 978-1-7281-5291-2/19/$31.00

Application of Machine Learning Techniques for Diagnosis of Diabetes Based on Iridology

1st Ratna Aminah
Department of Physics,
Faculty of Mathematics and Natural Sciences
Universitas Indonesia
Depok, Indonesia
ratna.aminah@sci.ui.ac.id

2nd Adhi Harmoko Saputro
Department of Physics,
Faculty of Mathematics and Natural Sciences
Universitas Indonesia
Depok, Indonesia
adhi@sci.ui.ac.id

***Abstract*— Complementary and alternative medicine (CAM) is a system and therapy in the medical field that works based on knowledge, abilities, and practice. CAM is used to maintain health, diagnose disease, or to prevent and treat mental and physical illness. This technique can predict and treat disease. At the same time, machine learning has been widely used in the application of the biomedical field as a tool for diagnosing disease. The purpose of this work is to validate the use of iridology as a valid scientific technique to diagnose diabetes disease. Iridology combined with machine learning to simplify the diagnose process. Iris images were captured using Camera Iriscope Iris Analyzer Iridology. The region of interest (ROI) was cropped according to the location of the pancreas organ on iridology chart. The Gray Level Co-Occurrence Matrix method has been implemented for feature extraction. Five different classifiers method is used to classify diabetic and non-diabetic classes. The results are then validated and evaluated by using the k-fold cross-validation and confusion matrix, respectively. The subject consisted of two groups: one was 16 subjects non-diabetic and 11 subjects diabetic. The results show that the best accuracy is 85.6%, with specificity is 0.90, and the sensitivity is 0.80.**

Keywords*— *Iridology; Diabetes; Disease Diagnosis, Image Processing; Machine Learning

I. Introduction

The growth of diabetes disease is very fast in recent years. The International Diabetes Federation (IDF) estimates that people worldwide with an age range of 20-79 years in 2017 are affected diabetes by around 146 million people living in rural areas and 279 million people living in urban areas and will increase to around 156 million people live in rural areas, and 473 million people live in urban areas. Based on IDF data, in 2017, Indonesia ranks 6th in the top ten countries with the most diabetic population, with 10.3 million people and will increase in 2045 to 10.6 million people [1].

Diabetes is a metabolic disease that occurs when the pancreas does not produce insulin, or when the pancreas produces insulin, but the body cannot use it effectively [2]. Diabetes was characterized by an elevated in blood glucose levels or commonly called hyperglycemia. Insulin is a hormone produced in the pancreas and serves to transport glucose from the bloodstream to body cells. Diabetes is divided into three types, namely type I, type II, and gestational diabetes (GDM). Type I diabetes is a condition when pancreas produce insulin in small amounts because of immune system attacks beta cells in the pancreas [3]. It commonly called insulin-dependent. The most common is type II diabetes, which characterized by hyperglycemia due to deficient of insulin production, and the body becomes resistant to insulin [4] while GDM is a condition when pregnant women have a high blood glucose level during their pregnancy [1].

Based on IDF data, of the 10.3 million people affected by diabetes in Indonesia, around 7.6 million people were not detected early. Delays in detecting diabetes will lead to severe damage to the blood vessels, eye, heart, nerves, and kidney[1].

Complementary and alternative medicine (CAM) is a system and therapy in the medical field that works based on knowledge, abilities, and practice. CAM is used to maintain health, diagnose disease, or to prevent and treat mental and physical illness.[5]. Many practitioners of modern conventional medicine do not believe that CAM is proven and orthodox [6]. However, CAM therapies are quite well known for treating disease such as diabetes. In some countries, CAM is widely used for health care such as in Africa, Australia, China, India, Indonesia, and Western countries. In Indonesia, CAM used by 40% of the entire population and used 70% of the rural population [7].

Iridology is a form of CAM technique that is believed that the patterns, shapes, colors, tissue damage, and other characters in the iris can be used to determine the health status of the body [8]. Iridology practitioners match their interpretations to iridology scheme. Bernard Jensen divides the iridology scheme approximately into 80-90 different zones, and each zone is a representation of specific body organs [9].

Advanced machine learning has been widely used in the application of the biomedical field as a tool for diagnosing disease. [10]. The purpose of this work is to validate the use of iridology as a valid scientific technique to diagnose diabetes disease. Iridology combined with machine learning to simplify the diagnose process. Table 1 shows some of the studies that have been carried out relating to iridology for disease detection.

TABLE I. SUMMARIZED APPLICATION OF IRIDOLOGY FOR DISEASE DETECTION

Application	Author	Feature Extraction Technique	Classification Technique
Diabetes	Bansal et al. [11]	2-D Discrete Wavelet Transform (DWT)	Support Vector Machine (SVM)
Alzheimer	Hernandez et al. [12]	First-order statistics	ZeroR, Naïve Bayes and Multi-layer Perceptron
Kidney disease	Hussein et al. [13]	Gabor Filter	Neural Network
Liver disease	Herlambang et al. [14]	Gray Level Co-Occurrence Matrix	Back-propagation Neural Network
Stomach disorder	Dewi et al. [15]	Principal Component Analysis (PCA)	Back-propagation Neural Network

In this paper, Gray Level Co-Occurrence Matrix for feature extraction and several classification algorithms such as Naive Bayes, Random Forest, Support Vector Machine, Ensemble Learning and k Nearest Neighbor for the classifier based on iridology has been proposed for diabetes prediction system. Then, the validation performed by calculating average accuracy from the k-fold Cross Validation method and evaluated by using Confusion Matrix.

II. System Design And Implementation

The process begins with data collection in the form of iris images acquisition and measurement of blood glucose level, then proceed with image pre-processing that consists of image enhancement, image localization, and image normalization. After that, proceed with segmenting Region of Interest (ROI), feature extraction, and classification. Block diagram of the system used for this research can be seen in Fig. 1

A. Data collection

Data were taken from 26 subjects, of which 15 were non-diabetic subjects (5 male and ten female), and 11 were diabetic subjects (6 male and five female). The data taken from each subject are the measurement of blood glucose levels for reference values and iris image acquisition. Random blood glucose test was performed on the subject around 1-1,5 hours after eating. Blood glucose measurements are carried out using a glucometer. The design of the measurement system in this present study consisted of a Camera Iriscope Iris Analyzer Iridology, PC, and software CV Advance Analysis System to display the iris image as shown in Fig. 2.

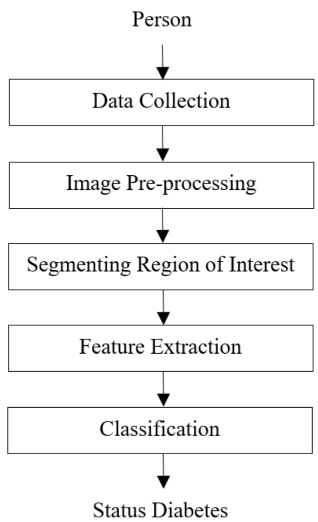


Fig. 1. Block diagram of the diabetes prediction system

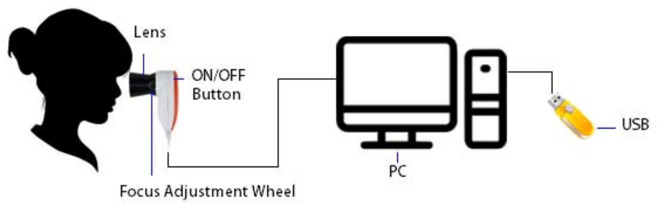


Fig. 2. Design of measurement system

The value of glucose sugar levels of subjects ranged from 89 mg/dL for the lowest value and 212 mg/dL for the highest value. The age range of subject diabetic is around 20-58 years, with an average age is 48 years. While for subject non-diabetic have an age range around 20-51 years with an average age is 28 years. The duration of diabetic varied from 1 year to 20 years with distribution below five years there are six people, 5-10 years there are two people, and more ten years there are three people.

B. Image pre-processing

Image pre-processing is an improvement process of the image into a suitable form to obtain some image features necessary for subsequent processing. It involves three steps, namely, image enhancement, image localization, and image normalization. Image enhancement aims to increases image quality and the dynamic range of the features into a suitable form so that the features can be detected easily. Image enhancement consists of three steps, namely, filtering, grayscale, and edge detection. Filtering process aims to sharpen the detail in the image using the Fast Fourier Transform. In process grayscale, the RGB image is changed to grayscale by adding up the R, G, and B value then divide it by 3 to get the desired grayscale image. The grayscale process aims to simplify the information on the image. In image enhancement, edge detection is also done using Canny Edge.

Image localization is the process of separating the iris from other parts of the eye by looking for the outer and inner circles of the iris. Many algorithms have been used in previous researchers to find the inner and outer boundaries of the iris, such as Integro-Differential Operator (IDO) that introduced by Daughman [16], Circular Hough Transform that proposed by Wildes [17], and active contour model [18]. In this study, Circular Hough Transform (CHT) was used to find iris boundaries in an image by looking for the radius coordinates and the center of the pupil and iris area. An edge map produced by the first derivative of the intensity value on eye image. The parameters of the circle were estimated by the circle that passes the maximum edge of the edge map [19].

Normalization is a process of eliminating inconsistent dimensions which cause by stretching of the iris because of various lightning levels, various imaging distances, camera rotation, tilted head position, and rotation of the eyes [11]. To solving that problem, Daughman proposed rubber sheet model that will change each point in the image from cartesian coordinates (x, y) into polar coordinates (r, θ) with the center of pupil expressed as a reference point as shown in Fig. 3. This method is modeled as:

$$I\left(x(r,\theta), y(r,\theta)\right) \rightarrow I(r,\theta) \tag{1}$$

where r lies in the interval [0,1] and θ varies from [0,2π] [13]. With,

$$x(r,\theta) = (1-r)x_p(\theta) + rx_l(\theta) \tag{2}$$

And

$$y(r,\theta) = (1-r)y_p(\theta) + ry_l(\theta) \tag{3}$$

(x, y) is cartesian coordinate, (r, θ) is normalized polar coordinate, x_p,y_p and x_l, y_l are the boundaries of coordinates iris and pupils throughout θ direction. Normalization produced images of size 360×270 pixels.

C. Segmenting Region of Interest

Image segmentation aims to partition the iris to get the region which is desired. It is essential to choose the region to get the appropriate features to be analyzed. The pancreas is an organ that related to diabetes because of its job to produce insulin. The region was cut from normalized iris image using Rubber Sheet.

The pancreas consists of three parts, namely the head, body, and tail. Based on iridology scheme, the head of the pancreas is located in the right eye in the area between 7 to 8 o'clock, while the tail of the pancreas is located in the left eye in the area between 4 to 5 o'clock, and body of the pancreas is located in the left eye in the area and 7 to 8 o'clock [19]. So that for the right eye will be selected one part and for the left eye will be selected two parts with the same size. The size of the segmented Region of Interest is 360 × 30 pixels, as shown in Fig. 4.

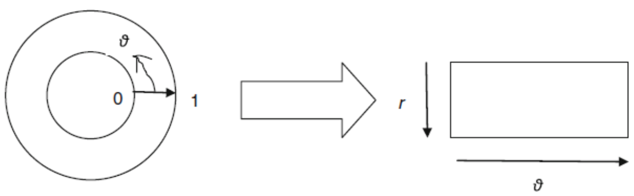


Fig. 3. Iris normalization process

Fig. 4. Segmenting Region of Interest

D. Feature Extraction

Feature extraction aims to get the feature from the data which are desired. Iris consists of blood vessels and nerves. When the body is not healthy, the texture on the iris will change. These changes can take the form of holes, fractures, hyperpigmentation, radii solaris, nerve rings, lines, and color of iris [20]. Many researchers have proposed feature extraction algorithms such as Gabor filters [16], 2-D discrete wavelet transform (DWT) [21], and Gray Level Co-Occurrence Matrix.

In this study, the Gray Level Co-occurrence Matrix (GLCM) has been implemented to extract the feature from iris images. GLCM is a 2nd order statistical method for checking the texture and considering spatial relations between pixels. [22]. GLCM texture picks up the relation between two pixels at a time, called the reference and the neighbor pixel. The features extracted in GLCM are contrast, correlation, energy, and homogeneity. Table 2 shows features in GLCM and their formulas.

E. Classification

In machine learning, classification techniques are used for grouping data according to the data class based on the features that have been obtained. Iris image that has been processed will be divided into training data and testing data. The conclusions of the system will be analyzed through testing data. In this study using five different classifiers, named as Naive Bayes (NB), Support Vector Machine (SVM), Random Forest (RF), Ensemble Learning (EL) and k Nearest Neighbor (kNN). The prediction system has been implemented using MATLAB 9.2. The results of the system have been validated using 10-fold cross-validation. All data are divided randomly into ten sections. Nine sections as training data are used to train the system, while the other one section as testing data is used to see the performance of the system. Each section has the same size.

TABLE II. FEATURE IN GLCM

Textural Features	Formulas
Contrast	$\sum_{i=1}^{N}\sum_{j=1}^{N}(i-j)^2 P(i,j)$
Correlation	$\dfrac{\sum_{i=1}^{N}\sum_{j=1}^{N}(ij)P(i,j)-\mu_i\mu_j}{\sigma_i\sigma_j}$
Energy	$\sum_{i=1}^{N}\sum_{j=1}^{N}P(i,j)^2$
Homogeneity	$\sum_{i=1}^{N}\sum_{j=1}^{N}\dfrac{P(i,j)}{1+(i-j)^2}$

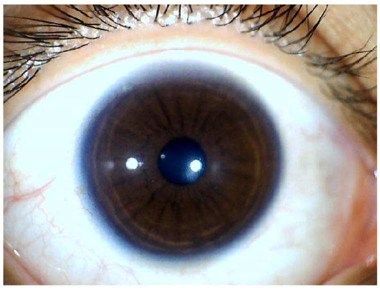
Hence, there are ten iterations for the training and testing process. Then the result evaluated using confusion matrix. The accuracy of the classification system is measured for all iterations, and the final accuracy is obtained from the average accuracy of each iteration. Besides, specificity and sensitivity were calculated to evaluate the performance of the system.

III. Result and Discussion

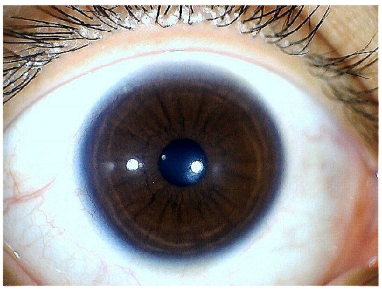
Fig. 5 shows a comparison of the original image with the image after the sharpening process. It can be seen that image after sharpening has more lucid in details, and the texture of the iris look more clearly. Because the detail in the image looks more clearly, more edge will be detected in the edge detection process. Consequently, in the process of finding the outer and inner boundaries of the iris or image, localization will be more comfortable. Fig. 6 shows the result of the localization process.

Table 3 shows the overall accuracy obtained using for a different classifier with variations in angular orientation at GLCM. The best accuracy for the Naive Bayes method is obtained when the angle is 90˚ and for Random Forest obtained when using three angles, 0;45;135. While the best accuracy for Support Vector Machine, Ensemble Learning, and k Nearest Neighbor is obtained when the four angles are used. From the overall variation, k Nearest Neighbor has the best classification accuracy compared to another method. The accuracy obtained is 85.6%. The best variation is also seen from the comparison of the specificity and sensitivity values.

Sensitivity and specificity both were calculated for the RF, EL, and kNN classifier as these methods gave the best accuracy amongst all methods. Table 4 and Table 5 shows the specificity and sensitivity of classifier performance, respectively. Maximum sensitivity and specificity have been obtained as 0.800 and 0.900 when using four angles and kNN classifier.



(a)



(b)

Fig. 5. Image before sharpening (a) and after sharpening (b)

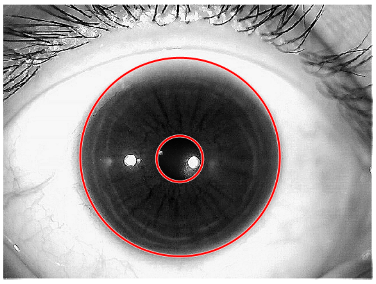


Fig. 6. Image localization using Circular Hough Transform

TABLE III. COMPARISON OF ACCURACY WITH VARIATIONS IN ANGULAR ORIENTATION FOR DIFFERENT CLASSIFIERS

Angular Orientation (˚)	Classifier accuracies				
	NB (%)	*SVM (%)*	*RF (%)*	*kNN (%)*	*EL(%)*
0	62.37	71.12	68.60	68.58	70.28
45	66.48	70.32	79.40	76.88	74.03
90	67.32	75.63	78.08	83.93	79.80
135	65.23	71.07	76.08	74.00	75.63
0;45	59.83	73.98	79.40	78.90	80.65
0;90	64.80	76.45	78.12	81.02	78.55
0;135	63.98	70.65	78.98	78.93	78.15
45;90	67.30	77.28	79.80	83.55	81.87
45;135	64.82	73.53	80.18	79.82	79.02
90;135	66.07	73.20	79.83	82.72	78.13
0;45;90	65.22	76.43	81.4	83.08	83.48
0;45;135	64.40	77.67	81.48	82.25	79.82
0;90;135	65.63	78.55	78.93	83.53	78.93
45;90;135	66.05	77.35	81.03	83.55	79.75
0;45;90;135	66.07	79.35	80.98	85.60	83.92

TABLE IV. COMPARISONS OF SPECIFICITY WITH VARIATIONS IN ANGULAR ORIENTATION FOR RF, ENSEMBLE LEARNING AND KNN METHOD

Angular Orientation (˚)	Classifier		
	Random Forest	*Ensemble Learning*	*kNN*
0	0.786	0.792	0.812
45	0.839	0.811	0.845
90	0.857	0.858	0.871
135	0.851	0.863	0.826
0;45	0.857	0.856	0.850
0;90	0.850	0.844	0.798
0;135	0.883	0.878	0.856
45;90	0.837	0.870	0.896
45;135	0.851	0.830	0.864
90;135	0.884	0.843	0.838
0;45;90	0.883	0.875	0.856
0;45;135	0.891	0.855	0.857
0;90;135	0.876	0.876	0.850
45;90;135	0.876	0.869	0.900
0;45;90;135	0.883	0.888	0.900

According to three parameters, namely accuracy, specificity, and sensitivity, the best variation is when angle orientation used four angles (0;45;90;135) with kNN method with accuracy is 85.60%, specificity is 0.900, and sensitivity is 0.800. Results obtained from this work shown that the GLCM as features extraction have a great ability to classify diabetic and non-diabetic based on iridology. Table 5 shows the comparison between the proposed model with existing iris diagnosis models for detecting disease.

TABLE V. COMPARISONS OF SENSITIVITY WITH VARIATIONS IN ANGULAR ORIENTATION FOR RF, ENSEMBLE LEARNING AND KNN METHOD

Angular Orientation (°)	Classifier		
	Random Forest	***Ensemble Learning***	***kNN***
0	0.489	0.544	0.463
45	0.672	0.614	0.636
90	0.651	0.694	0.784
135	0.606	0.603	0.592
0;45	0.685	0.717	0.703
0;90	0.660	0.685	0.797
0;135	0.629	0.615	0.673
45;90	0.729	0.727	0.729
45;135	0.706	0.716	0.660
90;135	0.652	0.672	0.761
0;45;90	0.694	0.762	0.763
0;45;135	0.683	0.691	0.741
0;90;135	0.635	0.876	0.800
45;90;135	0.689	0.673	0.717
0;45;90;135	0.685	0.751	0.800

TABLE VI. COMPARISON BETWEEN THE PROPOSED MODEL WITH EXISTING MODELS

Authors	**Disease**	**Classification accuracy (%)**
Purnama et al. [23]	Pancreas disorder	83
Sivasankar et al. [24]	Pulmonary disease	84.38
Samant et al.	Diabetes	89.63
Proposed model	Diabetes	85.6

IV. CONCLUSION

Iridology is an alternative medicine technique used in this work to classify status diabetes and non-diabetes using machine learning techniques. machine-learning has been widely used in the application of the biomedical field as a tool for diagnosing disease. CHT is used to find boundaries of the iris and Rubber sheet model is used to normalized iris image. The size of the segmented Region of Interest is 360 × 30 pixels. The proposed model uses Gray Level Co-Occurrence Matrix-based feature extraction with five different classifiers, named as Naive Bayes, Support Vector Machine, Random Forest, Ensemble Learning, and k Nearest Neighbor. The maximum accuracy is 85.6% with specificity, and sensitivity value is 0.900 and 0.800, respectively. In future work, analysis can be linked to lifestyle and environmental factors to explore further the consequences of diabetes on individual health and change in iris texture.

ACKNOWLEDGMENT

This research was carried out with financial support from Universitas Indonesia (PITTA Research Grant 2019).

REFERENCES

[1] IDF, *IDF Diabetes Atlas*. 2017.

[2] R. A. DeFronzo, E. Ferrannini, K. G. M. M. Alberti, P. Zimmet, and G. Alberti, *International Textbook of Diabetes Mellitus*. 2015.

[3] S. Canivell and R. Gomis, “Diagnosis and classification of autoimmune diabetes mellitus,” *Autoimmun. Rev.*, vol. 13, no. 4–5, pp. 403–407, 2014.

[4] A. Olokoba, O. Obateru, and L. Olokoba, “Type 2 Diabetes: a Review of Current Trends -,” *Int. J. Curr. Res. Rev.*, vol. 7, no. 18, pp. 61–66, 2015.

[5] Eurocam, *CAM 2020*. Belgium, 2014.

[6] E. Ernest, “Healing, hype or harm?,” *Societas*, 2008.

[7] H. T. Debas, R. Laxminarayan, and S. E. Straus, “Chapter 69 Complementary and Alternative Medicine,” pp. 1281–1292, 2004.

[8] A. Bansal, R. Agarwal, and R. . Sharma, “Pre-Diagnostic Tool to Predict Obstructive Lung Diseases Using Iris Recognition System. In Smart Innovations in Communication and Computational Sciences,” *Springer*, pp. 71–79, 2019.

[9] P. E. Harris, K. . Cooper, C. Relton, and K. . Thomas, “Prevalance of complementary and alternative medicine (CAM) use by the general population : A systematic review and update,” *Int. J. Clin*, vol. 66, pp. 924–939, 2012.

[10] C. Parmar, P. Grossmann, J. Bussink, P. Lambin, and H. . Aerts, “machine learning methods for quantitative radiomic biomarkers,” *Sci, Reports 5*, pp. 1–11, 2015.

[11] A. Bansal, R. Agarwal, and R. K. Sharma, “Determining diabetes using iris recognition system,” *Int. J. Diabetes Dev. Ctries.*, vol. 35, no. 4, pp. 432–438, 2015.

[12] F. Hernández, R. Vega, F. Tapia, D. Morocho, and W. Fuertes, “Early Detection of Alzheimer ’ s using digital image processing through Iridology , an Alternative method,” 2018.

[13] S. E. Hussein, O. A. Hassan, and M. H. Granat, “Assessment of the potential iridology for diagnosing kidney disease using wavelet analysis and neural networks,” *Biomed. Signal Process. Control*, vol. 8, no. 6, pp. 534–541, 2013.

[14] R. G. A. N. P. Herlambang, R. R. Isnanto, and A. Z. Ajub, “Application of liver disease detection using iridology with back-propagation neural network,” *ICITACEE 2015 - 2nd Int. Conf. Inf. Technol. Comput. Electr. Eng. Green Technol. Strength. Inf. Technol. Electr. Comput. Eng. Implementation, Proc.*, pp. 123–127, 2016.

[15] A. K. Dewi, A. Novianty, and T. W. Purboyo, “Stomach disorder detection through the Iris Image using Backpropagation Neural Network,” *2016 Int. Conf. Informatics Comput. ICIC 2016*, no. Icic, pp. 192–197, 2017.

[16] J. Daughman, “How iris recognition works?,” *IEEE Trans Circuits Syst Video Technol*, vol. 14, pp. 21–30, 2004.

[17] C. Tisse, L. Martin, L. Torres, and M. Robert, “Person identification technique using human iris recognition,” *Int. Conf. Vis. Interface*, 2002.

[18] N. Ritter, “Location of the pupil-iris border in slit-lamp images of the cornea,” *Int. Conf. Image Anal. Process.*, 1999.

[19] P. Samant and R. Agarwal, “Machine learning techniques for medical diagnosis of diabetes using iris images,” *Comput. Methods Programs Biomed.*, vol. 157, pp. 121–128, 2018.

[20] Z. Othman and A. . Prabuwono, “Preliminary study on iris recognition system: tissues of body organs in iridology,” *IEEE, EMBS, Conf. Biomed. Eng. Sci.*, 2010.

[21] X. He and P. Shi, “Extraction of complex wavelet features for iris recognition,” *19th Int. Conf. Pattern Recognit.*, pp. 1–4, 2008.

[22] A. Suresh and K. L. Shunmuganathan, “Image Texture Classification using Gray Level Co-Occurrence Matrix Based Statistical Features,” *Eur. J. Sci. Res.*, vol. 75, no. 4, pp. 591–597, 2012.

[23] I. Lesmana, I. Putu, E. Purnama, and M. Purnomo, “Abnormal condition detection of pancreatic beta-cells as the cause of diabetes mellitus based on iris image,” *2011 2nd Int. Conf. Instrumentation, Commun. Inf. Technol. Biomed. Eng.*, pp. 150–155, 2011.

[24] K. Sivanskar, M. Sujaritha, P. Pasupathi, and S. Muthukumar, “FCM based iris image analysis for tissue imbalance stage identification,” *Int. Conf. Emerg. Trends Sci. Eng. Technol.*, 2012.