disease, hypertension, overweight, joint and bone problems, etc.

This problem is caused by an unhealthy lifestyle [2]. An unhealthy lifestyle can lead to reduced physical activity and increased obesity. It can cause DM and generally occurs in almost all countries [3]. The prevalence of DM has increased very rapidly in recent years. Now, DM is a global health problem. The International Diabetes Federation (IDF) stated that therewere 425 million people with DM in 2017, with a projected 48% to 629 million by 2045.

Currently, the measurement of blood sugar levels and cholesterol levels is done by invasive methods. Blood sampling in an invasive method causes discomfort because it causes injury and pain [4]. Another method is needed, namely a non-invasive method to reduce discomfort.

In this paper, an alternative measurement is given usingimage processing technology. Early detection of DM and HC is done by using the iris image. In iridology, alternative disease diagnoses using the iris pattern, tissue weakness, eye color, damage, and other characteristics can show the patient’s systemic health.

Figure 1 depicts the iris map or Iridology chart. The eyeiris can be used to diagnose diseases, where each part ofthe body represents a representation of the area found inthe iris.



Traces of records relating to the intensity levels or deviations of organs caused by the disease are systematically recorded and patterned on the eye iris and surroundings. It can be used as a practical guide for diagnosing various diseases. Figure 1 shows an iridology map, a map of the eye iris, which is a detailed picture of the body’s overall condition.

On the right side of the body his health condition is depicted in the right eye, and the health condition of the left body is depictedin the left iris. There is much research related to disease diagnosis based on the image that has been done. The research is related to the iris image, including detecting abnormal conditions of pancreatic Beta-cells as the cause of Diabetes Mellitus by segmentation using integral differential operator (IDO). IDO is based on segmentation results which are used to obtain the region of interest (ROI) at 06.45-07.15 o'clock position, feature extraction with gray level co- occurrence matrix (GLCM), and adaptive neighborhood-based modified backpropagation (ANBMP) classification. The accuracy of determining the ROI of the pancreas in thisstudy was 87.5%, and the accuracy of detection of pancreatic damage was 83.3% [5].This study could only detect patients with Type 1 DM. Type 1 diabetes, once known as juvenile diabetes or insulin-dependent diabetes, is a chronic condition. In this condition, the pancreas makes little or no insulin[6]. DM can be detected by improving the image using Adaptive Median Filter . Iris segmentation was done by determining the coordinates of the center of the pupil and the iris along with their radius. The segmentation results were normalized using the rubber sheet model (RSM) and classification using the Support Vector Machine (SVM) method. The result of diabetesdetection accuracy was 75%. In this study, the determination of ROI was based on the iridology map and only detected one disease.

HC poses a significant threat to a person’s health. Although it is not considered a disease, this condition can be a secondary disease and contribute to various forms of the disease, especially CVD [4]. The prevalence of hypercholesterolemia in the 25-34 year age group is 9.3%, and the 55-64 year age group increases with age up to 15.5% [6]. Currently, CVD is the leading cause of death globally. According to WHO, 31% of all deaths worldwide are caused by CVD disease, from 17.5 million people who die each year. Early detection is necessary to reduce mortality from both causes of the condition, namely DM and HC.

Research on disease detection based on the eye iris has been carried out only to detect one disease or one organ abnormality. In this study, two abnormalities were detected at once, namely DM and HC. Thus, the iris is detected from the input image whether it was suffering from DM, HC, DM, and HC, or in normal condition. The feature extraction method, 3D Gray Level Co-occurrence Matrix (3D GLCM), ensures iris detection

## Materials and Methods

This section described the identification of DM and

HC based on iris images using 3-Dimensional Gray Level Co-Occurrence Matrix (3D-GLCM). The system architecture is shown in Figure 2.



Fig. 2 Iris identification system architecture

This paper proposed a feature extraction approach using 3D-GLCM and similarity-based classification.

## Dataset

The samples of iris images from one hundred people where samples were taken as many as one hundred patients with specifications of 25 patients suffering from DM, 25 patients with high cholesterol, 25 patients suffering from DM and high cholesterol, and 25 patients with normal conditions (blood sugar and cholesterol levels were on a normal scale).The iris image captured by using a 12MP iridology camera eye Iriscope. The first procedure in taking the iris image is carried out on patients who are willing to participate in the study and meet the inclusion criteria that have been determined in the selection of subjects. The inclusion criteria of the subjects in this study were males or females, aged >= 20 years, who did not have cataracts, their irises were never injured, and the iris was not photographed postoperatively. Second, it was done after the patient does a blood test to check random blood sugar and cholesterol levels. The iris image was taken in the right eye and the left eye three times each.

## Pre-Processing

Pre-processing is the most important part of producinga higher level of accuracy. In the preprocessing step, two processes were carried out. The first process was to convert the RGB input image into a grayscale image.The second process was to improve the grayscale image using the Adaptive Histogram Equalization (AHE) method.

### RGB to Grayscale Conversion:

The RGB to Grayscale conversion process in this study used the rgb2gray() function in MATLAB, where the rgb2gray function converted the RGB image to grayscale by eliminating hue and saturation information while maintaining luminosity. NTSC standard conversion formula used to calculate the effective exposure of a pixel as in equation.



* + 1. ***Image Improvement with AHE Method:*** Adaptive Histogram Equalization (AHE) is an image processing technique that increases the contrast to obtain an image with an intensity value which will make the pixel point solid black at the darkest pixel point and each brilliant white at the brightest one. In medical imaging, AHE is used for contrast enhancement techniques and produces excellent images. In AHE, the histogram equalization process is carried out on each block (tile). The blocks (tiles) are generated by dividing the image by the size of n x n. Each block size varies in size and gives different results. Where between tiles on some pixels can overlap. Equation

(2) is the formula for calculating the gray value using the AHE method.



where:

Ci - cumulative of i grayscale value of the originalimage;

Round - rounding operation to the nearest integer; K0 - output value of adaptive histogram equalization;

𝑘 - number of image grayscale bits;

𝑀 - height of the image;

𝑁 - width of the image.

## Feature Extraction

The feature extraction process was one of the important processes in recognizing the class of an image object, which aimed to measure the quantitative magnitude of the characteristics of each pixel. The results of this process represented the characteristics of an object that can distinguish object classes well. In this study, the feature extraction process used was the 3D-GLCMmethod. 3D-GLCM Forming, Texture analysis using 3DGLCM considered the relationship between three pixels called reference and two neighboring pixels.



Figure 4 shows the steps in the formation of the 3DGLCM matrix where iris image had been converted into a grayscale image, and image improvements had been made using AHE

Figure 5 illustrates how to form a co-occurrence matrix at a distance of d=1 pixel and direction =00 with a 7x7 grayscale image and with 4 gray value(0,1,2,3).

1. 7x7 grayscale image with 4 gray values;
2. co-occurrence matrix 4x4 layer k = 0;
3. co-occurrence matrix 4x4 layer k = 1;
4. co-occurrence matrix 4x4 layer k = 2;
5. co-occurrence matrix 4x4 layer k = 3



### 3D-GLCM Feature Extraction

After being formed, the 3D-GLCM matrix was at a distance of d and in all possible directions. The next step was to calculate the statistical characteristics of all the 3D-GLCM matrices that were formed.

In this study, six statistical characteristics were used, namely:

### Max Probability

Max Probability shows the size of the response strength of the co-occurrence matrix with a range of values [0,1], which is 0 to 1 [7]. The method of calculating the max probability is shown in equation (3):



where:

i = ith row in 3D-GLCM;

j = jth column in 3D-GLCM; k = kth layer in 3D-GLCM;

pijk = ijk-element in the normalized matrix.

### Entropy

Entropy shows a measure of the irregularity of the intensity distribution of an image in the co- occurrence matrix. Entropy = 0, if all elements of pijk are 0, and will be maximum when all elements of pijk are equal. The maximum value is 2log2K [7]. The method of calculating entropy is shown in equation (4) below:



where: Q = matrix size;

i = ith row in 3D-GLCM;

j = jth column in 3D-GLCM; k = kth layer in 3D-GLCM;

pijk = ijkth element in the normalized matrix.

It is assumed that the value of entropy = 0 if all elements ofpijk are 0.

### Energy

Energy is a feature to measure the concentration of intensity pairs in the co-occurrence matrix with a valuerange of [0,1], which is 0 to 1[7]. The energy value will increase if the pixel pairs that meet the requirements of the co-occurrence intensity matrix are concentrated in several coordinates and decrease if they are spread out. The method of calculating energy is in equation (5).



where: Q - matrix size;

i = ith row in 3D-GLCM;

j = jth column in 3D-GLCM; k = kth layer in 3D- GLCM;

pijk = ijkth element in the normalized matrix.

### Correlation

This feature shows the level of correlation between pixels in an image with a value range of 1 to -1. This measure is undefined if one of the standard deviations is 0 [7]. The way of calculating the correlation is shown in equation (6) as follows:



where: Q = matrix size;

i = ith row in 3D-GLCM;

j = jth column in 3D-GLCM; k = kth layer in 3D- GLCM;

pijk = ijkth element in the normalized matrix.

### Contrast

Contrast is a feature used to measure the strength of the difference in intensity in the image with a value range of 0 (when the co-occurrence matrix is constant) to (Q- 1)^2 [7].The contrast value increases if the variation in image intensity is high and decreases when the variation is low. The equation used to measure the contrast of an image is shown in equation (6) below:



where: Q - matrix size;

i - ith row in 3D-GLCM; j–jth column in 3D-GLCM; k - kth layer in 3D-GLCM;

pijk- ijkth element in the normalized matrix.

### Homogeneity

Homogeneity is used to measure the spatial proximity of the distribution in the diagonal elements of the cooccurrence matrix—the variation of image intensity with a value range of 0 to 1. The maximum is achieved when the co-occurrence matrix is in the form of a diagonal matrix[7]. The homogeneity value increases when the intensity variation in the image decreases. The method of calculating homogeneity is shown in equation (7) below:



where: Q - matrix size;

i = ith row in 3D-GLCM;

j = jthcolumn in 3D-GLCM; k = kth layer in 3D-GLCM;

pijk= ijkth element in the normalized matrix.

The statistical feature extraction process results are

used as input data in training to identify input patterns and pairs of output patterns (Figure 6).



## Classification

The classification step in this research was divided into two processes, namely the learning and testing process. Both processes have referred to the results of statistical feature extraction calculations. The statistical feature extraction parameters used are maximum probability (max probability), entropy, energy, correlation, contrast, and homogeneity. The training process in this study was based on the calculation results of statistical feature extraction using 3D- GLCM. This value becomes input data in the learning process. Figure 6 shows the flow of the learning process.

### Testing

The testing process was done by measuring the distance of each test image to the group value parameters resulting from the training process. The distance here 157 was used to determine the degree of similarity or dissimilarity of two feature vectors. The steps of the training process are shown in Figure 7.

Distance measurement was used by the Euclideandistance (ED) method.



ED is a metric used to measure the similarity of two vectors. ED calculates the square root of thedifference of two vectors, and the formula is shown in equation

(8) .



where: dij =Euclidean distance between i and j; i= ith data;

j=jth data;

## Results and Discussion:

This study tested the performance of the 3D-GLCM method on two iris datasets. Dataset I contained 100 images of the right iris for learning and testing.

Dataset II contained 100 images of the left iris for learning and for testing too. At the testing stage, the cross-validation (CV) technique has been carried out. The dataset is separated into two subsets, namely learning process data and testing process data. Tests carried out on the dataset have used 80 images as learning data, and 20 images have been used for testing so that in this study, a 5-fold CV has been used. Evaluation of the classification results was done by using the confusion matrix method. The confusion matrix has provided details of misclassification. The predicted class was shown at the top of the matrix, and the observed class is on the left. Each cell contained a number representing the actual number of cases of the class observed by the model for a given prediction class, shown in Figure 8.

This study was divided into four classes: DM class, HC class, DM and HC class, and normal class.

In the multi-class classification, sensitivity, specificity, and accuracy were calculated through the average sensitivity, specificity, and accuracy for each class, where the calculation formula is shown in equations (9) to (11).

where: k - number of classes;

TPi - true positive for the ith class; TNi - true negative for the ith class; FNi - false negative for the ith class; FPi - false positive for the ith class.

The number of positive data classified correctly by the system is called True positive. The number of negative data that are classified correctly by the system is called True negative. The number of positive data but reclassified incorrectly by the system is called False Negative. The number of negative data classified correctly by the system is called a false negative.

The obtained test results were used to calculate the sensitivity, specification, and accuracy. Figure 9a shows the test results on Dataset I, and Figure 9b shows the test results on Dataset II. The test results were carried out by measuring the value of sensitivity, specificity, and accuracy to differences in the image gray level. It was done by comparing four gray levels; they are the

gray level of the 4-bit gray-level which consists of 16 gray levels; the 5-bits gray-level, which consists of 32 gray levels; the 6-bits, which consist of 64 gray levels; and the 7 bits and 8 bits which consists of 256 gray levels.



## Conclusions

This paper proposes to identify two disease disorders at once in a 3D-GLCM algorithm where the output is the identification of early detection in one of the four possible clusters, namely DM, HC, DM, and HC or normal.

In this paper, identification with 3D-GLCM has been carried out by comparing the gray levels of 16, 32, 64, and 256. The test results show that the more gray levels contained in the image, the higher the sensitivity and accuracy values. While the specificity value is getting smaller, the 3D-GLCM method can get a good recognition rate.

The limitation of this study is the small number of datasets. However, this can be solved by using the cross-validation method. Another weakness is in the iris image that is generated during data retrieval. Where if the eye shape of the respondent is small/slanted, the iris taken will be covered by the eyelids, and if the respondent has thick eyelashes, it causes noise in the iris in the form of eyelashes. The existence of this weakness affects the value of accuracy, specifications, and sensitivity.

Therefore, to test the performance of the algorithm in future research, it is necessary to pay attention to the collection of iris data that is free from eyelid and eyelash noise.

## Acknowledgment

I would like to thank my guide Dr. D. Prashanth Kumar sir for esteemed guidance to write this paper and for helping to collect the required information.

## References

1. AN EFFICIENT MACHINE LEARNING APPROACH TO NEPHROLOGY THROUGH IRISRECOGNITION

[C.D.Divya,](https://link.springer.com/article/10.1007/s44163-021-00010-4#auth-C__D_-Divya) [H.L.Gururaj,](https://link.springer.com/article/10.1007/s44163-021-00010-4#auth-H__L_-Gururaj) [R.Rohan,](https://link.springer.com/article/10.1007/s44163-021-00010-4#auth-R_-Rohan) [V.Bhagyalakshm](https://link.springer.com/article/10.1007/s44163-021-00010-4#auth-V_-Bhagyalakshmi) [i,H.A.Rashmi,](https://link.springer.com/article/10.1007/s44163-021-00010-4#auth-H__A_-Rashmi) [A.Domnick](https://link.springer.com/article/10.1007/s44163-021-00010-4#auth-A_-Domnick) & [Francesco Flammini](https://link.springer.com/article/10.1007/s44163-021-00010-4#auth-Francesco-Flammini). Discover Artificial Intelligence volume 1, Article number: 10(2021).

1. SHAW J. E., SICREE R. A., and ZIMMET

P. Z. Global estimates of the prevalence of diabetes for 2010 and 2030. Diabetes Research and Clinical Practice, 2010, 87(1): 4–14.

1. GUARIGUATA L., WHITING D. R., HAMBLETON I., et al. Global estimates of diabetes prevalence for 2013 and projections for 2035. Diabetes Research and Clinical Practice, 2014, 103(2): 137–149.
2. RAMLEE R. A., AZIZ K. A., RANJIT S., and ESRO M. Automated Detecting Arcus senilis, symptom for Cholesterol Presence using Iris Recognition Algorithm. Journal of telecommunication , Electronic and Computer Engineering,2011,3(2):29-39.
3. LESMANA I. P. D., PURNAMA I. K. E., and PURNOMO H. P. Abnormal Condition Detection of Pancreatic Beta-Cells as the Cause of Diabetes Mellitus Based on Iris Image. Proceedings of the International Conference on Instrumentation, Communication, Information Technology and Biomedical Engineering, 2011, November: 4–9.
4. [Type 1 diabetes - Symptoms and causes -](https://www.mayoclinic.org/diseases-conditions/type-1-diabetes/symptoms-causes/syc-20353011) [Mayo Clinic.](https://www.mayoclinic.org/diseases-conditions/type-1-diabetes/symptoms-causes/syc-20353011)
5. GONZALEZ R. C. and WOODS R. E. Digital Image Processing. Third Edition, Pearson Prentice Hall, 2008.