Diagnosis of Diabetes Using Computer Methods: Soft Computing Methods for Diabetes Detection Using Iris

Piyush Samant, Ravinder Agarwal

Abstract—Complementary and Alternative Medicine (CAM) techniques are quite popular and effective for chronic diseases. Iridology is more than 150 years old CAM technique which analyzes the patterns, tissue weakness, color, shape, structure, etc. for disease diagnosis. The objective of this paper is to validate the use of iridology for the diagnosis of the diabetes. The suggested model was applied in a systemic disease with ocular effects. 200 subject data of 100 each diabetic and non-diabetic were evaluated. Complete procedure was kept very simple and free from the involvement of any iridologist. From the normalized iris, the region of interest was cropped. All 63 features were extracted using statistical, texture analysis, and two-dimensional discrete wavelet transformation. A comparison of accuracies of six different classifiers has been presented. The result shows 89.66% accuracy by the random forest classifier.

Keywords—Complementary and alternative medicine, Iridology, iris, feature extraction, classification, disease prediction.

I. INTRODUCTION

DIABETES mellitus has become an epidemic disease, and International Diabetes Federation (IDF) estimated that 381.8 million people are affected by diabetes mellitus. In India more than 62 million people are diagnosed with diabetes [1]. Diabetes mellitus is simply excess of blood sugar level. Pancreas, with its function of insulin secretion, is a fish shaped gland which is often associated with diabetes mellitus. Type 2 is the most common form of diabetes mellitus in which body cells are not able to utilize the insulin produced by the pancreas organ for the consumption of glucose. CAM techniques have shown their potential for deterrence of diabetes. Researchers [2], [3] have shown that CAM techniques are natural and effective for chronic diseases like diabetes.

Iridology is an old CAM technique which provides an alternative to mainstream diagnosis techniques. Its invasive and painful for the patients. Iridology makes use of color, shape, structures, pattern, marking, and other features of iris to get the information about the health status of individual. Famous iridologist Jensen [3] developed the iridology chart in which iris is divided into 80-90 different zones, and each zone of this chart represents a particular body organ. Table I shows potential research work where iridology and machine learning

Piyush Samant and Ravinder Agarwal are with EIED, Thapar University, Patiala 147004, India (phone: 0175-23933062; e-mail: piyush.samant@thapar.edu, ravinder_eeed@thapar.edu).

techniques are used for different biomedical application and disease prediction.

TABLE I
DIFFERENT IRIDOLOGY APPLICATIONS

Application	Segmentation and Normalization Algorithms	Classification Technique
Predict gender using Iris [4]-[6]	CHT [4]-[6] Daugman's rubber sheet Normalization [4]-[6] Geometric [4], statistical [5], Texture [4], 2-D DWT features [5] PCA [6]	Decision tree [4] SVM Neural Network [6]
Diabetes Detection [7]	CHT Daugman's rubber sheet normalization, Gabor filter and 2-D DWT based features	SVM
Iris Tumor Detection [8], [9]	Canny Edge detection [8] CHT [8], [9]	Histogram threshold [8], [9]
Pancreatic Beta-Cells as the cause of diabetes [10]	Quantitative features like texture measures	ANMBP
Early detection of diabetes [11]	Color image coding	ANN
Broken tissue pancreas through iridology [11]	Cropped ROI directly from iris	Broken tissue analysis through computer vision
Kidney Disease [12]	CHT Rubber sheet normalization	ANFIS
Healthcare Information [13]	Histogram Equalization, Gabor filter and 2D DWT Iris texture analysis from different regions of interest	ANN

CHT= Circular Haugh Transform, DWT= Discrete Wavelet Transformation, ROI= Region of Interest, PCA= Principle Component Analysis, SVM= Support Vector Machine, ANN= Artificial Neural Network, ANFIS= Artificial Neuro-fuzzy Inference System, ANMBP= Neighborhood based Modified Backpropagation using Adaptive Learning Parameters

In the present research work, old CAM technique (iridology) was combined with modern computer machine vision for diabetic diagnosis. Iris segmentation technique was used for iris segmentation and normalization. This piece of research uses to integrate different dispersed techniques of iridology as a portable medical diagnosis technique.

II. METHODOLOGY

To investigate the potential of iridology along with modern machine learning techniques, a database of 200 (100 diabetic and 100 non-diabetic) subjects was taken from I-SCAN 2 of crossmatch technologies by capturing the infrared images of both the eyes, which acquires gray infra-red images of size

640X480 of both iris simultaneously. Thereafter, pre-image processing and machine learning techniques were used.

After segmentation and rubber-sheet normalization, each iris was converted in to a 2D array of size 360x720. Two regions of interest *(ROIs)* were extracted from the right eye, and one ROI was extracted from left eye as per modern iridology chart of Jensen [14]. ROI was extracted corresponding to the position of pancreas organ in both of the iris. Total 63 features (statististical, textural and wavelet features) vector are randomly extracted from ROIs and stag together in order to make an array. For classification, six classification algorithms were accessed. Broadly, four major stages were followed to design diabetes detection model. Fig. 1 shows a flow chart of methodology and briefly discussed in subsections.
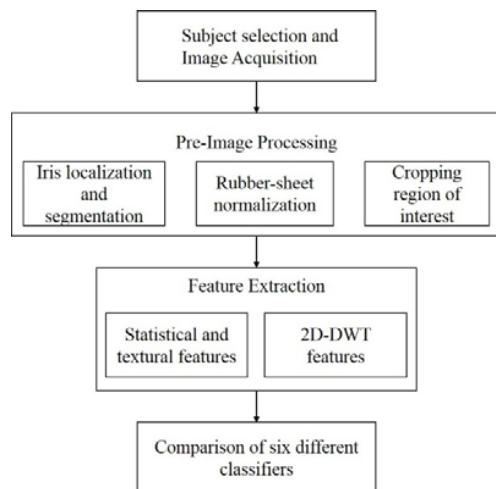

Fig. 1 Different stages for the proposed methodology

A. Subject Selection and Image Acquisition

A controlled group study of type II diabetic and healthy subjects was performed. Procedure for data acquisition was consulted by the university health center and duly authorized by university research board at Thapar University, Patiala. Care had been taken to consider to select the subject of different age group and gender. Table II shows the detail subject distribution. Diabetes diagnosed time of subjects varies from 6 months to 20 years.

TABLE II
DETAILED SUBJECT DISTRIBUTIONS

	No. of Male	No. of Female	Mean Age	Standard Deviation
Diabetic Subjects	58	42	53.32	9.9
Healthy subjects	56	44	52.86	10.63

B. Iris localizing and Segmentation

First step in iris segmentation was to see the inner and outer boundaries of iris. In recent years, various algorithms are proposed by the researchers for segmentation of iris from the digital image of eye [15]-[20]. The most imperative work in the early history of iris segmentation is Daugman's 1994 patent [15], [16]. It describes an effective iris segmentation

methodology in detail. It is fair to say that iris biometrics as a field has developed with the concepts in Daugman's approach becoming a standard reference model. Daugman [17]-[19] introduced Integro-Differential Operator (IDO) for IRIS segmentation, the operator search the circular boundary to detect iris within the digital image of eye. Wildes [20] proposed an segmentation algorithm in order to find out center and radius of IRIS region, based on circular Hough Transform (CHT). In the present research, an advanced and enhanced version of Daugman's IDO is presented. A simple CHT algorithm was employed to locate center coordinates and radius of inner and outer boundaries of IRIS. Thresholding followed by first order derivative was calculated and created an edge map. Now, circle parameters were estimated with the help of this edge map. Estimation was done by calculating the maximum edge points passing through the particular circle in the edge map.

C. Rubber-Sheet Normalization

Rubber-sheet normalization is nothing but transforming from Cartesian coordinate to polar coordinates. Basic idea was to convert iris in to a homogeneous rectangular form which is invariant to size, orientation, angle of camera, light illumination level, etc. Every single point (x,y) in the Cartesian coordinate of segmented iris was transformed into its corresponding polar coordinate (r, θ) as:

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

where r lies into a unit interval [0, 1], and angle θ varies from [0,2π]. 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_p, y_p) and (x_l, y_l) are the parameters of pupil and iris respectively in θ direction. A homogeneous 2D array of size 360x720 was obtained after applying the rubber sheet normalization algorithm.

D. Extraction of Region of Interest (ROI)

According to iridology chart [14], [21] the head, tail and body of pancreas organ, represents the three different segments into the iris. Head of pancreas represents in right eye between 36' to 39', while body and tail represents in left eye between 38' to 41' and 21' to 24', respectively. Therefore, these specific regions were cropped from the rubber sheet normalized rectangular representation of segmented iris. These ROI are of size 360x36. Finally, the features were extracted from these regions. Fig. 2 shows the cropped ROI for both left and right eye.

E. Extracting Features

IRIS consists of rich texture information and breaking of tissues of IRIS, which is directly associated with these texture features. Depending on the health of individual, changes can

be observed in these texture features into the ROI. At the same time, wavelet transforms show its great potential in various fields like in matching, biomedical, telecommunication, etc. Discrete Wavelet Transforms (DWT) are very suitable for non-stationary image analysis. Its ability of handling time space utilization and capacity of handling interference terms makes it appropriate for texture analysis [22], [23]. Total 63 statistical, texture, and DWT features were used to quantify the broken tissue information of the IRIS. Table III shows the statistical and textural features extracted.

(a)

(b)
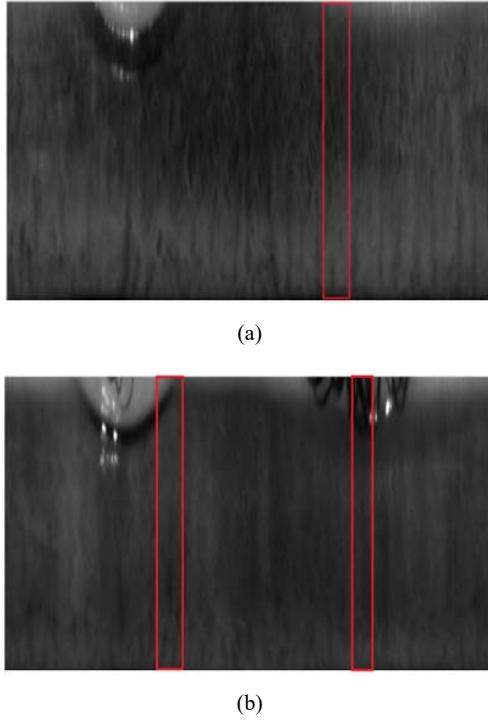
Fig. 2 (a) ROI for right eye (b) ROI for left eye

TABLE III
EXCRETED FEATURES FROM ROI

Feature Name	Formula
Mean Intensity	$\dfrac{1}{N}\sum_{i=1}^{N} X(i)$
Standard Deviation	$\left(\dfrac{1}{(N-1)}\sum_{i=1}^{N}(X(i)-\bar{X})^2\right)^{1/2}$
Entropy	$\sum_{i=1}^{N_1} P(i).\log_2 P(i)$
Contrast	$\sum_{i,j}|i-j|^2\, P(i,j)$
Correlation	$\sum_{i,j}(i-\mu_i)(j-\mu_j)\, P(i,j)$
Energy	$\sum_{i,j} P(i,j)^2$
Homogeneity:	$\sum_{i,j}\dfrac{P(i,j)}{1+|i-j|}$

Wavelet Features

Decomposing digital image into time and frequency components can be an effecting technique to analyze the lurking information within the image. Statistical and textural features were calculated for 3-Dentation DWT decompositions. Decompositions simply show high and low frequency components of ROI. Fig. 3 shows the schematic of the undecimated three-dimensional wavelet transform of ROI, where, ROI X (i,j) is decomposed into eight decompositions, by the means of low pass (L) and high pass (H) filtering.
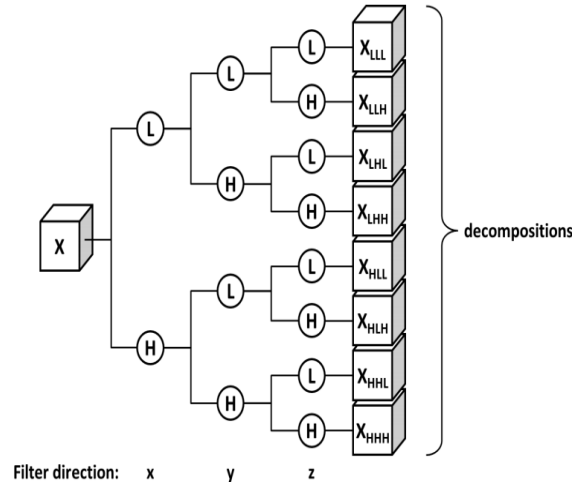

Fig. 3 Schematic of decomposition of 3D wavelet transform

F. Classification

In advance computing methods, classification techniques are known as a supervised learning assignment of concluding a decision from categorized training dataset. Classification algorithms infer the hypothesis which helps in predicting the labels from the testing dataset by analyzing the training dataset [24]. Here, six different classifiers were investigated, named as Binary Tree Model (BT), Random Forest (RF), Adaptive Boosting Model (AB), Support Vector Machine (SVM), Generalized Linear Models (GL), and Neural Network Model (NN). Classifiers were implemented in R package [25]. These classifiers belong to different fields of statistics and computer science and it is very difficult to understand each and every parameter of classifiers. Manual parameter selection leads to biasing in the performance index. So, parameter selection was done through cross validation for optimal performance of classifier. 70% of data were used for training the classifier and 30% were used for the testing purpose. Testing data were trained through repeated iteration technique, and validation was performed through 10-fold cross validation under ROC curve.

III. RESULT AND DISCUSSION

Bio-Medical imaging is regularly used as handy source of information for clinical study. It is a non-contact, non-invasive, and cost effective diagnostic tool. The technique adapted in this research work provides automated methods

with modern machine learning techniques to access the correlation between iris and diabetes. Fig. 4 shows the overall classification performances of six classifiers. Maximum accuracy of 89.66% was achieved by the RF classifier.

A. Area under the Curve (AUC) Analysis

AUC has the ability to access the performance of the classifier over its entire operating range. Fig. 5 depicts the ROC curve between false positive rate and true positive rate of various classifiers against the selected feature set.
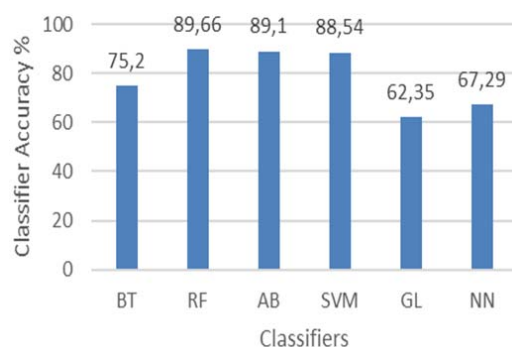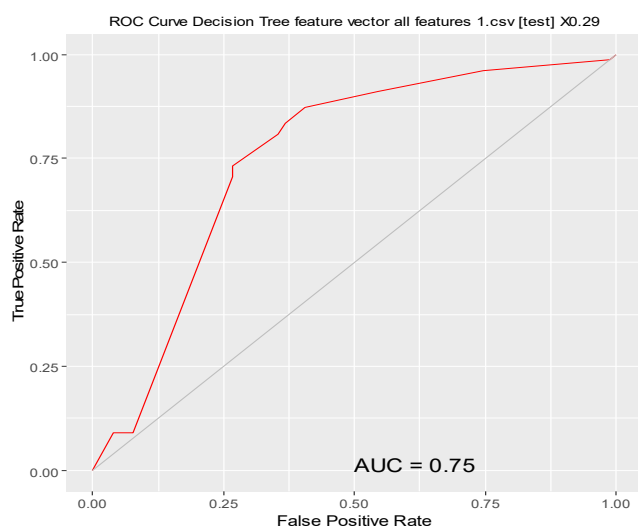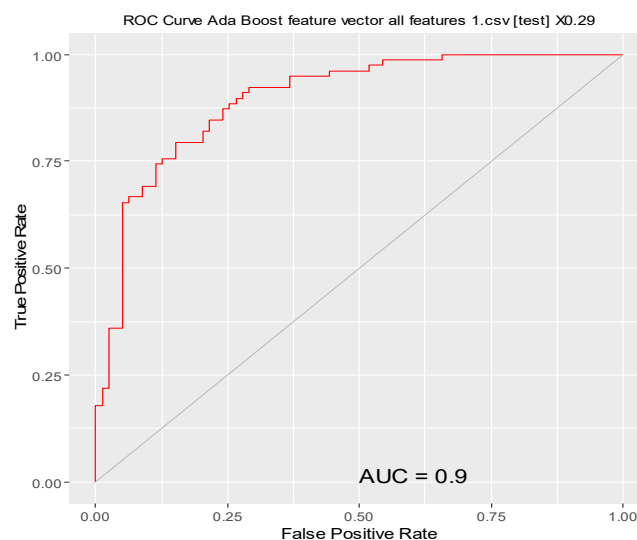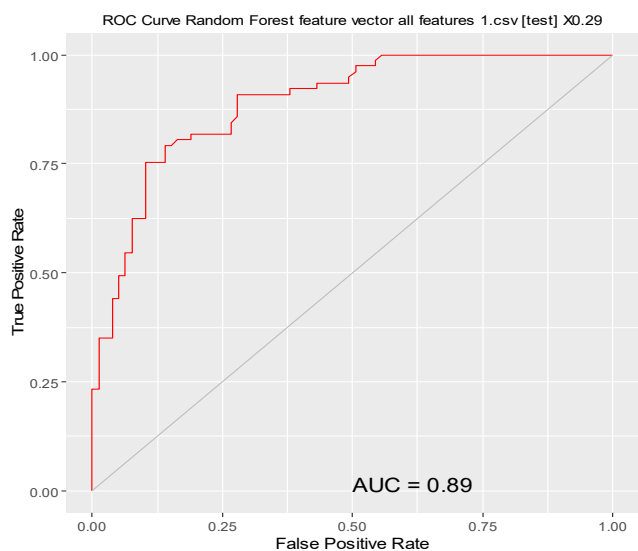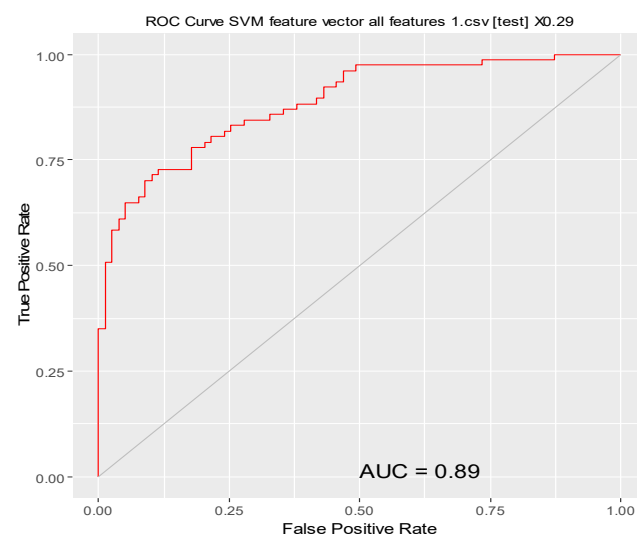

Fig. 4 Classifier performance
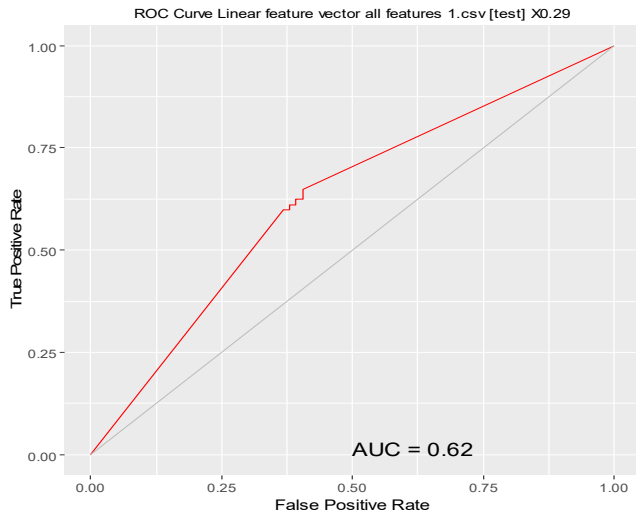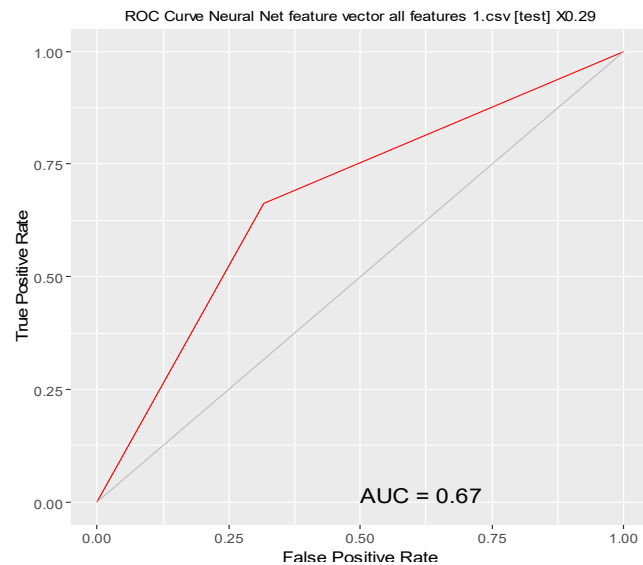

(a)

(b)

(c)

(d)

(e)

(f)

Fig. 5 Area under the curve for all classifiers (a) Binary Tree (b) Random Forest (c) Adaptive Boosting Models (d) Support Vector Machine (e) Generalized Linear Models (f) Neural Network

B. Comparison with Other Proposed Techniques

Numerous studies have been presented to determine diabetes and other related diseases by means of iridology. Table IV shows a comparative study of the proposed model with other existing models irrespective of sample size and classification accuracy.

TABLE IV
COMPARISON OF SAMPLE SIZE AND ACCURACY FOR EXISTING MODELS AND PROPOSED MODEL

Author	Disease	No. of Samples	Classification accuracy (%)
Wibawa and Purnomo [26]	Condition of pancreas organ	34	94
Lasmana et al. [10]	Pancreas disorder	50	83
Bansal et al. [7]	Diabetes	80	87.5
Sivasankar [27]	Pulmonary Disease	32	84.38
Hareva et al. [13], [28]	Healthcare Monitoring	32	90-95
Proposed model	**Diabetes**	**200**	**89.66**

IV. CONCLUSION

The potential of the diagnosis utility of the iridology along with machine learning modes had been studied. For the effective realization of feature based predictive analysis, it is essential to evaluate and compare various predictive modeling methods. The proposed technique has advantages over commercially available specialized diabetes diagnostic techniques as it is non-invasive and cost effective. At the same time, the benefit can be given to the remote areas where diagnosis facility is lagging for diabetes. Overall accuracy of 89.66% was achieved by the RF classifier. Despite these promising results, it is very essential to perform extensive study and investigation to validate iridology for diabetes diagnosis and to open an entirely new path for disease diagnostics.

REFERENCES

[1] J. Beagley, L. Guariguata, C. Weil, and A. A. Motala, "Global estimates of undiagnosed diabetes in adults," *Diabetes Res. Clin. Pract.*, vol. 103, no. 2, pp. 150–160, 2013.
[2] T. Middle and E. Journal, "Effectiveness of complementary and alternative medicine – Call for a 'black box' research agenda," *Middle Eur. J. Med.*, vol. 7, no. 8, pp. 239–240, 2005.
[3] P. E. Harris, K. L. Cooper, C. Relton, K. J. Thomas, and P. Harris, "Prevalence of complementary and alternative medicine (CAM) use by the general population : A systematic review and update," *Int. J. Clin. Pract.*, vol. 66, no. 10, pp. 924–939, 2012.
[4] V. Thomas, N. V. Chawla, K. W. Bowyer, and P. J. Flynn, "Learning to predict gender from iris images," *2007 First IEEE Int. Conf. Biometrics Theory, Appl. Syst.*, pp. 1–5, 2007.
[5] A. Bansal, R. Agarwal, and R. K. Sharma, "Predicting Gender Using Iris Images," *Res. J. Recent Sci.*, vol. 3, no. 4, pp. 20–26, 2014.
[6] S. Azilah and F. Paper, "Identification of vagina and pelvis from iris region using artificial neural network," *Teknologi*, vol. 76, no. 7, pp. 91–95, 2015.
[7] A. Bansal, R. Agarwal, and R. K. Sharma, "Determining diabetes using iris recognition system," *Int. J Diabetes Dev Ctries*, vol. 34, no. 4, pp. 432-438, 2015.
[8] L. W. Liam, Ali Chekima, L. Chung Fan and J. A. Dargham, "Tumor Detection using IRIS Pattern," 4th Annual Seminar of National Science Fellowship, 2004.
[9] A. Helwan, "ITDS: Iris Tumor Detection System using Image Processing Techniques," *International J. Sci. Eng. Res.*, vol. 5, no. 11, pp. 76–80, 2014.
[10] I. P. Dody, L. I. Ketut, E. Purnama, and M. Hery, "Abnormal Condition Detection of Pancreatic Beta-Cells as the Cause of Diabetes Mellitus Based on Iris Image," *2011 Int. Conf. Instrumentation, Commun. Inf. Technol. Biomed. Eng.*, pp. 150–155, 2011.
[11] J. F. Banzi and Z. Xue, "An Automated Tool for Non-contact, Real

Time Early Detection of Diabetes by Computer Vision," *Int. J. Mach. Learn. Comput.*, vol. 5, no. 3, pp. 225–229, 2015.

[12] 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.

[13] D. Hareva, S. Lukas, and N. Suharta, "The smart device for healthcare service: Iris diagnosis application," in *Eleventh International Conference on ICT and Knowledge Engineering*, 2013, pp. 1–6.

[14] B. Jensen, "Iridology Chart." (Online). Available: http://www.bernardjensen.com/. (Accessed: 01-Nov-2016).

[15] T. Zhu and Y. W. Zhu, Young, "Biometric personal identification system based on iris analysis," American Patent, 1994.

[16] J. G. Daugman, "High confidence visual recognition of persons by a test of statistical independence," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 15, no. 11, pp. 1148–1161, 1993.

[17] D. M. Rankin, B. W. Scotney, P. J. Morrow, and B. K. Pierscionek, "Iris recognition—the need to recognise the iris as a dynamic biological system: Response to Daugman and Downing," *Pattern Recognit.*, vol. 46, no. 2, pp. 611–612, 2013.

[18] J. Daugman, "Iris Recognition," *Am. Sci.*, vol. 89, no. 4, p. 326, 2001.

[19] J. Daugman, "How Iris Recognition Works," *IEEE teansactions circuits Syst. video Technol.*, vol. 14, no. 1, pp. 715–739, 2004.

[20] R. P. Wildes, "Iris recognition: an emerging biometric technology," *Proc. IEEE*, vol. 85, no. 9, pp. 1348–1363, 1997.

[21] B. Jensen and B. Jensen, *Iridology Simplified*, 5th ed. California USA: Iridologists International, 2011.

[22] C. Parmar, P. Grossmann, J. Bussink, P. Lambin, and H. J. W. L. Aerts, "Machine Learning methods for Quantitative Radiomic Biomarkers," *Nat. Publ. Gr.*, pp. 1–11, 2015.

[23] M. De Marsico, A. Petrosino, and S. Ricciardi, "Iris Recognition through Machine Learning Techniques: a Survey," *Pattern Recognit. Lett.*, 2016.

[24] M. Fern and E. Cernadas, "Do we Need Hundreds of Classifiers to Solve Real World Classification Problems?," *J. Mach. Learn. Res.*, vol. 15, pp. 3133–3181, 2014.

[25] M. Kuhn, "Building Predictive Models in R Using the Caret Package," *J. Stat. Softw.*, vol. 28, no. 5, pp. 1–26, 2008.

[26] A. D. Wibawa and M. H. Purnomo, "Early detection on the condition of Pancreas organ as the cause of diabetes mellitus by real time iris image processing," in *IEEE Asia-Pacific Conference on Circuits and Systems, Proceedings, APCCAS*, 2006, pp. 1008–1010.

[27] K. Sivasankar, "FCM based Iris Image Analysis for Tissue Imbalance Stage Identification," in *International Conference on Emerging Trends in Science, Engineering and Technology, Proceedings,* 2012, pp. 210–215.

[28] D. H. Hareva, B. B. Sitorus, S. Lukas, N. O. Suharta, C. Sciences, U. P. Harapan, and L. Karawaci, "Implementation of iridology application on smartphone," in *7th ICTS,* 2013, pp. 33–38.