Development of Diabetic and Related Disease Identification System Using IRIS

A Thesis

Submitted in the fulfilment of the requirement for the award of the degree of

DOCTOR OF PHILOSOPHY

in

ELECTRICAL AND INSTRUMENTATION ENGINEERING DEPARTMENT

Submitted by

Piyush Samant

(Registration No.:951404006)

Under the supervision of

Dr. Ravinder Agarwal

Professor, EIED, TIET

Electrical and Instrumentation Engineering Department
Thapar Institute of Engineering & Technology,
Patiala -147004, Punjab, India

Oct 2020

CERTIFICATE

I hereby certify that work which is being presented in the thesis entitled **"Development of Diabetic and Related Disease Identification System Using IRIS"** in partial fulfillment of the requirements for the award of degree of Doctor of Philosophy submitted in Electrical and Instrumentation Engineering Department, Thapar Institute of Engineering and Technology, Patiala is an authentic record of my own work carried out during a period from January, 2015 to November, 2019 under the supervision of Dr. RavinderAgarwal (Professor, Electrical and Instrumentation Engineering Department, Thapar Institute of Engineering and Technology, Patiala)

The matter embodied in this thesis has not been submitted by me for the award of any other degree of this or any other University.

(Piyush Samant)
Registration No.: 951404006

This to certify that the above made statement by the candidate is correct to the best of my knowledge and belief.

Dr. Ravinder Agarwal
Professor
Electrical and Instrumentation Engineering Department
Thapar Institute of Engineering and Technology, Patiala
Punjab-147001, India

ACKNOWLEDGEMENT

Before I express my heart-felt gratitude towards all my mentors, first and foremost, I would like to thank my greatest teachers: **God and my Parents**, who have given me the power to believe in myself and my passion to pursue this study. I could never have done this without their blessings.

I would like to express my gratitude to my supervisor, **Prof. Ravinder Agarwal** for motivation, patient guidance and every-time support. I am truly very fortunate to have the opportunity to work with him. I found his guidance to be extremely valuable. I am thankful to him that he has full confidence in my ability; this work would have not been completed without his vision and encouragement.

I would like to express my deepest gratitude to **Ms. Sakshi Bansal, Mr. Rohit Gupta, Dr. Atul Bansal, and Dr. Megha,** for their motivation and encouragement. I would also like to thank my all-fellow colleagues of **Biomedical Research Laboratory**, Electrical and Instrumentation Engineering Department, Thapar Institute of Engineering and Technology, Patiala for their encouragement and support. My sincere thanks to **Dr. Ish Padhi,** Rapid Laboratory, Patiala and all the subjects who helped in developing the database.

I am highly in debt to my sisters **Anjali** and **Varsha**, without whose unbiased sacrifice and support this work would not have been completed.

Although I have tried to express my gratitude to every person whose contributed in this work directly or indirectly, there may still be someone hiding behind the veils of my forgetful part of memory. Last but not the least; I would like to thank all such souls.

(Piyush Samant)

SUMMARY

The advancements in medical imaging and information technology have paved the way for the use of computers in diagnosis. It has also become essential to reduce the mortality rate by early detection of diseases. Computer-Aided Diagnostic (CAD) systems are being developed to diagnose diseases from different medical imaging modalities. Interpretation of these images presents challenges to the expert as the affected regions' patterns may look similar for one or more diseases. CAD systems have been developed to diagnose the disorders present in the lungs, brain, liver, and spinal cord. Physicians can use CAD systems as a second opinion in decision making and treatment planning.

Complementary and Alternative Medicine (CAM) techniques for the diagnosis are also prevalent. In the modern era of computers, medical practitioners believe that the combination of CAM techniques with CAD can accomplish medical science. Iridology is a CAM technique which is generally based on the concept of neural pathways between the body and the iris. The iridologist assessed the health condition, which interprets patterns, shapes, rings, colors and pigmentation markings, fibers, structures, and changes in the iris. In this thesis, an attempt is made to correlate the specific areas of the iris with diabetes. The related disease condition of the individual as diabetes is a chronic disease that occurs either when the pancreas does not produce enough insulin or when the body cannot effectively use the produced insulin. Insulin is a hormone that regulates blood sugar. Hyperglycaemia or raised blood sugar is a common effect of uncontrolled diabetes and, with time, leads to severe damage to many of the body's systems, especially the nerves and blood vessels.

This research aims to develop an efficient CAD system for automatic screening of diabetes and related diseases with high accuracy using machine learning such that an alternative diabetes diagnosis method can be proposed. Image-based CAD systems generally consist of seven phases-database creation, image pre-processing, segmentation, finding region of interest, feature extraction, optimizing the features, and classification for decision making.

For this research work, we have developed our data-set of infrared eye images along with meta-data. Meta-data includes the diabetes clinical blood test reports, details of all the physiological parameters, and patients' medical history. A supervised learning model for the

diagnosis of diabetes can be trained. The success of a computer-based system depends both on the features and classification method. An efficient set of textural features decides the accurate diagnosis, and an appropriate classification method provides the potential to produce a correct classification.

In the first research work, the diagnosis of diabetes is made based on the first-order statistical features and discrete wavelet features from the specific areas of the iris. The investigation was performed over a close group of 338 subjects (180 diabetic and 158 non-diabetic). According to the iridology chart, the region of interest from the iris image was cropped as the zone corresponds to the position of the pancreas organ. Statistical, textural, and discrete wavelength transformation features were extracted from the region of interest. The results show the best classification accuracy of 89.63% calculated from random forest classifier. Maximum specificity and sensitivity are absorbed as 0.9687 and 0.988, respectively. Results have shown the proposed model's effectiveness and diagnostic significance for a non-invasive and automatic diabetes diagnosis.

The second work presents a detailed comparative analysis of classification techniques to diagnose type 2 diabetes and its duration using the combination of features of a specific area of iris and physiological parameters. A set of 334 subjects is investigated. Subjects were divided into diabetic and non-diabetic groups. Moreover, the diabetic group was classified into three different subgroups according to the diabetic state's duration. Statistical features, Gray Level Co-occurrence Matrix (GLCM), and Gray Level Run Length Matrix (GLRLM) based features were extracted from the specific areas of iris. Nine classifiers of different application areas were selected. Subsequently, six parameters (accuracy, precision, sensitivity, specificity, F-score, and area under the curve) of each classifier have been computed and analyzed. The analysis provides promising results with more than 95% accuracy. Similarly, the same features have been further explored for another study to diagnose Diabetes Kidney Disease (DKD). This experiment achieved an accuracy of 94.5% with the combination of first-order statistical features, GLCM, and GLRLM based features.

In the previous study, the sclera-based person identification technique presented is a relatively new biometric technique and needs to explore more. The human eye's sclera contains the unique blood vessel patterns that make it a potential tool for personal identification. A fast and robust sclera segmentation algorithm and an impressive feature set

for recognition are presented. The most versatile data-set in image quality (UBIRIS V1) was selected for segmentation and recognition purposes. For sclera segmentation, an unsupervised algorithm is presented based on pixel mapping and cauterized grayscale. After that, from the segmented sclera, global features and Gabor wavelet transform based are extracted. Finally, for the recognition, a multi-class support vector machine classifier is applied. The recognition results are presented in terms of overall system accuracy and sensitivity. It is notionally proved that the proposed technique is reliable and accurate for executing sclera-based recognition.

TABLE OF CONTENT

Certificate i
Acknowledgement ii
Summary iii
Table of Content vi
List of Abbreviations ix
List of Figures xi
List of Tables xii

INTRODUCTION 1
1.1 Diabetes 1
1.1.1 Type 1 Diabetes 2
1.1.2 Type 2 Diabetes 3
1.1.3 Gestational Diabetes 3
1.2 Diagnosis of Diabetes 3
1.2.1 Blood Glucose Tests 4
1.3 Alternative Methods for Diabetes Diagnosis 5
1.4 A Generalized System for Image Based Disease Diagnosis 6
1.5 Motivation of the Study 7
1.6 Gap Identified 8
1.6.1 Improved Algorithm for Extracting the Iris and Sclera 8
1.6.2 Application in Bio-medical Interest 8
1.6.3 Need of Robust Feature Vector 8
1.7 Objectives 8
1.8 Layout of the Thesis 9

STATE OF THE ART 11
2.1 Diabetes and Its Prevalence 11
2.2 Diabetes and Physiological Parameters 14
2.3 Iris and Disease Diagnosis 15

2.4 Iris and Sclera Localization and Segmentation **17**
2.4.1 Iris Segmentation 18
2.4.2 Sclera Segmentation 21

DIAGNOSIS OF DIABETES USING IRIS IMAGES **23**
3.1 Introduction **23**
3. 2 Material and Methodology **25**
3.2.1 Subjects Selection and Data Collection 26
3.2.2 Iris Image Processing 27
3.2.3 Feature Extraction and Selection Methods 28
3.2.4 Classification 30
3.3 Analysis **31**
3.4 Results and Discussion **32**

DIABETES DIAGNOSIS USING IRIS-BASED FEATURES AND PHYSIOLOGICAL PARAMETERS **37**
4.1 Introduction **37**
4.2 Methodology **40**
4.2.1 Subject Selection and Division 40
4.2.2 Physiological Features 42
4.2.3 Eye Image Acquisition and Pre-image Processing 42
4.2.4 Extraction of Features From Iris 44
4.2.4.1 Statistical features 44
4.2.4.2 Gray Level Co-Occurrence Matrix (GLCM) features 44
4.2.4.3 Gray Level Run Length Matrix (GLRLM) features 45
4.2.5 Feature Selection 46
4.2.6 Classification 47
4.2.7 Statistical Analysis 48
4.3. Results and Discussion **50**
4.3.1 Performance Evaluation for PCA 50
4.3.2 Performance Evaluation of Modified t-test 54
4.3.3 Comparison With Existing Techniques 59

ASSESSMENT OF KIDNEY PROBLEM IN THE DIABETIC PATIENTS USING IRIS **63**
5.1 Introduction **63**
5.2 Methodology **65**
5.2.1 Data-set Collection 65
5.2.2 Pre-Image Processing Techniques 65
Figure 5.1 Block diagram of the adapted methodology 66
5.2.4 Feature Extraction and Experiment Design 67
5.2.5 Classification 67
5.3 Result and Discussion **67**
5.3.1 Classifier Analysis by FOS Features 67
5.3.2 Classifiers Performance Analysis by GLCM Features 67
5.3.3 Classifiers Performance Analysis by GLRLM Features 68
5.3.4 Classifiers Performance Analysis Using Combined Feature Vector 68

SCLERA SEGMENTATION AND RECOGNITION **72**
6.1 Introduction **72**
3.2 Proposed Method **73**
6.2.1 Data Set 73
6.2.2 Pre-Processing and Sclera Segmentation 73
6.2.3 Feature Extraction 74
6.2.4 Classification and Matching 75
6.3 Result and Discussion **75**

CONCLUSION AND FUTURE WORK **78**
7.1 Conclusion **78**
7.2 Future work **79**

LIST OF PUBLICATIONS **80**

REFERENCE **81**

LIST OF ABBREVIATIONS

IDF	International Diabetes Federation
DM	Diabetes Mellitus
WHO	World Health Organization
NCD	Non-Communicable Diseases
ICMR	Indian Council of Medical Research
ATP	Adenosine Triphosphate
FA	Fatty Acids
GDM	Gestational Diabetes Mellitus
FPG	Fasting Plasma Glucose
BMI	Body Mass Index
IDO	Integro-Differential Operator
CHT	Circular Hough Transform
NIR	Near Infrared
ANMBP	Neighborhood based Modified Back-propagation using Adaptive Learning Parameters
FRR	False Rejection Rate
SPS	Super Pixel Segmentation
LP	Laplace Pyramid
CDA	Circu-Differential Accumulator
WT	Watershed Transformation
DWT	Discrete Wavelet Transform
SVM	Support Vector Machine
LS-SVM	Least Square Support Vector Machine
RBF	Radial Basis Function
GDA	Generalized Discriminate Analysis
CAD	Computer Aided Diagnosis
ACE	Angiotensin Converting Enzyme
CAM	Complementary and Alternative Medicine
MPE	Molecular Pathological Epidemiology
HCC	Hepatocellular Carcinoma
IR	Infra-red
ROI	Region of Interest
GLCM	Gray Level Co-occurrence Matrix
GLRL	Gray Level Run Length
AI	Artificial Intelligence
DIP	Digital Image Processing
FV	Feature Vector

PCA	Principal Component Analysis
SRE	Short Run Emphasis
LRE	Long Run Emphasis
GLN	Gray level Non-uniformity
RP	Run Percentage
RLN	Run Length Non-uniformity
GRE	Low Gray Level Run Emphasis
HGRE	High Gray Level Run Emphasis
FN	False Negative
AUC	Area Under the Curve
TPR	True Positive Rate
FPR	False Positive Rate
MCC	Matthews correlation coefficient
SEV	Specific Explained Variance
LDA	Linear Discriminate Analysis
CC-NN	Cascade Correlation Neural Network
DN	Diabetic Nephropathy
DKD	Diabetic Kidney Disease
GFR	Glomerular Filtration Rate
MSVM	Multiclass Support Vector Machine
GWT	Gabor Wavelet Transformation
LBP	Local Binary Patterns

LIST OF FIGURES

Figure 1.1	Prevalence of Diabetes Estimated by IDF
Figure 1.2	A generalized block diagram of an image-based disease diagnosis system
Figure 3.1	Summarized methodology for diabetes diagnosis using iris
Figure 3.2	Percentage distribution of diabetic subject as duration of diabetic state
Figure 3.3	Classifier performances according to duration of diabetic state
Figure 4.1	Proposed methodology
Figure 4.2	Comparison of accuracies of classifiers with different SEV (in %) of PCA
Figure 4.3	Performance comparisons of the accuracies of classifiers with different number of features ranked by the modified t-test
Figure 5.1	Block diagram of the adapted methodology
Figure 5.2	Extraction of Region of Interest (ROI)
Figure 5.3	Block diagram of the experiment designed for combined feature vector
Figure 6.1	Flowchart of proposed methodology
Figure 6.2	Flow chart of the sclera segmentation algorithm
Figure 6.3	Graphical user interface designed for visualization of segmentation and recognition results
Figure 6.4	Comparisons of Accuracy and Sensitivity of the LBP-PCA and GWT features

LIST OF TABLES

Table 2.1	Summary of the studies by the IDF for status of diabetes
Table 2.2	A summarized iris based disease diagnosis system
Table 3.1	Detailed subject distribution
Table 3.2	Extracted statistical features
Table 3.3	Extracted texture features
Table 3.4	Feature selection methods and specification
Table 3.5	Classification accuracies of different classifiers against feature selection methods.
Table 3.6	Specificity and sensitivity of different classifiers against best accuracies
Table 3.7	Comparison of proposed model with existing models
Table 4.1	Distribution of diabetic subjects according to the duration of diabetic state
Table 4.2	Classification and detailed distribution of subjects according to the gender and age
Table 4.3	Physiological features and their corresponding types
Table 4.4	Extracted GLCM features
Table 4.5	Comparison of different SEV of PCA for average accuracy and precision
Table 4.6	Comparison of different SEV of PCA for the average sensitivity and specificity
Table 4.7	Comparison of different SEV of PCA for the average F-score and AUC
Table 4.8	Comparison of average accuracy and precision of different number of features ranked by the modified t-test
Table 4.9	Comparison of average sensitivity and specificity of different number of features ranked by modified t-test
Table 4.10	Comparison of average F-score and AUC of different number of features ranked by the modified t-test
Table 4.11	Performance comparison with existing techniques
Table 5.1	Average accuracy, sensitivity and specificity for FOS features
Table 5.2	Average accuracy, sensitivity and specificity for GLCM features
Table 5.3	Average accuracy, sensitivity and specificity for GLRLM features
Table 5.4	Average accuracy, sensitivity and specificity using combined feature vector

Chapter 1

Introduction

1.1 Diabetes

Diabetes Mellitus (DM) is one of the most common metabolic diseases and is characterized by the glucose tolerance capability of the body. A continuous increased blood glucose level can leads to microvascular and macrovascular complication and increases the risk of heart, eyesight, kidney and nerve system damage [1]. In the past few decades as lifestyle of the population changed, the prevalence of the diabetes is increased with a faster rate. International Diabetes Federation (IDF) estimates the prevalence of the diabetes worldwide as shown in Figure 1.1 [2].

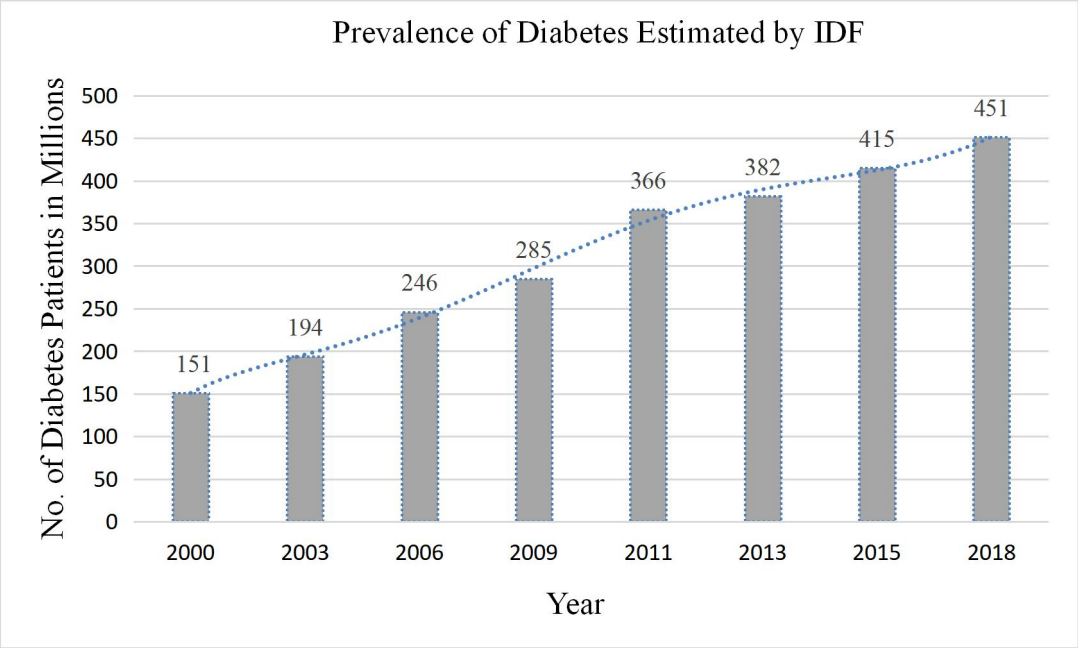


Figure 1.1 Prevalence of Diabetes Estimated by IDF

Our body converts the carbohydrates of the food into glucose, which provides energy for work to different parts of the body. When the body does not consume enough glucose, insulin helps to absorb the excess glucose to reduce blood glucose levels. Insulin is an important hormone which secretes by the Beta (β) cells of the pancreas organ [3]. Insulin consumes the glucose with the help of three different processes named as Glycolysis, Glycogenesis, and

Lipogenesis. Glycolysis is the process of converting glucose into the Adenosine Triphosphate (ATP). ATP is an electrochemical form that stores energy [4]. Glycolysis is an irreversible process which, converts glucose into glycogen, i.e., short term energy storage in the liver or muscles. It's a reversible process. Lipogenesis is the process of converting glucose into lipids and Fatty Acids (FA), to store the energy long term. It's an irreversible process and remains as adipose tissues or fatty layers in the body. Also, when the glucose level in the body is too low, the pancreas organ secretes another hormone called glucagon. Glucagon helps the liver to release stored glucose. Glycogen undergoes a reversible process of glycogenolysis, which releases glucose from the glycogen. Glucose also is made by another reversible process called gluconeogenesis, in which amino acids are converted into glucose. Together, insulin and glucagon maintain blood glucose in the body, and this state is called homeostasis. The excess of glucose level in the body is called Hyperglycemia, and the shortage is termed as Hypoglycemia. If the blood glucose concentration is lower than 70 Mg/DL, the person will have Hypoglycemia. It results in many disorders, including awkwardness, the dilemma in talking, bewilderment, unconsciousness, seizures, or death. A continuous hunger, sweating, unsteadiness, and weakness can also be seen in the patients suffering from Hypoglycemia. Any person is called Hyperglycemia if the blood glucose concentration is over 120 Mg/DL after overnight fasting. If Hypoglycemia remains for a longer period, it results in eye, nerve, and kidney disease; in a more generalized form, this condition is called diabetes. Diabetes is broadly classified mainly into 3 categories according to the function of insulin, i.e., type 2 diabetes, type 1 diabetes, and gestational diabetes.

1.1.1 Type 1 Diabetes

Type 1 diabetes occurs when β cells of pancreas organ does not produce enough insulin to consume the glucose. Any person suffering from type 1 diabetes has to take insulin time to time to avoid complications of having high blood glucose level for a longer period of time. Around 10% diabetic cases are of type 1 diabetes and it is a childhood chronic disease. Type 1 diabetes is the result of an auto-immune reaction, in which β cells of the pancreas organ are destroyed by the immune system. The main reason of type 1 diabetes is hereditary. Life style is not the reason of type 1 diabetes but in order to maintain the type 1 diabetes maintaining a healthy lifestyle is must. Till now type 1 diabetes cannot be prevented but with a healthy lifestyle and with proper medication and timely testing it can be maintained. An increase in thirst and urination, increased hunger, fatigue, sudden weight loss, and, blurred vision are

some of the major symptoms of the type 1 diabetes [5]. Over the period of time type 1 diabetes may leads to heart disease, kidney disease, nerve damage, foot problem, eye problem, depression *etc*.

1.1.2 Type 2 Diabetes

Type 2 is a ubiquitous form of diabetes in which the body becomes resistive or insensitive to the insulin, so the blood glucose level remains high for a longer period. Type 2 is the most common form of diabetes and around 85%-90% of diabetic patients having type 2 diabetes. Type 2 diabetes is a progressive disease and directly associated with the modified lifestyle, genetic, and family-related risk factors. Typically type 2 diabetes is seen in adults over the age of 45 years, but nowadays, many cases among the children, young adults, and adolescents also been diagnosed. If a person is living with a modifiable lifestyle, including high blood pressure, obesity, and poor nutrition, the possibility of having type 2 diabetes greatly increases [6]. Symptoms of type 2 diabetes are excessive thrust, fatigue, hunger, skin infection, and blurred vision *etc.*

1.1.3 Gestational Diabetes

Gestational Diabetes Mellitus (GDM) occurs during pregnancy and usually disappears after giving birth. Gestational diabetes can happen at any stage of time but mostly found in the second half, which is after 5th month. GDM generally does not show any specific symptoms, although some women feel fatigued, increased thirst, and dry mouth. However, these symptoms are common during the pregnancy period. Therefore, the only way to diagnosis GDM is a regular blood sugar test. Most of the women having GDM give birth to a healthy baby, but in some cases, the baby can grow more significant than the normal, which can increase the difficulties during delivery or maybe a cause of premature delivery. GDM increases the risk to women to have type 2 diabetes [7].

1.2 Diagnosis of Diabetes

DM is a combination of metabolic diseases caused by the extended time high blood glucose level. Diabetic patients suffer from one or more complications due to damage to the main functional organs by chronic Hyperglycemia. It's been estimated that complications in the functional organs of diabetic patients start accruing 5 to 10 years before the actual diagnosis.

Therefore, early diagnosis of diabetes is essential to prevent the damage and degrade the complications due to diabetes.

Clinical tests based on blood, urine, saliva, or sweat as the test samples are used for the diabetes diagnosis. Other than these modalities interstitial fluid, also uses ocular fluid and breathe as a reliable tool for the diabetes diagnosis. All of these physiological fluids have different glucose concentrations. It was observed that urine and sweat contain less glucose concentration. Despite the non-invasive method, studies related to these have been omitted, and in the present scenario, it is not practiced clinically for the diabetes diagnosis.

1.2.1 Blood Glucose Tests

The concentration of the blood glucose level is calculated by the blood glucose tests to check diabetes. There are different clinical tests by which screening of diabetes can be done.

Glycated Haemoglobin (A1C) test is a standard blood glucose test which does not require any fasting and specific preparation for the test. It indicates the blood sugar level of the last two-three months. The percentage of blood glucose is directly attached to the hemoglobin, which is the percentage of oxygen-carrying the protein in the red blood cells. Blood glucose level is directly associated with the hemoglobin attached to it and is directly prostrate. If the A1C level during two different tests comes out as 6.5 or higher, the subject will be diabetic. It the A1C value comes out below 5.7, it means the subject is healthy.

Moreover, if the test values come out between 5.7 and 6.4, then the subject is pre-diabetic or at a higher risk of having diabetes. Although, A1C test is a straightforward and handy method for diabetes diagnosis, the results from this test are not considered consistent. In particular specific medical conditions, the results may be biased and make the A1C method inaccurate. This method is not so popular among medical practitioners.

In the random blood glucose test, the blood sample is taken at any time, i.e., regardless of the food taken. The blood glucose level calculated as 200 milligrams per decilitre (mg/dL) or higher the subject will say to be diabetic.

In the fasting blood glucose test, the blood sample is taken in the morning after overnight fasting of at least twelve hours. If the fasting sugar level comes out less than 100 mg/dL, then

it is normal. If it comes out in between 100 to 125 mg/dL, then the subject will be called pre-diabetic and for above 125 mg/dL.

Oral blood glucose test is a standard test to screen diabetes; it is a popular and clinically accepted. In this test, the first blood glucose concentration is measured after overnight fasting. Then a sugary drink is fed, and after periodically two hours, blood glucose tests have been tested. If glucose level comes out less than 140 mg/dL, then the subject is healthy. If the glucose level is less than 140 mg/dL and greater than 200 mg/dL, than it is a healthy and diabetes condition, respectively. If glucose level comes out in between these two readings, it is called a pre-diabetic state.

1.3 Alternative Methods for Diabetes Diagnosis

Complementary and Alternative Medicine (CAM) therapies are quite famous for chronic diseases like diabetes, arthritis, *etc*. and became very popular in the past few decades. Many mainstream medical practitioners believe these therapies are unorthodox, but in many countries, CAM is used widely for health-care applications. Harris *et al.* [8] found that around 52% and 38% general population are using CAM in Australia and the USA, respectively.

Iridology is an alternative medicine science that correlates iris patterns, colors, breakage, tissue weakness, and other characteristics, which can acquire evidence about a patient's systemic health. It reveals weakness or braking in tissues long before the symptoms appear. Iridology practitioners match their interpretations of the iris chart in which iris is divided into several zones corresponding to specific body organs [9].

The history of iridology was started from the 19th century, Hungarian physician Ignatz Von Peczely observed dark-colored streak in the iris of an owl with a broken leg is almost similar of a man with a broken leg [10]. After that, he started observing and found that that the dark streak grew blur and disappeared as man's leg healed. Later in 1950 onwards, Dr. Bernard Jensen, an American naturopath, took iridology forward. He proposed the iridology charts for the researchers and practitioners. They have explored that the minute details of the iris contains handy features and have information about the primary health condition.

1.4 A Generalized System for Image Based Disease Diagnosis

A generalized image based disease diagnosis system is also called Computer Aided Diagnosis (CAD) system. In a CAD based system a sequential combination of image processing techniques and supervised machine learning technique are applied. On the basis of the learning of the classifier from the labelled images the prediction are made on unknown images regarding disease. Figure 1.1 shows a basic idea of a generalized image based disease diagnosis model.

The first step in an image-based diagnosis system is a database collection of the images. The selection of imaging modalities is also crucial and depends on the kind of problem undertaken.

The next step after image acquisition is pre-image processing to improve the image data [11]. Pre-image processing techniques are used to remove the noise and for image enhancement. After that, image segmentation of Region of Interest (ROI) [12–14] is performed, which is an integral part of the pre-image processing technique. Image segmentation is an essential operation for the meaningful analysis and interpretation of the acquired image. It is a crucial and essential component of image analysis and diagnosis system. Segmentation is one of the most complicated tasks in image processing, which determines the accuracy of the final diagnosis. Notably, while dealing with the medical images where the information is to be explored only from a specific ROI, for example in the iris-based diagnosis the central region of interest is iris part, so segmentation technique has to be applied to segment out the iris from the complete image of the eye.

After segmentation of the ROI, significant features are to be calculated using mathematical measurements. In the images based disease diagnosis system, features are classified into different categories like shape and size based features, statistical features, color-based features, texture-based features, *etc*. For example, to study breast cancer and skin cancer, shape-based features are explored [15, 16]. Likewise, for tissue-related studies, statistical features are explored. Appropriate feature extraction helps in the elimination of unnecessary, redundant features from the dataset, which helps to give a fast and accurate result.

Classification for prediction is a well-known data mining technique that uses training data to build up a model based on the labeled data set. After that, the trained model is applied to the testing data to predict the disease. Numerous classification algorithms are proposed for image-based disease diagnosis model [17].

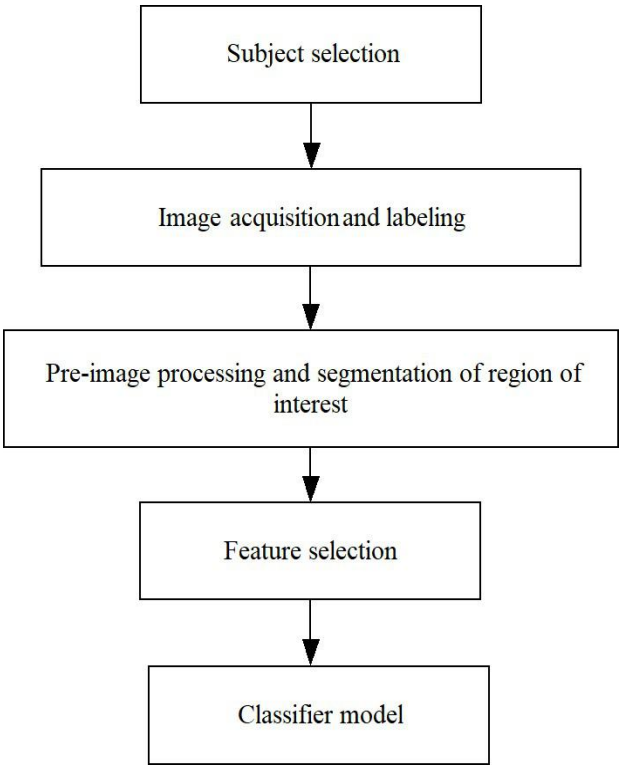


Figure 1.2 A generalized block diagram of an image-based disease diagnosis system

1.5 Motivation of the Study

The present lifestyle of people is not healthy due to work pressure and ignorance of the small health problems, which is very harmful in the long term. Chronic disease like diabetes is a slow poison for the body which harm the major body organs slowly and results in multi-organ failure. Worldwide almost all countries are facing challenges in the diagnosis of diabetes. The developed nations have proper diagnostic facilities; however, advanced diagnostic facilities are not readily available in undeveloped and developing nations. In India, more than 60% of the population lives in rural areas. Therefore, there is a need for a simple and effective diabetes diagnosis method [18]. The modern machine learning techniques have opened the door the unorthodox diagnosis methods. Motivation is to explore these unorthodox methods with modern machine learning techniques so that a novel and accurate diabetes diagnosis system can be proposed.

1.6 Gap Identified

Based on the literature survey, some gap in the studies has been identified as:

1.6.1 Improved Algorithm for Extracting the Iris and Sclera

Iris of human eye contains rich information about physiology of human health, to extract iris from the human eye is most important step in this context. The various researchers already propose different algorithms. There is scarcity of comparative analysis of existing algorithms. Based on parameters, more efficient protocol need to be proposed. Sclera the white color part of eye also contains fruitful parameters for physiology of human health, but still not efficiently exploited yet by the researchers.

1.6.2 Application in Bio-medical Interest

Most of the research work in iris recognition systems are mainly focused on biometrics. Few models had been proposed for clinical applications to diagnose the disease. So, it is an open area of research for the identification of diseases at discrete levels like diabetes, jaundices, *etc*.

1.6.3 Need of Robust Feature Vector

The model for disease identification should be robust so that any general person can use it without consulting a medical practitioner. The algorithm should be robust enough so that the effects of artifacts and biases (uneven light effect, shaking of hand /camera, low-quality images, *etc*.) can be rectified. The existing techniques lack in terms of the robustness of the feature vectors.

1.7 Objectives

a) To develop an efficient algorithm to extract iris and sclera from captured image

b) Development of robust feature of iris images for decision making environment

c) To apply model for identification of diabetes and other related disease

d) A comparative analysis with the existing method

1.8 Layout of the Thesis

In Chapter 2, the literature part has been discussed. This chapter is divided into four sections. The first section discusses diabetes and prevalence worldwide. In the second section, the work related to diabetes and its relationship with the physiological parameters are discussed. In the third section, the studies done concerning the relationship between iris and disease diagnosis are explored. In the last section of this chapter, studies related to the segmentation of iris and sclera, parts are discussed.

The third chapter evaluates the diagnostic validity of iridology for the diagnosis of type 2 diabetes using soft computing methods. The investigation was performed over a close group of diabetic and non-diabetic subjects. The experiment is designed in such a way that maximum information from the iris related to the diabetes state can be extracted, and a standard set of steps can be predefined for this kind of disease diagnosis system. Also, the proposed study opens the door for the new non-invasive and accurate diabetes diagnosis system.

Chapter 4 presents a detailed comparative analysis of machine learning-based classification techniques to diagnose type 2 diabetes using the combination of iris-based features and physiological parameters. This study shows that the combined Feature Vector (FV) of the iris-based features and physiological features gives more accurate and sophisticated results. At the same time duration of the diabetic state is also considered for the analysis so that the proposed model can perform the more robustness and variability.

In chapter 5, a pilot study for the diagnosis of diabetic kidney disease (DKD) is presented. The significant steps of the generalized image-based diagnosis system are used. The primary purpose of this study is to explore more related diseases to diabetes and to verify the standard procedure used in chapter 3 and 4 for this type of iridology based diagnosis system.

Chapter 6 is an additional study of the sclera segmentation and recognition system. In this chapter, a robust algorithm for the sclera segmentation is presented. Additionally, the contribution of the global features for person recognition is presented.

Chapter seven provides a few broad conclusions to the proposed work by highlighting the significant findings and contributions, along with a discussion of possible future research

direction which arises from the research work undertaken in this thesis. This process aims to simplify the correlation between the image regions of interest and iridology charts, as well as to eliminate the errors introduced by the human factor.

Chapter 2

State of the Art

In this chapter, the relevant literature has been consolidated. This chapter is divided into three sections. In the first section, studies related to diabetes and its prevalence has been discussed. In the second section, studies related to the role of iris on disease diagnosis, especially diabetes, have been described. The last section presents studies related to iris and sclera localization and segmentation.

2.1 Diabetes and Its Prevalence

A vast literature has been presented to study the status of diabetes globally as well as in India. The International Diabetes Federation (IDF) and the World Health Organization (WHO) keep a keen eye on the prevalence of diabetes and adopted a transparent procedure to estimate the progression of diabetes worldwide and different countries. Guariguata *et al.* [19] have been presented a detailed study prevalence of diabetes for 2013 and estimated for 2035 in the adults of age between 20-79. The study was performed over 744 data sources of 130 countries. It has been observed that there will be a considerable increment in the diabetes population, especially in developing countries. Another interesting comprehensible study has been presented by the Noncommunicable Diseases (NCD) Risk Factor collaborator group [20] to presents the trend of diabetes since 1980 worldwide. The study was performed over 4.4 million participants from 751 population-based studies. In both studied, it was found that diabetes has increased more rapidly in developing and undeveloped countries as compared to developed or high-income countries. Studies have been found that diabetes cases in adults were increased to 422 million in 2014 from 108 million in 1980. The minimum growth rate was observed in north-west Europe while northern Africa, middle-east and middle Asia had shown maximum growth in the prevalence of diabetes.

Fermandes *et al.* [21] estimated the expenditure on diabetes and found that about $ 162 billion had been spent on diabetes globally. The study concludes that diabetes is a substantial economic burden on the global health-care system, which needs a proper and effective management system for public health to reduce financial burden.

Various models have been presented so far to estimate the diabetic population in the future. Ogurtsona *et al.* [2] Performed an exhaustive study to predict the diabetic population for 2040, according to the trend up-to 2015. All the studies related to the diabetes population have been included, and then using the analytic hierarchy process, most relevant studies have been selected from all the countries. After that, extrapolation has been applied over the selected data to make the prediction model, and finally, a logistic regression model has been applied to generate the estimate of the diabetic population. The model has been predicted that around 643 million population will be diagnosed diabetic by 2040. A similar kind of study has been performed by Whiting *et al.* [22] to predict the diabetic population in 2030 from the available database of 2011. Authors have been estimated that around 522 million people will be affected by diabetes by 2030. Every year IDF contributes to review studies by the name of diabetes atlas in high impact journals those estimates the prevalence of diabetes in the future. From 2009, a total of 14 global and 7 regional studies have been published so far. Table 2.1 summarizes all the above mention studies by the IDF.

Table 2.1 Summary of the studies by the IDF for status of diabetes.

Sr. No.	**Description**	**Year**
Global Studies		
1	A survey on diabetes was conducted by IDF and WHO on the world population, and estimation was made for 2010 and 2030. It was observed that the prevalence of diabetes will be 6.4% in 2010 and will hike to 7.7% by 2030 [23].	2009
2	On the basis of different peer reviewed available literature, data of diabetes was collected by IDF. Methodology had been proposed using logistic regression to estimate the prevalence of diabetes [24].	2011
3	Diabetes data was collected using total 565 data sources from 110 different countries for 2011 and estimated the diabetes population for 2030. Study observed that undeveloped and underdeveloped countries will face maximum increment in diabetes [22].	2011
4	Burdon of metabolic risk on the world population due to diabetes had been estimated [25].	2012
5	An update of world diabetic population has been published [26]	2012

6	With the help of available literature, diabetic population of 2013 had been calculated and using logistic regression estimation was done for 2035. Total 744 data sources from 130 countries had been included for this study [19].	2013
7	An global estimation of type 1 diabetes in children had been performed, it was found that around 500,000 children of age less than 15 years diagnosed as diabetic type 1. Region wise study also been performed and found that European children are more affected by type 1 diabetes [27].	2013
8	On the basis of available literature, IDF estimated the undiagnosed case of diabetes and found that around 45% cases of diabetes remained undiagnosed worldwide [28].	2013
9	Guariguata *et al.* [29] estimated the world wide cases of hyperglycemia in pregnancy due to diabetes.	2013
10	Linnenkamp *et al.* [7] proposed methodology to estimate the hyperglycemia in pregnancy.	2013
11	First documented work published by IDF to estimate the global deaths attributable due to diabetes in 2010 and 2011. Study had shown that maximum increment found in Western Pacific areas due to diabetes [30].	2013
12	An updated estimation by IDF estimate the global deaths attributable due to diabetes in 2015 and compared with previous studies [31].	2015
13	IDF reviewed total 540 data sources from 111 countries and estimated the diabetes population for 2015 and 2040 [2].	2016
14	IDF estimated the total expenditure on diabetes for the year 2014 globally. It was found that total 1583 to 2842 USD on diabetes[21].	2017
Regional Studies		
1	IDF estimated the diabetes cases in Africa for year 2013 [32].	2009
2	IDF estimated the diabetes cases in Europe for year 2013 [33].	2009
3	IDF estimated the diabetes cases in Middle-East and North Africa for year 2013 [34].	2009
4	IDF estimated the diabetes cases in North America and Caribbean Region for year 2013 [35].	2009
5	IDF estimated the diabetes cases in South and Central America for	2009

	year 2013 [36].	
6	IDF estimated the diabetes cases in South-East Asia and Caribbean Region for year 2013 [37].	2009
7	IDF estimated the diabetes cases in Western Pacific Region for year 2013 [38].	2009

As it is already discussed that mostly the undeveloped and developing countries are most affected by the diabetes. India is also one of them, and diabetes is not homogeneous in India. At present, around 41 million Indians have been diagnosed as diabetic, but and there is a vast difference between urban and rural populations [39]. Therefore, undoubtedly diabetes is one of the primary healthcare problems in India. According to the report by the Indian Council of Medical Research (ICMR), middle, east, and south Indian states have more cases of diabetes as compared to northern India [40]. Monica *et al.* [18] already shown that the condition of diabetes will become worst in India in the near future, and lifestyle in the urban areas is a significant factor for this. A developing country like India will have to face significant emotional, economic, and social burdens to deal with diabetes. Seema *et al.* [41] investigated that Body Mass Index (BMI) plays a significant role in diabetes, and adults having low BMI have a high risk of being diabetic.

2.2 Diabetes and Physiological Parameters

Physiological parameters along with modern computer methods have shown a great potential for the risk assessment of the diabetes. Modern machine learning techniques have achieved a very high accuracy in diagnosing the diabetes using physiological parameters. Hasan *et al.* [42] performed a comparative analysis on pima Indian database for diabetes diagnosis. Age, BMI, blood-pressure, glucose concentration level and other physiological parameters had been used as features to train a multi-layer neural network for the diabetes diagnosis. Pima Indian is only database so far to have information about above mentioned physiological parameters related to diabetes health of an individual. A number of studies [43–46] have been presented models to predict diabetes using pima Indian database. Dwivedi presented [46] a comparative analysis of different classifiers to predict the diabetes using pima Indian database. Polat *et al.* [43] classify the diabetes and non-diabetes classes of UCI diabetes data-set. For feature selection purpose Principal Component Analysis (PCA) and for classification purpose Adaptive Neuro-Fuzzy inference System (ANFIS) has been utilized. Authors have

achieved as improved accuracy of 89.47%. Polat *et al.* [44] proposed a cascade technique to diagnosis diabetes using pima Indian data base. A two stage based on system Generalized Discriminate Analysis (GDA) and Least Square Support Vector Machine (LS-SVM) have been proposed. GDA used to analysis the discriminate between features while LS-SVM used to classify the two classes of diabetic and non-diabetic people. Heydari *et al.* [47] collected the physiological diabetic data of a huge population (2536) of Iran and presented a exhaustive comparative study of classification algorithms. Physiological parameters, along with modern computer methods, have shown excellent potential for the risk assessment of diabetes.

2.3 Iris and Disease Diagnosis

The iris-based diagnosis is an interesting and reliable tool in biomedical imaging. It provides an access to diagnosis even those chronic diseases which either hard to diagnose with pathological techniques. Liljequist [48] had identified changes in iris and correlated those changes with different illnesses. Soediono [10] investigated that earlier, the practitioners examined by the microscopes, magnifying glasses, and other equipment to observe iris tissue change and examine the health status of an individual. Iridology is an alternative diagnostic technique which analyzes the iris colors, patterns, shape, and other tissue construction like characteristics to predict the health condition of the individual. Jensen [49] divided iris into 80-90 zones corresponding to different body organs. Salles *et al.* [50] found that literature for iridology is quite limited, and it's essential to have more grim and extensive studies for this.

Jae Young *et al.* [51] explained that iris could help as a supplementary source to diagnose the medical condition, the authors have investigated the connection between iridological constitutions and Angiotensin Converting Enzyme (ACE) polymorphism in hypertensive. In the modern era of Computer-Aided Diagnosis (CAD), many researchers used the basics concept of iridology along with computer-based diagnostic techniques for disease diagnosis. Qiu. Sun *et al.* [52] proposed a model to predict ethnicity using iris images with the help of the Ada-Boost algorithm and achieved an accuracy of 85.95%. The model attempts to predict whether a subject is Asian or not. Thomas *et al.* [53] applied machine learning methods to develop models, whose predict gender based on the texture features of the iris. Bansal *et al.* [54] utilized the CHT model and Support Vector Machine (SVM) algorithm for predicting gender.

Table 2.2 A summarized iris based disease diagnosis system

Author	Application	Pre-image processing techniques	Remarks
Jae young *et al.* [51]	Relationship between iris structure and ACE gene polymorphism	For the assessment a system for iris analysis Bexel Irina was developed	Abdomina connective tissue weakness, Neurogenic, Cardio-renal connective tissue weakness features have been extracted from iris Total 166 subjects were investigated
V. Thomas *et al.* [53]	Predicted gender using iris	Canny Edge Detector and Circular Haugh Transformation used for segmentation and texture encoding Daugman's rubber sheet model, FFT, Gabor Filter for normalization Geometric and Texture features were studied	LG EOU200 system for data acquisition, 10-fold cross validation method was used C4.5 decision tree, (initially on SVM and Neural Network)
Atul Bansal *et al.* [54]	Predicted gender using iris	The CHT was used for segmentation, Daugman’s rubber sheet model for normalization Combines 2-D DWT based features with statistical features	I-SCAN-2 iris scanner used for data acquisition Accuracy of system was compared with Polynomial, Gaussian and RBF kernel functions of SVM classifier
Lin Ma *et al.* [55]	Gastrointestinal diseases	Manual blind clustering has been applied Manual classification was performed by six different persons having some medical background	Deformation in the geographical features due to gastrointestinal disease Pupil based features (pupil roundness and pupil largeness) were extracted
Passarella *et al.* [56]	Identified colon disorder	CHT used for localization and simple edge detection used for	Total 60 student volunteers were evaluated for the experiment

		segmentation	92% accuracy was achieved to classify colon disorder and results have been reconfirmed by the clinical test
Hussein *et al.* [57]	Diagnosed kidney disease	Histogram Equalization to normalize the contrast of the image CHT for localization and segmentation	Neuro-fuzzy classifier was used Wavelet, entropy and energy of the specific regions of the iris have been extracted
R. Ramlee *et al.* [62]	Cholesterol present in the blood vessels	Iris recognition models used for localization and segmentation	To determine the cholesterol Ostu threshold method has been used

Lin Ma *et al.* [55] studied geometrical structure changes in the iris patterns due to gastrointestinal diseases. Manual clustering, followed by the manual classification, had been performed to explore geometrical structure-based features of the iris. Passarella *et al.* [56] explored the specific iris regions to identify colon disorders. Hussein *et al.* [57] diagnosed kidney disease with the help of iris images. A total of 166 people was evaluated for the assessment. The authors had extracted wavelet, entropy, and energy features and applied neuro-fuzzy to classify. Ramlee *et al.* [58] estimated the cholesterol in the blood vessel by using iris images. John Daugman's [59] and Libor masek's [60] iris recognition methods have been applied for the localization and segmentation of iris. Features of the iris and iridology chart, a MATLAB program, was made to identify the presence of cholesterol in the body. However, further analysis performed to estimate the exact range or level of cholesterol in blood vessels [58, 61, 62]. Table 2.2 represents a summarized study of the iris-based disease diagnosis system.

Atul Bansal et al. [63] presented an SVM based model to determine diabetes, using Gabor filter and 2-D DWT based features. The authors have performed a close group study on 80 subjects and performed the analysis of the different kernels of the SVM classifier with the combination of wavelets. The authors have achieved 87.5% accuracy in classifying the diabetic and non-diabetic subjects using the radial basis kernel of SVM. Lesmana et al. [64] proposed a computer method based iris assessment technique to detect insulin deficiency from beta cells secreted from the pancreas. Neighborhood-based Modified Back-propagation

using Adaptive Learning Parameters (ANMBP) technique has been used to classify the abnormalities in the iris due to insulin deficiency using quantitative and textural features. J. F. Banzi et al. [52] evaluated the broken tissues and color change of the iris for early detection of diabetes. Salles et al. [53] reported a longitudinal study to observe the change in the sign of the cross of Andreas in the specific areas of the iris due to Diabetes Mellitus (DM) for four years. Although Iridology is very useful for diagnosing chronic diseases, few negative conclusions have also been reported regarding the iridology study. Earnst [162] has reported that a high false-positive rate in the manual diagnosis indicates some kind of suspect related to the iridology study. In another study,Cockburn [163] attempted to validate the medical diagnosis with Iridology in 15 subjects. The author has concluded that it does not appear to have validity in the context of conventional medicine. Knipschild [164] showed that Iridology is not a useful diagnostic of the gal-bladder disease. The experiment was conducted on a total of 78 subjects. These results have either performed over a small data set, and the advanced machine learning techniques have not been explored to prove the hypothesis. However, with the advancement of computer vision and machine learning techniques, no such negative study has not been reported.

2.4 Iris and Sclera Localization and Segmentation

Various studies related to the iris and sclera segmentation system have been discussed in this section. Localization and segmentation is a crucial step in diagnosis as well as biometric applications. A detailed overview of the state of art methods has been consolidated and discussed in detail.

2.4.1 Iris Segmentation

The iris is known to be the most popular and reliable tool for biometric. Recent studies have shown its importance in the disease diagnosis [65]. Iris segmentation is the process of segmenting out the iris part of the image of an eye for further operations. The segmentation process deals with eyelids and external factors like light illumination and reflections. Ignorance of these factors may lead to segmentation error and finally results in incorrect recognition or diagnosis. In this section, a detailed discussion over the state of the art methods of iris segmentation has been performed. This study will help to compare different iris segmentation methods based on their performance, effectiveness, and techniques. The

first literature over iris localization was by Daugman in 1992-93 [66], since then it has come a long way. Based on the technique used, advanced segmentation techniques have been divided into three categories methods based on classical circular approaches, methods based on advanced non-circular approaches, and methods based on the active counter algorithm.

In the methods based on classical circular approaches, the most significant work in biometric was done by Daugman. His publication [66] and a patent [67] describe the iris segmentation in detail using Integro Differential Operator (IDO). Daugman is known as the father of iris biometric, and iris recognition had been developed as an established biometric method after his contribution. These publications of Daugman became the standard reference model for iris biometrics. Later on, in 2004 [59], Daugman further stated that eye images should be captured under Near Infra-Red (NIR) illumination, as NIR illumination helps to explore the detailed information of pigmented iris also the visible light tends to observe the melanin pigment. Proenca *et al.* [68] achieved better performance for the noisy database by decreasing the False Rejection Rate (FRR) of the IDO algorithm. Radman et al. proposed a reliable and fast algorithm, which overcome the problem of sizeable computational time in localizing the iris parameters in IDO. In the first phase, pupil center is roughly estimated by circular Gabor filter, then by using IDO, actual parameters of iris and pupil are localized in small areas. Finally, the upper and lower eyelids portions were extracted by the line wire technique. Although, the proposed technique given the faster result, but authors have analyzed it only over one data set only.

Wildes [69] had developed a different system then IDO approach at Sarnoff Labs for iris biometrics. Wildes designed a digital camera with the help of a diffuse source and polarization to capture images in the low light and to localize the iris foundries. Authors also proposed an iris localization method based on Circular Hough Transform (CHT), and later on CHT based iris segmentation has become very popular among researchers. Wildes *et al.* got two patents [70, 71] to described their automated CHT based iris segmentation method. The main disadvantage of the system was that the subject d to self position his/her eye according to the camera position. Later on, in another research article system was updated in terms of hardware setup and iris evaluation scheme [72]. Later on, other CHT based iris segmentation frameworks also presented [73, 74]. Li Ma *et al.* [73] approximated the iris region by projecting the image in the vertical and horizontal direction, then iris parameters have been

estimated using edge detection and CHT technique. Abdul Jalil et al. [75] utilized CHT for color image segmentation, CHT has been applied to all three planes separately, and then iris boundaries have been estimated. Lin Wan *et al.* [76] proposed a circle based method for non-ideally captured iris using anisotropic diffusion. To improve the computational complexity, Laplace Pyramid (LP) framework has been incorporated for localizing the inner and outer boundaries of the iris. Although the above-mentioned algorithms give a reliable and fast technique for iris segmentation, these algorithms have been adopted assumption that iris is circular. Later on, many algorithms have been employed the non-circular techniques. In the methods based on advanced non-circular approaches, He *at al.* [12] presented an non-circular, fast, and reliable iris segmentation framework. Ada-boost cascade iris detector has been used to estimate the rough position of the iris center. Afterward, by using an elastic pulling and pushing model, the iris boundaries have been extracted. Zuo *et al.* [77] demonstrated an ellipse fitting and thresholding based robust iris segmentation method. Jan *et al.* [13] proposed a set of algorithms to overcome the problem of poor recognition due to uneven light illumination, reflection, and non-ideal iris invisible and well as infra-red light. Authors have utilized a two-fold technique based on Circu-Differential Accumulator (CDA) and gray statistics for iris localization. Abate *et al.* [78] presented a new iris segmentation technique, which was based on watershed transformation and specially designed for mobile devices. Watershed Transformation (WT) has been applied in two steps, in the first step WT used to binarize the image and in the second step WT used to analyze the sub-regions of the different zones of the iris. Further, WT has been utilized by Frucci *et al.* [79] for more accurate recognition.

Methods based on an active counter algorithm in the iris segmentation has been emerged in the last decade only but already utilized by many researchers for non-ideal iris segmentation. Daugman [80] suggested a consistent method for modeling the inner and outer boundaries of iris using active counters. Vasta *et al.* [81] proposed a curve evolution approach using a modified, modified Mumford–Shah method. Banerjee *et al.* [14] proposed an intensity-based unsupervised fully automatic algorithm the center point of the pupil has been detected using intensity-based profiling and then with the help of active geodesic counters inner and outer boundaries have been estimated. Talebi *et al.* [82] proposed an active counter-based two-step algorithm for iris segmentation. Noise has been eliminated by using the median filter then the modified active counter has been used for iris segmentation. Roy *et al.* [83] proposed a robust

iris segmentation algorithm for images captured in the random environment. Un-controllable factors like eyelids, eyelashes, uneven intensity, non-uniform light reflection, etc. have been considered. To evaluate the intensity information of the local regions and to accurate estimation of the iris boundaries, a modified crevasse curve evaluation technique has been utilized. Authors have been defined as a data fitting energy function and two filling functions, in terms of counters and to approximate the intensities of the image, respectively. Susitha *et al.* [84] utilized the Super Pixel Segmentation (SPS) algorithm for iris segmentation, the advantage of using SPS that it does not work on the estimation of iris boundaries like IDO and CHT.

2.4.2 Sclera Segmentation

The sclera is the white part of the eye, which contains unique blood vessel patterns. The sclera is not only used for biometric application but also many diseases can be predicted with the help of sclera. Therefore, accurate segmentation of the sclera is an essential task.

Derakhshani *et al.* probably reported the first work for sclera recognition [85, 86]. Authors have demonstrated a manual sclera segmentation method, which later on improved by semiautomatic and automatic techniques by Crihalmeanu *et al.* [87] and Zhou *et al.* [88] respectively. A semiautomatic method was proposed on non-ideal images by Crihalmeanu *et al.* [87] was based on K-means clustering and then manual correction of eye laces. Kangrok Oh *et al.* [89] performed ostu's thresholding method to detect sclera and to segment out the sclera region low pass filtering followed by histogram equalization has been used in the HSV space. Khosravi *et al.* [90] used active counter models based on Time Adaptive Self Organizing Map (TASOM) to localize the internal boundaries of the sclera region. Thereafter, by utilizing neighborhood characteristics, the boundaries have been estimated. Das *et al.* [91] proposed a bank of 2D decomposition of the haar wavelet that enhances the tenuous vessels of the sclera. Zhou *et al.* [92] proposed a segmentation method based on HSV color space and by using up and down-sampling. Zhou *et al.* [92] used a Sobel operator to emphasize the region of interest and then based on the hypothesis that there is a huge intensity difference between both sclera area and iris, ostu's thresholding technique has been applied for the segmentation. To estimate the iris boundaries, authors have applied the Integro Differential Operator (IDF). A new robust method of sclera segmentation for color images was proposed in [85].

The sclera is also used for the disease diagnosis. A. Laddi *et al.* [93] utilized the sclera region of the eye for the detection of jaundice. Images of both the eyes of 330 patients suffering from jaundice and 90 healthy peoples have been acquired. Color information of that region in L, a, and b values as per the CIE Lab color model was extracted. The image processing algorithms include steps like image segmentation, morphology, and color analysis.Neonatal jaundice, the yellow coloration of the skin and sclera caused by hyper-bilirubinemia, is the most common condition confronting neonatologists. About 60% of term and 80% of preterm infants diagnosed by jaundice in the first week of life [41]. T. S. Leung, 1, K. Kapoor *et al.* [42] presented a novel diagnosis method for neonatal jaundice; authors have proposed exploiting the yellow discoloration in the sclera processing the color values of sclera pixels to predict the Total Serum Bilirubin (TSB) levels. Three-color indexes for red, green, and blue, represented by R, G, and B were calculated by averaging 900 neighboring pixels (30 × 30 regions) centered on the predefined pixel coordinates. In their early research, a mobile application was developed to monitor jaundice in newborn babies [43]. Bili-Cam, a low-cost system that uses smart-phone cameras to assess newborn jaundice, was developed and evaluated on 100 newborns. Images of sclera vessel patterns seem to be defocused or blurry. Most importantly, the vessel structure in the sclera is multi-layered, and it contains a complex nonlinear deformation. Eliza Gail Maxwell [44] presented a comparison between Contrast Limited Adaptive Histogram Equalization (CLAHE) and Gabor filter for sclera blood vessel enhancement.

In this chapter, literature related to diabetes and its diagnosis using iris has been surveyed. The survey is being performed in four stages: In the first stage the literature for diabetes and its global status have been studied, the second stage is about the role of physiological parameters in diabetes, in the third stage literature on iris-based diabetes diagnosis have been reviewed. In this section, different features of specific regions of the iris and classifiers employed by the researchers have been reviewed. Finally, in the fourth section, pre-image processing techniques for iris and sclera segmentation have been studied. Moreover, iris segmentation algorithms have been divided into three types based on the algorithms used for segmentation.

Chapter 3

Diagnosis of Diabetes Using Iris Images

Complementary and alternative medicine techniques have shown their potential for the treatment and diagnosis of chronic diseases like diabetes, arthritis, *etc*. At the same time, digital image processing techniques for disease diagnosis is a reliable and fastest-growing field in biomedical. The proposed model in this chapter evaluates the diagnostic validity of iridology for the diagnosis of type 2 diabetes using soft computing methods. The investigation was performed over a close group of a total of 338 subjects (180 diabetic and 158 non-diabetics). Infrared images of both the eyes were captured simultaneously. The region of interest from iris was cropped as zone corresponds to the position of pancreas organ according to the iridology chart. Statistical, texture, and discrete wavelength transformation features were extracted from ROI. The results show the best classification accuracy of 89.63% calculated from the RF classifier. The highest specificity and sensitivity were absorbed as 0.9687 and 0.988, respectively. Results have exposed the effectiveness and diagnostic significance of the proposed model for a non-invasive and automatic diabetes diagnosis.

3.1 Introduction

The prevalence of diabetes mellitus increased very fast in the past few years, and it has now become a global health problem. International Diabetes Federation (IDF) estimated that worldwide 381.8 million people are affected by diabetes, and about 591.9 million people will be affected by this disease with a massive hike of 55% by 2030 [64]. Diabetes mellitus is simply excess of blood sugar level. Pancreas organ, a fish-shaped gland with its function of insulin secretion often associated with diabetes mellitus. Pancreas organ performs both the functions, exocrine, and endocrine of glands. It produces and releases the digesting juices to the intestine as well as controls the blood sugar level by producing glucagon and insulin [94].

Type 2 is a prevalent form of diabetes in which the body becomes resistant to insulin, and more insulin needs to bring glucose level down. Therefore, the pancreas organ needs to produce more insulin. In the later stages, the pancreas organ does not produce enough insulin

to consume glucose. Pancreas organ starts dysfunctioning before its actual physical symptoms are observed. Therefore, early diagnosis of diabetes is a big challenge, and late diagnosis may lead to severe and advert effects on health. CAM therapies are quite famous for chronic diseases like diabetes, arthritis, *etc*. and became very popular in the past few decades [8]. Many mainstream medical practitioners believe these therapies are unorthodox, but in many countries, CAM is used widely for healthcare applications. Harris *et al.*[8] found that around 52% and 38% general population are using CAM in Australia and the USA, respectively. The authors also found a significant increasing trend in the population using CAM.

A number of diabetic research had performed a large population set to improve the increased risk factor on other diseases due to diabetes [95–98]. Iridology is an alternative medicine science that correlates iris patterns, colors, tissue weakness, breakage, and other characteristics, which acquires evidence about a patient's systemic health. It reveals weakness or breakage in tissues long before the symptoms appear. Iridology practitioners match their interpretations of the iris chart in which iris is divided into several zones corresponding to specific body organs [49]. A typical iris chart divides the iris approximately 80–90 zones [99]. Salles *et al.* [50], in their study, found that literature available about iridology is quite limited, and it's necessary to have some more severe and extensive studies in this context. Jae Young *et al.* [51] explained that it helps as a supplementary source to diagnosis the medical condition. The authors investigated the relationship between iridological constitutions and ACE polymorphism in hypertensive. Advanced image processing and data mining techniques have been utilized as a powerful disease diagnostic tool in biomedical applications [47, 100]. Ramlee *et al.* [58, 61, 62] used iris recognition algorithms to detect cholesterol present in the blood vessels. Salles *et al.* [101, 102] studied the prevalence of iridology signs for diabetes mellitus and verified these signs for disease risk factors. Chaskar and Sutaone [103] diagnosed diabetes from iris using PCA and neural networks as a classifier.

This chapter deals with new directions and methodologies to assess the potential of iridology for diagnosis of diabetes, along with modern computer imaging and machine learning techniques.

3. 2 Material and Methodology

An attempt is made to evaluate the feasibility of diabetes diagnosis with the help of iris and machine learning techniques. Basic idea is to establish a standardized procedure for diabetes diagnosis along with conventional clinical diagnostic methods. Figure 3.1 shows basic steps followed for the development of the iris based diagnostic model. These steps are further briefly discussed as follows:

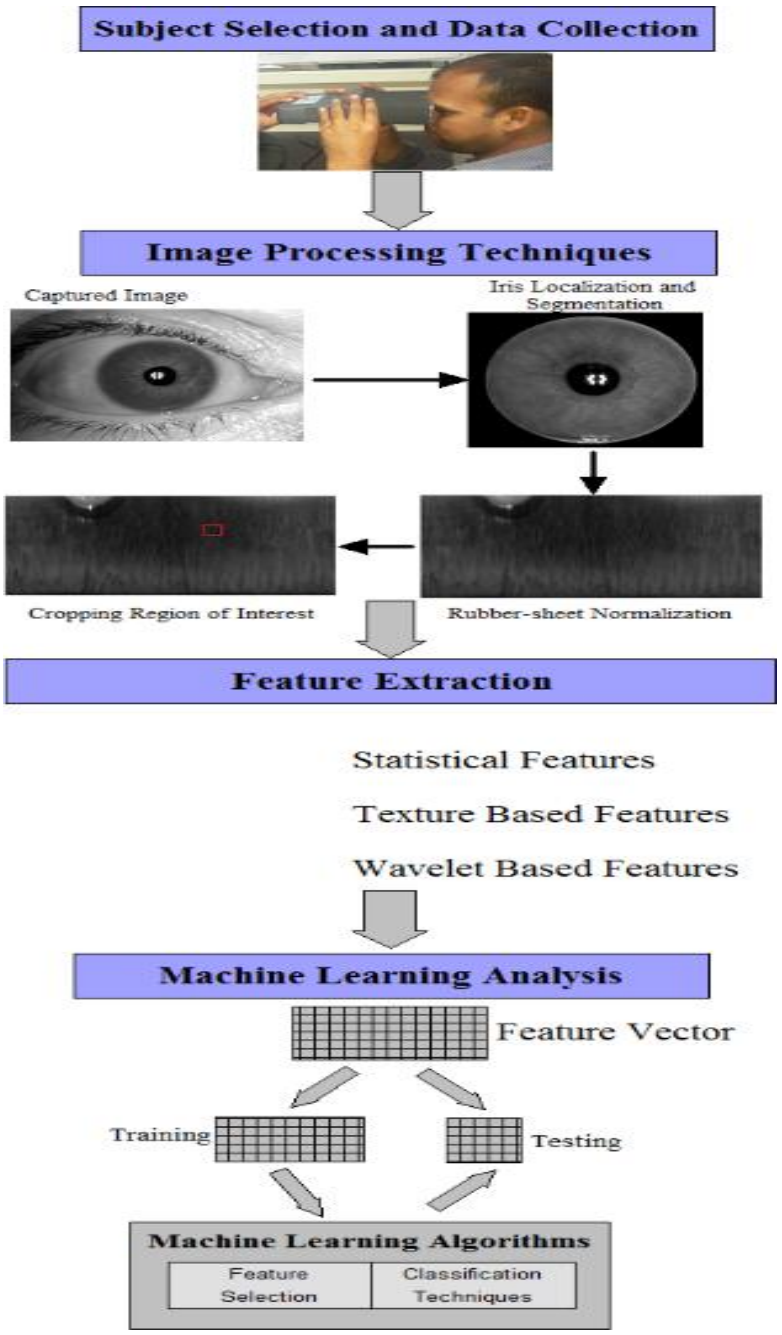


Figure 3.1 Summarized methodology for diabetes diagnosis using iris

3.2.1 Subjects Selection and Data Collection

A total of 338 (180 diabetic and 158 non-diabetic) subjects were investigated in the present study. Only type 2 diabetic patients were considered for the investigation. The protocol followed for the data acquisition was approved by the Institute Research Board, Thapar Institute of Engineering and Technology, Patiala. Subject selection was based on three factors, namely, gender ratio, average age, standard deviation, and duration of the diabetic state of the age (Table 3.1). The duration of the diabetic state varied from 1 year to 25 years. The mean duration of a diabetic state was 7 years. Diabetic subjects are further subdivided into five different groups according to the duration of diabetes state, as shown in Table 3.1. Figure 3.2 shows the distribution of diabetic subjects according to the duration of diabetes state.

Table 3.1 Detailed subject distribution

	No. of Male	**No. of Female**	**Gender ratio**	**Average age**	**Standard Deviation**	**Total**
Diabetic subjects	102	78	1.307	53.32	8.56	180
Healthy subjects	91	67	1.35	52.86	9.98	158

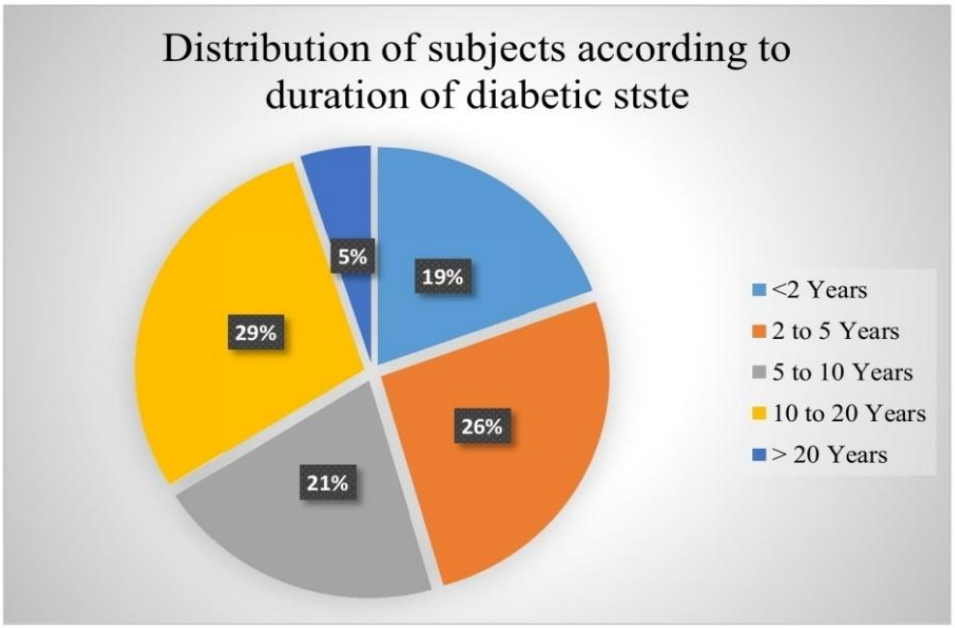


Figure3.2 Percentage distribution of diabetic subject as duration of diabetic state

Over-night fasting and 2-hour postprandial blood glucose test was performed on the subjects. Iris images were acquired using I-SCAN-2 (make cross match-technologies, USA). Iris scanner acquires gray infra-red (IR) images of size 640×480 of left and right iris simultaneously. IR imaging is a non-contact and non-ionizing imaging modality. It's best suitable for the assessment of soft tissues and broken tissues, due to their different absorption and scattering at the IR wavelength.

3.2.2 Iris Image Processing

Image processing is used to extract suitable features form the region of interest from the image of the eye so that the chosen features can be extracted. Sub-steps for iris image processing in the current research work is iris localization and segmentation, rubber-sheet normalization, and cropping region of interest.

Segmentation is a process of locating the inner and outer boundaries of the iris in the image of the eye. Daugman [59, 66, 104] introduced IDO for iris localization. The operator follows a circular path to detect the iris boundaries. Sankowski *et al.* [105] presented a novel algorithm for iris segmentation from eye images taken in visible and near-infrared lights and proposed a unique reflection localization algorithm with high reliability. Wildes [72] suggested an automatic segmentation algorithm to calculate the parameters of iris, which was based on CHT [106]. Li *et al.* [107] proposed a new region-based distinction formula that was realized by a simple finite difference for re-initialization. Radman *et al.* [108] proposed a fast and robust segmentation algorithm that find out iris parameters precisely. In this chapter, a modified CHT algorithm is used to calculate the radius and center coordinates of the inner and outer boundaries of iris. Initially, the first derivative of the intensity of the image is calculated, and then thresholding is applied to generate the edge map. Thereafter, the parameters of the circle are estimated by the circles passing through the maximum edge points in the edge map.

To map the extracted iris into a fixed rectangular representation rubber-sheet normalization [59, 66, 109] algorithm is performed. The rectangular output is invariant to the dimensional irregularities, position, orientation, camera angle of the iris, and pupil within the image of the eye. This algorithm also normalizes other factors like the optical angle of the camera, illumination level, and angular inconsistencies. Each point within the iris region in the

Cartesian coordinate (x, y) of the eye image I(x, y) remapped into homogeneous polar coordinate (r, θ) as:

$$I\{x(r,\theta), y(r,\theta)\} \rightarrow I(r,\theta) \tag{3.1}$$

where, r lies into an unit interval [0, 1] and angle θ varies from [0, *2n*] with

$$x(r,\theta) = (1-r)x_p \ + rx_l \ (\theta) \tag{3.2}$$

$$y(r,\theta) = (1-r)y_p \ + ry_l \ (\theta) \tag{3.3}$$

(x_p, y_p) and (x_l, y_l) are parameters of pupil and iris respectively in θ direction. A homogeneous 2D array of size 360×720 was obtained using rubber-sheet normalization algorithm in the present research work.

ROI is cropped from the rubber sheet normalized iris and is considered according to the position of head, tail, and body of the pancreas organ in the iris, as suggested in the modern iridology chart [99]. According to the iridology chart, the head, tail, and body of the pancreas are represented by the three different segments. The head of the pancreas is represented in the right eye between 7 to 8 o'clock, while the body and tail are represented in the left eye between 7 to 8 and 4 to 5 o'clock, respectively

3.2.3 Feature Extraction and Selection Methods

Depending on the diabetic health condition of the individual, changes are observed in statistical and textural features from the region of interest. A total of 180 features are extracted to quantify the broken tissue information of the iris. These features are divided into three groups first-order statistics, textural features, and wavelet features.

To describe the distribution of intensities within the ROI, first-order statistics features have been calculated. First-order statistics features contribute to the information associated with the gray level distribution of the ROI. Table 3.2 shows the extracted first-order statistical features and their formulas [110].

Texture features are also included along with statistical features to provide information associated with the position of the numerous grey levels of an image. Features are calculated

from the respected Grey Level Co-occurrence Matrix (GLCM). Table 3.3 shows the extracted texture features and their respected formulas [111, 112].

Table 3.2 Extracted statistical features

Statistical Features	**Formulas**
Mean Intensity	$\frac{1}{N}\sum_{i=1}^{N} x(i)$
Standard Deviation	$\left(\frac{1}{(N-1)}\sum_{i=1}^{N}(x(i)-\bar{x})^2\right)^{1/2}$
Entropy	$\sum_{i=1}^{K} P(i).log_2 P(i)$
Histogram Intensity features	Five histogram intensity levels

where, x_i denotes the two dimensional image of N scalar observation and P is first order histogram having N_l intensity levels.

Table 3.3 Extracted texture features

Textural Features	**Formulas**
Contrast	$\sum_{i,j} \lvert i-j\rvert^2 P(i,j)$
Correlation	$\sum_{i,j} (i-\mu_i)(j-\mu_j)P(i,j)$
Energy	$\sum_{i,j} (P(i,j)^2$
Homogeneity	$\sum_{i,j} \frac{P(i,j)}{1+\lvert i-j\rvert}$

where, P(i,j) is co-occurrence matrix having N discrete levels, μ is mean of P(i,j), marginal row probabilities $P_x(i)$= is defined as $\sum_{j=1}^{N} P(i,j)$ and marginal row probabilities $P_y(i)$= is defined as $\sum_{i=1}^{N} P(i,j)$, μ_x is mean of P_x, μ_y is mean of P_y, σ_x and σ_y are standard deviation of P_x and P_y respectively.

The wavelet transformation shows great potential in various fields like matching, biomedical application, telecommunication, *etc*. Discrete Wavelet Transforms (DWT) is suitable for non-stationary signal and image analysis for its utilization ability of time-space and capacity of handling interference terms [112, 113]. By decomposing region of interest into time and frequency components, the textural information can be analyzed more effectively. 2- D DWT is calculated, in which ROI X(i,j) is decomposed into 4 "harr" mother wavelet decompositions. These decompositions are low and high-frequency functions and labeled as XLL, XHL, XLH, XHH, where L and H are low and high-frequency functions, respectively, e.g., XLH is calculated as high pass component in the Y-direction of result from low pass filtering in the X-direction. Thereafter, for each decomposition and possible combinations of these decompositions, first-order statistics and texture features are computed.

Feature selection is important for complexity reduction and for fast computing in machine learning methods. Feature selection techniques also remove feature redundancy and improve classification reliability. Fisher score discriminator, t-test, and chi-square test are three popular filter-based feature selection methods which, have been used in the present research work. Feature selection methods have been selected based on computational efficiency, simplicity, and popularity in literature for biomedical applications. A scoring criterion has been calculated for each feature. The ranking of features is done based on the scoring criterion. Scoring criterion or relevance index (J) for the different techniques is calculated [100] and shown in Table 3.4.

3.2.4 Classification

In advance computing methods, classification techniques are used for concluding a decision from the categorized data set. The classification algorithm concludes a hypothesis by analyzing the training data set that helps in predicting the classes from the testing data set. A number of classifiers which belongs to various fields of computer application and data mining have already been proposed [17]. In the present research work, six different classifiers,

named as Binary Tree Model (BT), Support Vector Machine (SVM), Adaptive Boosting Model (AB), Generalized Linear Models (GL), Neural Network Model (NN) and Random Forest (RF) have been used. Classifiers are implemented in the R package [114]. Classifiers have been trained using the repeated 10 fold cross-validation technique.

Table 3.4 Feature selection methods and specification

Feature Selection Method	**Formula for Relevance Index or Scoring Criteria**	**Specification**
Fisher Score	$J_{fisher}(X_k) = \frac{\sum_{m=1}^{2} n_m (\mu_{k,m} - \mu_k)^2}{\sum_{m=1}^{2} n_m \sigma_{k,m}^2}$	X_k = feature to be evaluated, μ_k =overall mean of feature to be evaluated, m = no of samples in m^{th} class, $\mu_{k,m}$= mean of (X_k) on m^{th} class and $\sigma_{k,m}^2$ = variance of (X_k) on m[th] class
t-test	$J_{chi-square}(X_k) = \sum_{i=1}^{r} \sum_{m=1}^{2} \frac{(n_{im} - \mu_{im})^2}{\mu_{im}}$ where, $\mu_{im} = \frac{n_{*m} n_{i*}}{N}$	μ_1 and μ_2 = means of two classes σ_1 and σ_2 =variance of two classes
Chi-square test	$J_{chi-square}(X_k) = \sum_{i=1}^{r} \sum_{m=1}^{2} \frac{(n_{im} - \mu_{im})^2}{\mu_{im}}$ where, $\mu_{im} = \frac{n_{*m} n_{i*}}{N}$	n_{im} =no of samples with i[th] feature value in m[th] class, n_{i*} =no of samples with i[th] feature value n_{*m} =no of samples in class m and N no of samples

3.3 Analysis

The prime motive of using different feature selection and classification methods is to investigate and compute the best suitable feature selection and classification algorithm according to the present data set. Depending upon the scoring criterion of each feature selection method, 10 to 50 top-ranked features are selected, with an increment of 10 features

(i.e., 10, 20, 30, 40, 50 features). Classifiers are trained using these features and applied to the testing data to predict the classes. Thereafter, accuracy, sensitivity, and specificity are calculated to evaluate the performance of the classifiers. To generalize the classifier performance, 10 fold cross-validation technique has been applied. Sensitivity or True Positive Rate (TPR) defined as the ratio of correctly identified diabetic patients, which should be maximum. False Positive Rate (FPR) is another characteristic of the classifier, which is defined as the ratio of diabetic patients, incorrectly recognized as non-diabetic. The percentage of FPR should be minimum for better classification. FPR is also written as (1-specificity). Sensitivity and specificity both are very important parameters for performance evaluation of any classifier. Specificity is even more critical in the present case because identifying a diabetic person as non-diabetic leads to a big mistake.

3.4 Results and Discussion

Classification accuracies are calculated for all six classifiers. Table 3.5 shows the accuracies in % of a different classifier against selected features. Best classification accuracies are calculated using the t-test feature selection method as 89.63%, 89.38% for 40 features, and 89.97%, 89.18% for 50 features by RF and AB classifiers, respectively. Sensitivity and specificity both are calculated for the 40 and 50 top-ranked features for AB, RF, and SVM classifier as these combinations gave the best classification accuracy amongst all calculated accuracies.

Table 3.6 shows the specificity and sensitivity of classifier performance against selected features. Maximum sensitivity and specificity have been found as 0.8687 and 0.8980 for the t-test feature selection method for 40 features using RF classifier. However, maximum accuracy was calculated as 89.97% using the t-test feature selection method for 50 features, but for this accuracy, specificity is calculated as 0.7942. Results obtained from the presented research shown that the statistical, textural, and DWT features have a great ability to classify diabetic and non-diabetic people through iris images. The t-test feature selection, method, and RF classifier are found best suitable.

Table 3.5 Classification accuracies of different classifiers against feature selection methods.

Classifier Accuracies						Feature Selection Method	No. of Selected Features
BT	RF	AB	SVM	GL	NN		
64.73	73.28	72.68	71.98	50.98	51.4	Fisher score	10 features
65.18	78.97	78.35	76.32	52.98	55.94	t-test	
62.32	70.57	70.83	72.18	51.28	53.28	Chi-square test	
68.98	78.94	76.39	75.93	53.73	57.28	Fisher score	20 features
70.73	82.32	81.73	80.73	54.23	58.38	t-test	
68.7	73.28	75.03	74.98	53.98	55.38	Chi-square test	
72.37	80.54	81.36	82.52	55.37	62.95	Fisher score	30 features
73.18	87.58	87.11	85.38	58.39	62.73	t-test	
70.95	79.97	81.18	82.14	57.28	59.53	Chi-square test	
73.18	82.9	82.73	86.73	61.15	65.09	Fisher score	40 features
74.14	89.63	89.38	87.93	61.88	66.94	t-test	
71.08	87.37	88.73	84.78	60.93	64.87	Chi-square test	
74.63	83.54	83.83	87.28	61.97	66.84	Fisher score	50 features
75.2	89.97	89.18	88.53	62.35	67.29	t-test	
71.73	88.59	89.1	85.48	60.32	65.37	Chi-square test	

Table 3.6 Specificity and sensitivity of different classifiers against best accuracies

Classifier						Feature Selection Method	No. of Selected Features
RF		AB		SVM			
Specificity	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity		
0.7944	0.9182	0.8489	0.8987	0.8731	0.9171	Fisher score	40 features
0.9687	0.988	0.8423	0.9903	0.8128	0.9423	t-test	
0.8213	0.9237	0.9037	0.9448	0.9111	0.9328	Chi-square test	
0.8498	0.9842	0.9142	0.9537	0.8213	0.9321	Fisher score	50 features
0.8942	0.9973	0.827	0.9978	0.8827	0.9423	t-test	
0.9178	0.9728	0.9138	0.9747	0.8384	0.9779	Chi-square test	

Further analysis has been performed between non-diabetic subjects and diabetic subject of different groups (Table 3.1) to check the performance of proposed algorithm for subjects having different duration of diabetic state. To minimize over-fitting, same number of samples have been selected randomly from non-diabetic group as diabetic sub-group. Figure 3.3 shows the average classification accuracy and standard deviation of 10 fold cross validation

technique. The whole data set is divided into 10 random subsets. Then kept one subset as a holdout set or test set and the remaining nine subsets for the training set. Train the model on the training set and evaluate it on the test set. The same process is repeated for every subset as the testing set and evaluate every model. The final evaluation score calculated as the average of all the evaluation scores.

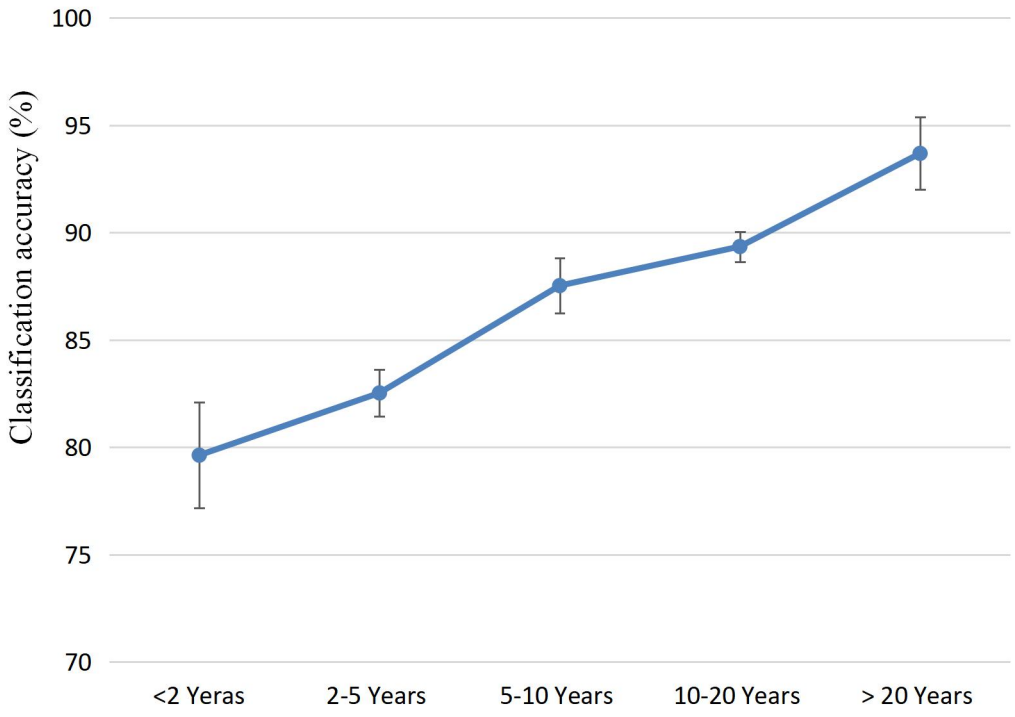


Figure 3.3 Classifier performances according to duration of diabetic state

Medical imaging is a common and reliable tool for information gathering in clinical practice. In the present research, an attempt has been made to develop an automated technique for screening of diabetes using iris. For the practical realization of the iris feature-based diagnostic system, it is required to evaluate and compare different feature selection and classification algorithms. Three filter-based feature selection techniques are employed as filter-based techniques are more efficient and avoid over-fitting. Furthermore, filter-based feature selection techniques are classifier independent; hence it allows separation of feature selection and classification components in prediction modeling. It increases the generalization of the components and overall analysis. Also, for classification analysis, six

different classification techniques are compared. Technically and theoretically, these classifiers belong to the different fields of applications and statistics. Therefore, it is not very easy to realize all classifier parameters in an unbiased way.

Classifier performance shows that the t-test based feature selection technique yields the highest classification accuracy almost for all classifiers. The t-test feature selection technique is a simple univariate method, which is based on the feature ranking. It also handles the redundancy of features, and hence it helps in reducing the computational complexity of the classifier. Almost all the feature selection methods RF classifier gives the highest accuracy. It could be a point of argument that the performance of the classifier/s can be further improved by the different combinations of parameter configuration. However, an exhaustive investigation for parameter tuning could improve the performance of the classifier, but it needs to deal with high computational complexity and exhaustive search. Nevertheless, it is noted that the RF classifier has performed satisfactorily for other applications in the biomedical domain [115].

The results indicate that choosing the t-test feature selection method with RF classifier increases the overall prediction performance for diabetes diagnosis using computer methods. In addition, diabetic subjects are further sub-divided into the duration of the diabetic state. Figure 3.3 clearly shows that the proposed algorithm has good accuracy in classifying the subjects having a duration of diabetic state for less than 2 years. The performance of the classifier increases when subjects having larger diabetic duration added to the analysis. Therefore, it can be observed that the proposed algorithm has good ability to classify diabetes in an early stage as well as for a long duration. This sub-grouping of diabetic subjects can also be performed according to diabetic phenotype and along with MPE. The proposed algorithm can identify the overall risk for the specific diabetic subtype. Table 3.7 shows the comparison of the present study with other proposed models. Our proposed model gives a better classification with a relatively bigger sample size.

Table 3.7 Comparison of proposed model with existing models

Author	Disease	No. of Subjects	Classification accuracy (%)
Wibawa *et al.* [116]	Condition of pancreas organ	34	94
Purnama *et al.* [64]	Pancreas Disorder	50	83
Bansal *et al.* [63]	Diabetes	80	87.5
Sivasankar *et al.* [117]	Pulmonary Disease	32	84.38
Hareva *et al.* [118, 119]	Healthcare Monitoring	32	90-95
Proposed model	Diabetes	338	89.63

The proposed methodology in this chapter provides an automated tool with machine learning techniques to observe the correlation between change in the features of the specific areas of iris with diabetes mellitus. Also, the method can be established as an alternative diagnostic tool for the quick scanning of diabetes. The technique used for personal identification checks the epigenetic patterns. At the same time, this chapter explores the fine and minute tissue description in the specific areas of the iris texture through the statistical parameters of the wavelet decomposition. Despite these promising results, the proposed model was performed on a systemic disease with ocular indicators. This research's main contribution is not only limited to mapping the changes in the features of the iris and diabetic health condition of the individual but also contributes to both iris diagnosis and recognition by separating the required iris features into the global structure and locally statistical features.

Chapter 4

Diabetes diagnosis using iris-based features and physiological parameters

Digital image processing and advanced machine vision techniques are popular for the diagnosis of disease(s) in biomedical science. This chapter presents a detailed comparative analysis of machine learning based classification techniques to diagnose type 2 diabetes using the combination of iris based features and physiological parameters. A set of 334 subjects is investigated. Subjects are divided into diabetic and non-diabetic groups. Moreover, the diabetic group is classified into three different subgroups according to the duration of the diabetic state. Statistical features, Gray Level Co-occurrence Matrix (GLCM), and Gray Level Run Length Matrix (GLRLM) based features are extracted from the specific areas of iris. Nine classifiers of different application areas are selected and subsequently, six parameters (accuracy, precision, sensitivity, specificity, F-score and area under the curve) of each classifier are computed and analyzed. The analysis provides promising results with more than 95% of accuracy. The proposed technique can be used as a non-invasive and non-contact diabetes diagnosis tool.

4.1 Introduction

A Computer Aided Diagnosis (CAD) for the health-care applications helps in decision-making for the medical practitioners. In recent years, CAD techniques along with the modern Artificial Intelligence (AI) methods have shown great achievements for disease(s) diagnosis, and provide a more robust, indispensable, and reliable system [120]. Medical imaging diagnosis models based on special feature extraction and supervised classification algorithms are famous due to their high reliability and accuracy. Therefore, the field of medical image analysis is attracting more researchers and medical practitioners [121].

Numerous researchers have explored the unorthodox imaging modalities using CAD techniques. For instance, Kanawong *et al.* [122] developed a mobile-based application for analyzing the images of tongue. Pang *et al.* [123] diagnosed appendicitis by analyzing the color of tongue and external characteristics of its blood vessels. Goyal *et al.* [124] developed

a health diagnosis system by analyzing the wrist pulse. Moreover, researchers have also shown potential work in eye based medical diagnosis system [125–127].

Iridology is a diagnosis approach that observes changes in the patterns, textures, colors and structure of the local areas of iris [128]. Iridology believes that minute lurking details of the specific areas of iris reflect the health conditions of an individual. Iridological signs with features have also been explored to diagnose the urological, cardiovascular [129], kidney abnormalities [130], and digestive problems in patients.

Iris is a colored circular part of the eye, which controls the volume of the light that enters into the eye. Iris is made up of tissues and surrounds the pupil. Normally, iris appears lighter than pupil. Iris comprise of rich textural characteristics that make iris diagnosis a useful tool in pathology. The outer surface of the iris is divided into two zones namely outer ciliary zone and central pupillary zone. The border between these two zones is called collarette. Iris is divided into two layers named as posterior leaf and anterior leaf. The posterior leaf consists of the dilator muscle, posterior pigmented epithelium, and sphincter muscle. The anterior leaf is more superficial then posterior leaf, it consists of connective tissue stroma with cells, blood vessels, and nerves, which supply the sphincter and dilator; however, there is no epithelial layer in primate species [57]. Main purpose of the iris is to control the light entering to the eye, by contraction and dilation process. Dilator and sphincter muscles are involved in controlling the size of the iris by regulating the amount of light that enters into the pupil. In low light, dilation of the pupil provides a way for increasing the number of photons reaching to the retina, which gives the edge to slower dark adaptive mechanisms. If the eye exposes to the bright light, the pupil constriction can reduce retinal illumination quickly [57]. The size of the iris varies due to these variations. This irregularity in the iris surface characteristics leads to a major problem in iris recognition system [4]. In fact, it is debatable that slight change occurs in the special iris parameters because of inconsistencies. Therefore, to ensure the satisfactory performance in iris recognition, the system needs to re-enroll the user on regular basis [129]. An operation, injury, and indirect infection due to any disease or change in the health problem are commonly directed towards dilator. Hence, complex non-linear deformation in the iris can be observed with the help of the sphincter muscle of the iris. The iris consist of five different layers of complex fibrous tissues and has a rich textural patterns of pigment spots, tiny holes, filaments, and strips [131]. Almost 70% of the iris part

comprises of nerve fibers [109]. These iris nerves are directly connected to the human's nervous system and brain.

Health status, functional change of organs, hormonal changes and blood circulation system reflect in the anatomical characteristics of the iris [132]. Lin Ma *et al.* [55] presented a medical analysis using geometric features of iris like pupil roundness, pupil largeness, and roundness of collarette boundaries. Young *et al.* [51] studied the relationship between angiotensin and iridological constitution. Angiotensin converts into enzyme gene polymorphism, which is the most popular biomarker to study the vascular disease. Hussein *et al.* [57] evaluated kidney disease in a controlled close group study by means of iridology. Ramlee *et al.* [58, 61, 62] utilized iris recognition methods to find out the cholesterol in blood vessels. Bansal *et al.* [63] predicted diabetes using wavelet features of specific areas of iris and Support Vector Machine (SVM). In recent years, researchers have presented various models for diagnosing diabetes using iris based features along with modern computer methods [23–29].

Physiological parameters have also attained high accuracy and reliability for the prediction of diabetes with the help of machine learning techniques [46, 47, 136]. Dwivedi [46] presented a detailed analysis of computational techniques to predict the diabetes using Pima-Indian diabetes database. Heydari *et al.* [47] performed a study on more than 2500 subjects, in which physiological features have been used to predict the type 2 diabetes. M. Sudha [137] provided a comparative study of classifiers for the diagnosis of breast cancer, fertility, and heart disease using physiological parameters. The disease diagnosis applications have also been developed using physiological parameters using machine learning techniques [32–33].

However, many researchers have proposed techniques for the diagnosis of diabetes using iris or physiological features, individually [25, 63, 135, 140, 142]. However still there is a need to explore the combination of the iris-based features and physiological features for diabetes diagnosis. Therefore, an accurate and reliable soft computing-based diabetes diagnosis technique has been proposed in this chapter, which utilizes the combination of physiological as well as iris-based features. It is also observed from the literature that an increase in the duration of diabetic state leads to increase in acute as well as chronic complications. Hence, it is of utmost importance to estimate the duration of diabetic state in the diabetic patients. Majority of the population in undeveloped and developing nation(s) do not have the basic

facility of diabetes diagnosis. In India, more than 30% diabetic people remain un-diagnosed [46]. Therefore, it is also important to estimate the prevalence of diabetes accurately so that corresponding necessary action(s) can be taken timely.

In this chapter, total 334 subjects are considered, and divided according to their diabetic health condition. IR images of both the eyes have been captured. As suggested by the iridology chart, the specific areas of the iris are cropped using Digital Image Processing (DIP) techniques. Statistical, Gray Level Co-occurrence Matrix (GLCM), and Gray Level Run Length (GLRLM) matrix based features have been extracted from cropped specific areas of the iris (both left and right). Physiological parameters of the subjects and features of specific areas of the iris are combined together in order to make a complete Feature Vector (FV) of total 113 features. Principal Component Analysis (PCA) and modified t-test feature selection techniques have been applied for selecting the features. Further, nine classifiers of different areas of applications are evaluated. The evaluation is done by calculating and comparing the six parameters to find the best suitable classifier as per the type of data set created in this work.

4.2 Methodology

The methodology is designed to standardize the procedure for diagnosing the diabetes non-intensively. Figure 4.1 shows the block diagram of proposed methodology, which includes a combination of iris-based features and physiological parameters to predict diabetes of an individual. A detailed description of the block diagram is explained in the following subsections:

4.2.1 Subject Selection and Division

Subject selection is a crucial factor in disease diagnosis. In this research work, total 334 subjects are screened. To diagnosis the diabetes, fasting and 2-hour postprandial tests have been performed. Total 84 and 250 subjects are diagnosed as non-diabetic and diabetic, respectively. Moreover, diabetic subjects are reclassified according to their duration of diabetic state as shown in Table 4.1. The distribution of age and gender related factors within diabetic and non-diabetic subjects are represented in Table 4. 2.

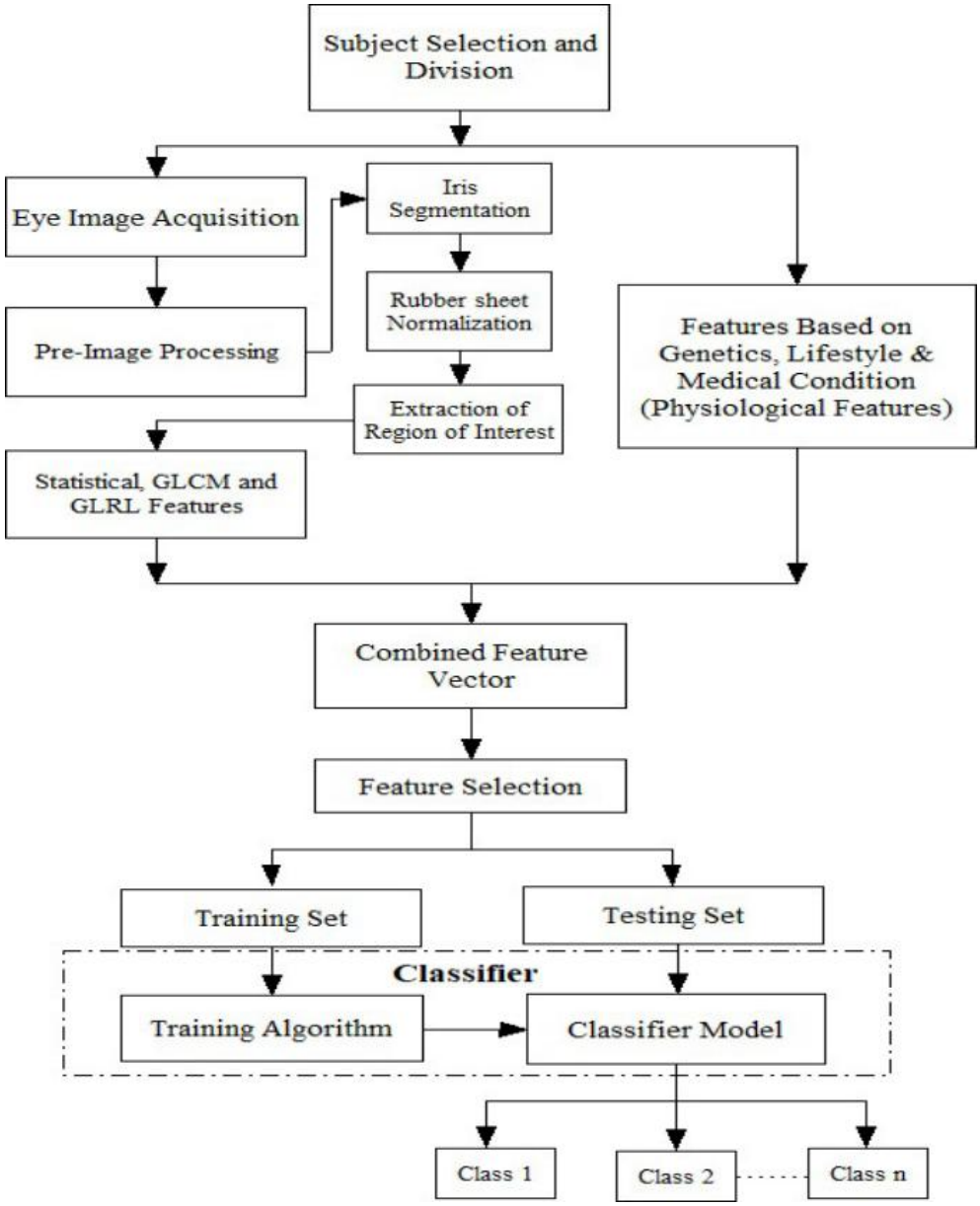


Figure 4.1 Proposed methodology

Table 4.1 Distribution of diabetic subjects according to the duration of diabetic state

Duration of Diabetic State	**Number of Subject**
< 2 Years	84
2-10 Years	84
> 10 Years	82

Table 4.2 Classification and detailed distribution of subjects according to the gender and age

Class	Diabetic Health Condition and Duration		No. of Male	No. of Female	Gender Ratio	Average Age	Standard Deviation of Age	Total Subjects
Class 1	Non-Diabetic		48	36	1.33	53.32	8.56	84
Class 2	Diabetic	< 2 Years	45	39	1.15	51.36	9.89	84
Class 3		2-10 Years	47	37	1.27	52.69	6.86	84
Class 4		> 10 Years	48	34	1.5	55.84	7.05	82

4.2.2 Physiological Features

Physiological features are related to one's genetics, lifestyle, and medical condition. Prior research works [46, 47] have proven that the physiological features play an important role in diabetes prediction. Meng *et al.* [143] presented a detailed study on the analysis of physiological features to predict the diabetes. Table 4.3 provides a list of all physiological features and their corresponding types considered in this work. It shows that 6 out of 20 physiological features are numeric (age, height, weight, BMI, systolic blood pressure, and diastolic blood pressure) while the remaining 14 features are nominal (categorical). It is necessary to transform the features from categorical to numeric so that they can be applied to the training algorithm of classifier(s) and this process is known as continuization. Binary encoding technique has been applied for continuization because there is no intrinsic ordering into the categories of the nominal features [59].

4.2.3 Eye Image Acquisition and Pre-image Processing

Images of both eyes of all the subjects are captured along with fasting and 2-hour postprandial blood glucose tests. Images are captured using iris scanner I-SCAN-2 by Cross Match Technologies, USA. The scanner captures IR images of both eyes simultaneously using IR spectroscopy technique. It is a fast-photographic technique which captures images in special resolution on the basis of tissue intrinsic composition. These special resolution IR

images are affected by the different morphological heterogeneity levels of tissues and cells. Also, IR wavelength has different absorption and scattering constants for the soft and broken tissues. Hence, IR spectroscopy is useful for the study of tissue breaking [47]. After capturing the images of eyes, pre-image processing techniques are applied to obtain the ROI from iris. Pre-image processing techniques used in this chapter are iris localization and segmentation, rubber-sheet normalization, and cropping of ROI**.**

Table 4.3 Physiological features and their corresponding types

S. No.	**Feature Name**	**Type**
1	Sex	Nominal
2	Age	Numeric
3	Height	Numeric
4	Weight	Numeric
5	BMI	Numeric
6	Family history of diabetes	Nominal
7	History of gestational disease	Nominal
8	History of high blood pressure	Nominal
9	History of used drug for high blood pressure	Nominal
10	Systolic blood pressure	Numeric
11	Diastolic blood pressure	Numeric
12	Smoking	Nominal
13	Drinking	Nominal
14	Stress at work	Nominal
15	Exercise or workout	Nominal
16	Fatigue level	Nominal
17	Condition of eye vision	Nominal
18	Thyroid in family	Nominal
19	Heart\ Lung\ Kidney\ Liver problem	Nominal
20	Taking medicine for any chronic disease	Nominal

The main parameters of iris are center and radius of inner as well as outer circular parts. Localization and segmentation processes are used for locating iris parameters within the image of the eye and segment out the localized iris, respectively. Although, a number of

segmentation algorithms have already been addressed in [59, 144, 145]. However, the most popular and acknowledged iris segmentation algorithm based on CHT [69] is used in this work.

Rubber-sheet normalization algorithm converts the circular iris into a fixed rectangular section [59]. The output of rubber-sheet normalization model is invariant to camera direction, alignment, and position. It also removes the external factors like uneven illumination and dimensional irregularities [80]. In this research work, Daugman's rubber-sheet model [59] is used to convert segmented iris into a rectangle of constant size of 360×720 (pixels).

After implementing rubber-sheet normalization, ROI was cropped according to the position of head, tail, and body of the pancreas organ as suggested in the iridology chart [146]. Different sections of the iris represent head, tail, and body of the pancreas organ. In the right eye in between 7 to 8 o'clock shows the head of pancreas organ, while in the left eye in between 4 to 5 and 7 to 8 o'clock shows the tail and body of pancreas organ, respectively. Finally, the head, tail, and body are cropped out from the normalized rectangular iris.

4.2.4 Extraction of Features From Iris

To evaluate the diabetic health condition of an individual, statistical features and textural features (GLCM and GLRLM) have been extracted from all three ROI's. These features are further discussed as follow:

4.2.4.1 Statistical features

In this chapter, 5 statistical features have been calculated directly from ROI, namely mean intensity, contrast, correlation, standard deviation, and entropy [40]. These features help to study the nature of gray pixels within the ROI and also, provide information of the gray level intensities of ROI.

4.2.4.2 Gray Level Co-Occurrence Matrix (GLCM) features

GLCM is a spatial gray level dependence matrix which characterizes, quantifies, and explores the distribution of gray level intensities. Statistical features only provide the information about gray level intensities, whereas in order to quantify the inter-pixel relationships GLCM based features have been extracted from ROI. If a gray level image $I(x, y)$ has G gray levels,

then the number of possible pairs in gray level would be $G \times G$. The frequency of each gray level pixel pair within the image has been calculated and stored in a matrix, namely, GLCM of order $G \times G$. Total 19 GLCM based features were extracted in this work, as given in Table 4.4.

For instance, energy measures the textural uniformity of an image while entropy calculates its complexity. Contrast shows the local variations of an image and is calculated as the difference between the highest and lowest values of a continuous set of pixels. Variance of an image is a measure of heterogeneity whereas correlation measures its gray level linear dependencies.

Table 4.4 Extracted GLCM features

Auto-correlation	Cluster Prominence	Cluster Shade
Difference Variance	Sum of Squares Variance	Correlation
Difference Entropy	Maximum Probability	Dissimilarity
Homogeneity	Information Measure of Correlation 1	Sum Entropy
Sum Average	Information Measure of Correlation 2	Contrast
Entropy	Sum Variance	Energy
Inverse Difference		

4.2.4.3 Gray Level Run Length Matrix (GLRLM) features

GLRLM matrix characterizes the gray level distance in different directions and represented as a number of contiguous intensities of same gray level. If the total number of pixels in the given image $I(x, y)$ are N, total number of gray levels are G, and R be the longest run of a single pixel, then GLRLM matrix will be of size (G×R). Each element $p(i, j|\theta)$ of the matrix gives the total number of occurrence of run having length j and i gray levels in a given direction θ. From this matrix, 7 features namely Short Run Emphasis (SRE), Long Run Emphasis (LRE), Gray level Non-uniformity (GLN), Run Percentage (RP), Run Length Non-uniformity (RLN), Low Gray Level Run Emphasis (LGRE), and High Gray Level Run Emphasis (HGRE) are extracted and defined in Equations 4.1–4.7.

$$SRE = \sum_{i=1}^{G}\sum_{j=1}^{R} \frac{p(i,j|\theta)}{j^2} / \sum_{i=1}^{G}\sum_{j=1}^{R} \frac{p(i,j|\theta)}{1} \tag{4.1}$$

$$LRE = \sum_{i=1}^{G}\sum_{j=1}^{R} j^2 \times p(i,j \mid \theta) / \sum_{j=1}^{R} p(i,j \mid \theta) \tag{4.2}$$

$$GLN = \sum_{i=1}^{G}\left(\sum_{j=1}^{R} p(i,j \mid \theta)\right)^2 / \sum_{i=1}^{G}\sum_{j=1}^{R} p(i,j \mid \theta) \tag{4.3}$$

$$RLN = \sum_{j=1}^{R}\left(\sum_{i=1}^{G} p(i,j \mid \theta)\right)^2 / \sum_{i=1}^{G}\sum_{j=1}^{R} p(i,j \mid \theta) \tag{4.4}$$

$$RP = \frac{1}{N}\sum_{i=1}^{G}\sum_{j=1}^{R} p(i,j \mid \theta) \tag{4.5}$$

$$LGRE = \sum_{i=1}^{G}\sum_{j=1}^{R} \frac{p(i,j \mid \theta)}{i^2} / \sum_{i=1}^{G}\sum_{j=1}^{R} \frac{p(i,j \mid \theta)}{1} \tag{4.6}$$

$$HGRE = \sum_{i=1}^{G}\sum_{j=1}^{R} i2 \times p(i,j \mid \theta) / \sum_{i=1}^{G}\sum_{j=1}^{R} p(i,j \mid \theta) \tag{4.7}$$

Therefore, total 31 (5 statistical, 19 GLCM and 7 GLRLM) features from one ROI were extracted to quantify and explore the condition of tissues in the iris for predicting the diabetic health condition of an individual.

4.2.5 Feature Selection

For one subject, total 113 features (31×3 ROI's + 20 physiological features) are extracted. Feature selection method(s) reduces computational complexity and feature redundancy, while it increases the classification reliability. Feature selection techniques are used on the basis of computational efficiency, simplicity, and popularity in biomedical applications. Two filter based feature selection methods, namely modified t-test [147] and PCA [45] have been applied and compared in this work.

Simple student t-test calculates the ratio between class mean difference and variability to find the significant difference between two classes, but it is limited only for two class problem only. For multi class problems, statistical t-score (Equation 4.8) is calculated as the difference between the mean of each class and mean of the remaining classes [147]. Equations 4.8-4.10 show the calculation of t-score using modified t-test for a multi class problem.

$$t_{in} = \frac{\bar{x}_{in} - \bar{x}_i}{M_n(S_i - S_0)} \tag{4.8}$$

$$s_i^2 = \frac{1}{N-n}\sum_{i=1}^{n}\sum_{j\in n}(x_{ij} - x_{in})^2 \tag{4.9}$$

$$M_n = \sqrt{(1/\eta_n - 1/N)} \tag{4.10}$$

where, N = number of samples; n = number of classes; η_n= number of samples in class n; $\bar{x}$ i = mean of i^{th} feature; t_{in}= t-value of i^{th} feature of nth class; $\bar{x}$ in = mean of i^{th} feature from nth class; x_{ij} = i^{th} feature of jth sample; S_i = standard deviation within class; and S_0 be the median value of $S_i\epsilon$ features. Ranking of the features have been done according to the t-score.

PCA is a well-known statistical tool that convert a set of highly correlated variables into linear uncorrelated variables using orthogonal transformation [148].

4.2.6 Classification

In supervised learning methods, classification techniques are used to conclude a decision from the labeled FV. The performance of classifier depends on the nature and type of data set. In the presented work, results of 9 different classifiers of three families are calculated and compared. These classifiers are given below:

a) Two decision tree classifiers named as Complex Tree (CT) and Median Tree (MT)
b) Four kernels of Support Vector Machine (SVM) classifiers named as Quadratic (QSVM), Cubic SVM (CSVM), Fine Gaussian (FGSVM), and Median Gaussian (MGSVM)
c) Three types of Ensemble Classifiers named as Boosted Tree (EBoT), Bagged Tree (EBaT), and Subspace K Nearest Neighborhood (SKNN)

All the classifiers belong to different application areas, and demand an exhaustive study for manual selection of parameters. As the manual selection of parameters can lead to biasing, therefore, classification learner application of MATLAB is used to perform classification. For the optimal performance of classifiers, repeated iteration technique is used for training and validated through 10-fold cross validation technique. For the validation technique, the complete data set has been divided into 10 different groups. Thereafter, 9 groups are combinedly used to train the classifier and remaining 1 group is selected for testing. This procedure is followed repeatedly for different training and testing data-sets. Thus, each group is exactly used once for testing and 9 (10-1) times for training. Finally, the evaluation score has computed as an average of all the calculated scores.

4.2.7 Statistical Analysis

The proposed model is evaluated using quantitative measures. To evaluate the performance of the classifiers, six performance measures are calculated using the following parameters of the classifiers:

a. Number of samples belong to a particular class and predicted correctly are counted as True Positive (TP)
b. Number of samples belong to a particular class and predicted absent in that class are counted as False Positive (FP)
c. Number of samples which do not belong to a particular class but predicted correctly are counted as True Negative (TN)
d. Number of samples which do not belong to a particular class but predicted present in that particular class are counted as False Negative (FN)

Thereafter, based on these parameters, following six measures are calculated for the performance evaluation of the classifiers:

- The accuracy of classifier is the degree to which prediction of a particular class matches to the true value. Accuracy is the most important parameter which shows the percentage of correctly classified subjects and is calculated using Equation 4.11.

$$Accuracy = \frac{1}{N}\sum_{i=1}^{N}\frac{TP(i)+FN(i)}{TN(i)+TP(i)+FP(i)+FN(i)}\times 100 \qquad (4.11)$$

where, N = number of samples

Accuracy alone does not contain enough information to conclude a decision about the performance of the classifier. Therefore, in this work other parameters are also included and analyzed while selecting the classifier.

- Precision specifies the detailed characteristics of features. It analyzes the number of times the prediction of classifier is correct and is calculated using Equation 4.12.

$$precision = \frac{1}{N}\sum_{i=1}^{N}\frac{TP(i)}{TP(i)+FP(i)} \tag{4.12}$$

- True Positive Rate (TPR) or Sensitivity describes that how many samples of a particular class are correctly classified. It should be maximum for the better classification performance and is calculated using Equation 4.13.

$$Sensitivit\ y / TPR = \frac{1}{N}\sum_{i=1}^{N}\frac{TP(i)}{TP(i)+FP(i)} \tag{4.13}$$

- False Positive Rate (FPR) is another important characteristic for the assessment of classifier performance which describes that how many subjects are incorrectly classified (not belonging to a particular class). It should be minimum for the better classification and can be written as (1- specificity) where Specificity is calculated using Equation 4.14.

$$Specificit\ y /(1-FPR) = \frac{1}{N}\sum_{i=1}^{N}\frac{TN(i)}{TN(i)+FP(i)} \tag{4.14}$$

- Area Under the Curve (AUC) is a plot between TPR and FPR of each class. It is a measure of prediction performance of the classifier. The overall AUC has been considered as an average AUC for each class.

- F-score is also a significant measure of the classification performance which describes the relationship between sensitivity and precision of the classifier. It is calculating using Equation 4.15.

$$F-score = \frac{2\times TP}{(2\times TP+FN+FN)} \tag{4.15}$$

4.3. Results and Discussion

The process of diabetes diagnosis is performed in two parts. The first part incorporates two feature selection methods, namely PCA and modified t-test. In the second part, classification algorithms are applied on the selected features. The performance of classifiers is evaluated in terms of accuracy, precision, sensitivity, specificity, AUC, and Matthews correlation coefficient (MCC). Total nine classifiers of different families are applied and further, compared for their performance analysis.

4.3.1 Performance Evaluation for PCA

The performance of each classifier is evaluated for different Specific Explained Variance (SEV) in % of PCA. Figure 4.2 shows a comparison of accuracies of various classifiers for different SEV of PCA. It depicts that QSVM and EBaT classifiers attained the maximum classification accuracy of 95.81% with 100% SEV of PCA. The accuracies of all the classifiers are further analyzed using statistical analysis, which concludes that there is no significant change in the classification accuracies of all classifiers (except SKNN) up to 99% SEV of PCA. The SKNN classifier shown a significant change (p-value= 0.005) in the accuracy up to 99% SEV of PCA. However, 100% SEV of PCA shows a major change in the accuracies of all classifiers.

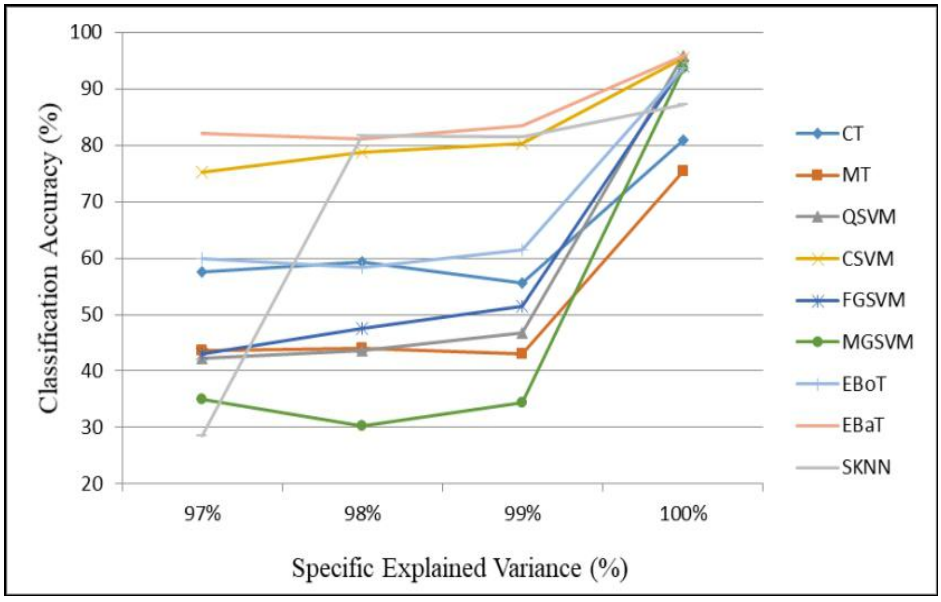


Figure 4.2 Comparison of accuracies of classifiers with different SEV (in %) of PCA

Table 4.5 Comparison of different SEV of PCA for average accuracy and precision

SEV in % of principal Components / Classifier	Accuracy (%)				Precision (%)			
	97%	**98%**	**99%**	**100%**	**97%**	**98%**	**99%**	**100%**
CT	57.49	59.28	55.69	80.84	50.78	52.19	48.15	71.17
MT	43.71	44.01	43.11	75.45	38.53	38.27	38.00	67.49
QSVM	42.28	43.71	46.71	**95.81**	37.89	40.42	42.83	**85.29**
CSVM	75.15	78.74	80.24	**95.51**	69.46	71.92	72.77	**85.04**
FGSVM	43.11	47.60	51.50	**94.01**	38.39	42.58	47.39	**82.67**
MGSVM	35.03	30.24	34.43	**93.71**	35.05	28.30	34.25	**82.41**
EBoT	59.88	58.38	61.38	**93.71**	53.64	52.95	55.85	**83.18**
EBaT	82.04	81.14	83.50	**95.81**	75.28	73.14	75.26	**84.19**
SKNN	28.44	81.74	81.44	87.13	27.76	73.98	73.82	78.05

Table 4.6 Comparison of different SEV of PCA for the average sensitivity and specificity

SEV in % of principal Components / Classifier	Sensitivity (%)				Specificity (%)			
	97%	**98%**	**99%**	**100%**	**97%**	**98%**	**99%**	**100%**
CT	52.54	53.31	50.47	73.54	75.22	76.28	73.75	87.89
MT	40.28	40.22	39.22	68.92	65.11	65.33	64.66	85.74
QSVM	37.59	39.26	42.03	**86.99**	64.26	65.78	68.07	**93.98**
CSVM	67.30	70.94	72.15	**86.70**	85.18	86.78	87.44	**93.86**
FGSVM	38.77	43.09	46.47	**85.63**	64.70	68.42	71.82	**93.09**
MGSVM	30.81	26.47	30.08	**85.37**	60.82	55.81	60.91	**92.98**
EBoT	54.15	52.77	55.29	**84.98**	76.72	75.91	77.62	**93.07**
EBaT	73.73	72.69	75.16	**87.00**	88.21	87.66	88.86	**93.76**
SKNN	25.66	73.15	72.85	78.50	52.39	88.03	87.90	90.51

Table 4.7 Comparison of different SEV of PCA for the average F-score and AUC

SEV in % of principal Components / Classifier	F-Score (%)				AUC			
	97%	**98%**	**99%**	**100%**	**97%**	**98%**	**99%**	**100%**
CT	50.98	52.35	48.69	72.11	0.83	0.82	0.80	0.93
MT	38.60	39.11	37.54	67.91	0.71	0.69	0.53	0.91
QSVM	37.19	38.82	41.70	**85.85**	0.67	0.70	0.71	**0.98**
CSVM	65.45	69.32	70.80	**85.62**	0.86	0.89	0.89	**0.98**
FGSVM	38.36	42.65	46.39	**83.56**	0.70	0.71	0.73	**0.97**
MGSVM	30.51	25.53	28.76	**83.29**	0.59	0.58	0.59	**0.97**
EBoT	52.37	51.46	53.80	**83.70**	0.81	0.85	0.83	**0.98**
EBaT	72.29	70.85	73.50	**85.33**	0.96	0.95	0.97	**0.99**
SKNN	25.45	70.99	70.94	77.54	0.95	0.95	0.96	**0.98**

Table 4.5 illustrates the average values of accuracies and precision with different SEV of PCA for all classifiers. The precision analysis represented almost same nature as shown by the accuracy. Each classifier exhibited a small change in the precision, up to 99% SEV of PCA while, for 100% SEV of PCA these classifiers shown large increment in precision. QSVM, CSVM, and EBaT classifiers attained maximum precision of 85.29, 85.04, and 84.19, respectively.

Table 4.6 shows the average values of sensitivity and specificity with different SEV of PCA for all classifiers. It is clear that QSVM, CSVM, and EBaT classifiers attained the maximum average sensitivity and specificity. Table 4.7 shows the average values of F-score and AUC with different SEV of PCA for all classifiers. QSVM, CSVM, and EBaT classifiers performed best among all classifiers. From Tables 4.5-4.7, it is clear that QSVM classifier shown best performance among all classifiers with 100% SEV of PCA.

PCA is the most effective and widely used technique for dimension reduction. However, it needs a detailed analysis before choosing it for the feature selection. PCA evaluates the significance and contribution of each feature of FV.

Figure 4.2 along with Tables 4.5-4.7 clearly indicate that there is a significant change in all the parameters of classifiers when SEV varies from 99% to 100%. It is observed that only 19 features contributed to obtain 99% SEV, while total 105 features combinedly provided 100% SEV. The main purpose of using PCA is to improve the computational complexity by reducing the feature size, but the aforementioned results concluded that PCA is unable to improve the computational complexity for attaining the best results. Hence, PCA is under-performing for our data set.

4.3.2 Performance Evaluation of Modified t-test

A detailed statistical analysis is implemented to evaluate the performance of the modified t-test. Figure 4.3 shows a comparison of the accuracies of different classifiers with different number of features. These features are sorted according to the t–score as described in Equation 4.10. The classification accuracies of all classifiers increase up to 30 number of features. Thereafter, there is almost negligible increment (or decrement in few cases) in classification accuracies when the number of features increased above 30. Thus, top 30

ranked features were enough to get the best accuracy while, PCA was using 105 features for the same as discussed in Section 4.3.1.

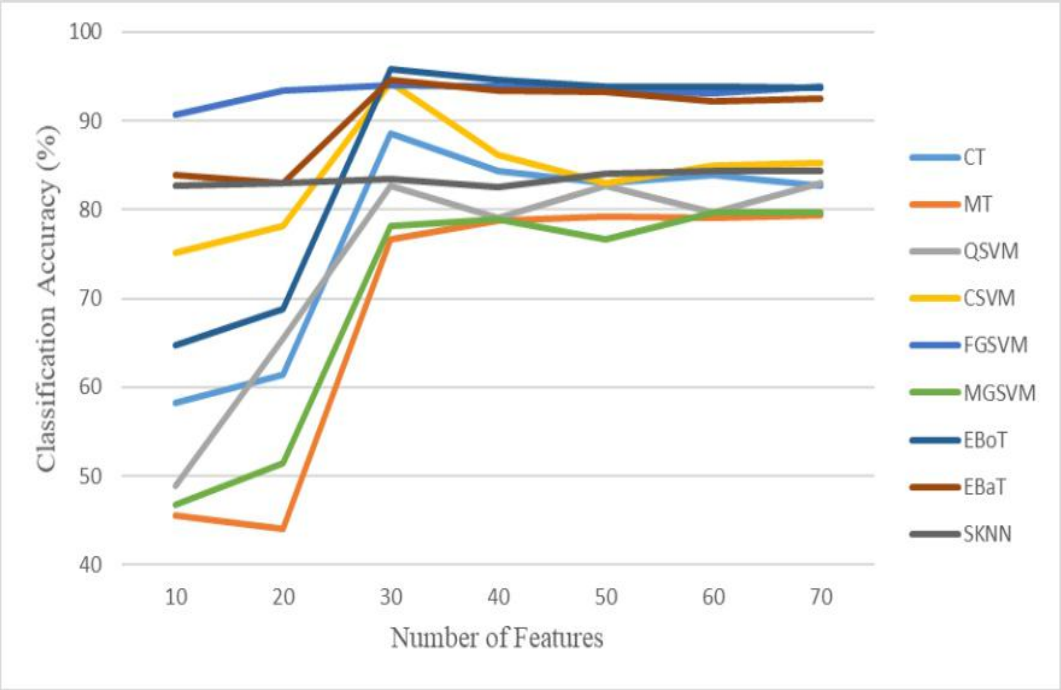


Figure 4.3 Performance comparisons of the accuracies of classifiers with different number of features ranked by the modified t-test

Table 4.8 Comparison of average accuracy and precision of different number of features ranked by the modified t-test

Parameter & No. of Features / Classifier	Accuracy (%)							Precision (%)						
	10	**20**	**30**	**40**	**50**	**60**	**70**	**10**	**20**	**30**	**40**	**50**	**60**	**70**
CT	58.30	61.38	88.62	84.43	82.93	83.83	82.63	50.64	53.85	79.51	75.70	74.54	74.83	73.92
MT	45.51	44.01	76.65	78.74	79.17	79.04	79.34	39.24	37.89	69.41	70.87	70.96	70.79	73.32
QSVM	48.80	65.48	82.63	79.04	82.63	79.64	82.99	42.90	57.44	73.66	72.49	74.18	67.63	74.74
CSVM	75.15	78.14	**94.31**	86.23	82.93	85.03	85.33	66.28	68.76	**83.56**	78.41	75.62	77.21	77.64
FGSVM	90.72	93.41	**94.01**	93.83	93.35	93.08	93.83	79.61	82.27	**82.67**	77.45	77.29	77.35	77.18
MGSVM	46.71	51.51	78.14	78.93	76.65	79.64	79.64	42.32	46.08	69.95	68.77	69.60	71.96	71.95
EBoT	64.67	68.86	**95.81**	94.62	93.89	93.82	93.69	55.90	59.82	**85.01**	82.56	81.92	81.26	80.51
EBaT	83.83	82.93	**94.61**	93.42	93.23	92.22	92.51	75.48	74.67	83.46	81.22	80.58	79.78	82.38
SKNN	82.63	82.93	83.50	82.61	84.13	84.37	84.43	74.17	74.69	81.93	75.49	75.49	75.44	75.20

Table 4.9 Comparison of average sensitivity and specificity of different number of features ranked by modified t-test

Parameter & No. of Features / Classifier	Sensitivity (%)							Specificity (%)						
	10	**20**	**30**	**40**	**50**	**60**	**70**	**10**	**20**	**30**	**40**	**50**	**60**	**70**
CT	52.17	54.94	81.01	76.98	75.35	76.24	75.09	75.52	66.34	69.09	85.10	91.81	67.98	79.30
MT	40.52	39.35	70.75	71.88	72.31	72.32	71.99	77.55	65.23	80.06	86.46	92.89	71.37	81.68
QSVM	44.21	59.35	75.79	71.66	74.84	71.97	75.06	91.58	87.07	89.15	**93.37**	93.09	87.39	**93.98**
CSVM	67.47	70.38	**85.79**	77.82	74.76	76.84	76.93	89.91	87.71	87.75	**90.40**	89.06	86.02	92.54
FGSVM	82.36	85.04	**85.63**	75.92	75.72	75.25	75.12	89.27	87.80	88.95	89.05	89.06	86.67	**92.47**
MGSVM	42.76	46.6	71.99	67.98	69.51	72.49	72.07	89.56	87.75	83.92	89.97	88.26	87.95	91.95
EBoT	57.61	61.54	87.13	82.97	83.04	81.46	80.66	89.14	88.10	89.18	**90.01**	89.06	87.77	91.60
EBaT	75.37	74.48	86.09	81.99	80.98	80.87	83.96	75.52	66.34	69.09	**85.10**	**91.81**	67.98	79.30
SKNN	74.25	74.55	84.07	76.02	76.02	76.61	76.24	77.55	65.23	80.06	86.46	**92.89**	71.37	81.68

Table 4.10 Comparison of average F-score and AUC of different number of features ranked by the modified t-test

Parameter & No. of Features / Classifier	F-Score (%)							AUC						
	10	**20**	**30**	**40**	**50**	**60**	**70**	**10**	**20**	**30**	**40**	**50**	**60**	**70**
CT	50.91	53.16	79.92	76.06	74.53	75.14	74.12	0.80	0.80	0.96	0.94	0.94	0.94	0.93
MT	39.41	38.13	69.69	71.05	71.24	71.09	71.73	0.71	0.68	0.92	0.93	0.91	0.91	0.91
QSVM	43.46	58.14	74.52	71.30	74.15	68.10	74.50	0.72	0.82	0.94	0.92	0.90	0.92	0.93
CSVM	65.81	68.80	**84.39**	77.43	74.54	76.41	76.60	0.86	**0.88**	0.97	0.93	0.92	0.93	0.94
FGSVM	80.55	83.08	**83.56**	73.80	73.80	73.59	72.82	0.96	**0.97**	0.98	0.94	0.93	0.92	0.93
MGSVM	41.94	45.82	70.66	68.10	69.32	71.97	71.73	0.83	0.76	0.91	0.91	0.91	0.92	0.94
EBoT	55.63	60.10	85.80	**82.32**	82.08	80.76	79.92	0.82	0.85	0.98	0.98	0.97	0.98	0.96
EBaT	74.05	73.15	84.51	**80.98**	80.25	79.90	**82.84**	0.94	0.95	**0.99**	0.98	0.98	0.98	0.98
SKNN	72.71	72.99	82.61	75.37	75.37	75.70	75.34	0.97	0.98	0.99	0.96	0.96	0.97	0.96

Tables 4.8-4.10 show that QSVM, CSVM, BaT, and EBoT performed best among all classifiers. There is always a significant increment in all the classification parameters whenever the number of selected features increases up to 30, whereas there is a minimal change in each parameter of classifiers whenever the numbers of features are increased above 30. Therefore, from the comparative analysis shown in Tables 4.8-4.10, it is depicted that the modified t-test feature selection technique performed better as compared to PCA.

In this research work, the complete FV has been prepared with the combination of physiological parameters and features of the specific areas of iris. Thus, it is important to check the presence of the physiological parameters in the obtained top 30 ranked features using modified t-test feature selection technique. It is found that BMI, systolic blood pressure, family history of diabetes, and age of subject are present in the ranked features.

Further, AUC analysis is carried out over the top performing classifiers for each class individually. The analysis shown maximum area for class 1, which indicates that the classifiers performed better in classifying the non-diabetic class (Class 1) from the diabetic classes (Classes 2, 3, and 4). The accuracy of classifiers for any individual diabetic class is also extremely important for the evaluation of diabetes diagnosis system. Based on the individual class performance only, the comments on class characteristics can be made and generalized. In addition, FN for classes 2, 3, and 4 with class 1 is crucial because predicting a diabetic person as non-diabetic is an erroneous result. Therefore, evaluation parameters like precision, sensitivity, and F-score become more essential for the classifier selection. The high classification rate attained in results also justifies the role of features like statistical, GLCM, and GLRLM along with physiological parameters for diabetic and non-diabetic subjects.

4.3.3 Comparison With Existing Techniques

The computed results (Figure 4.3 and Tables 4.8-4.10) have been compared with the existing techniques. The comparison is done based on feature types (physiological and/or iris based) and parameters of classifiers. Table 4.11 shows a comparison between our computed results with the result other state of the art methods.

Table 4.11 Performance comparison with existing techniques

Author	Physiological Parameters	Parameters of Specific Areas of Iris	Classifier and its Parameters						
			Classifier	Accuracy (%)	Precision	Sensitivity (%)	Specificity (%)	AUC	F-Score
Meng *et al.* 2012 [143]	√	×	LR	76.13	-	79.59	72.74	-	-
			ANN	73.27	-	82.18	64.49	-	-
			DT	77.87	-	64.49	75.13	-	-
Heydari *et al.* 2015 [47]	√	×	SVM	81.19	-	0.8309	0.6253	-	-
			ANN	97.44	-	0.9749	0.9687	-	-
			CT	95.03	-	0.9784	0.6697	-	-
			KNN	90.85	-	0.9989	0.0091	-	-
			BN	91.60	-	0.9457	0.6192	-	-
Bansal *et al.* 2015 [63]	×	√	SVM	87.5	-	0.95	0.90	-	-
Dwivedi 2017 [46]	√	×	ANN	77	0.80	0.86	0.60	-	0.83
			SVM	74	0.80	0.82	0.61	-	0.81
			KNN	73	0.78	0.82	0.56	-	0.80
			NB	75	0.82	0.80	0.67	-	0.81
			LR	78	0.80	0.89	0.59	-	0.84
			CT	78	0.73	0.77	0.57		0.77

Samant *et al.* 2018 [149]	×	√	BT	75.20	-	-	-	-	-
			RF	89.97	-	0.9973	0.9178	-	-
			AB	89.38	-	0.9978	0.9138	-	-
			SVM	88.53	-	0.9779	0.9111	-	-
			GL	62.35	-	-	-	-	-
			ANN	67.29	-	-	-		-
Present Chapter	√	√	CT	88.62	79.51	81.01	69.09	0.96	79.92
			MT	76.65	69.41	70.75	80.06	0.92	69.69
			QSVM	82.63	73.66	75.79	89.15	0.94	74.52
			CSVM	94.31	83.56	85.79	87.75	0.97	84.39
			FGSVM	94.01	82.67	85.63	88.95	0.98	83.56
			MGSVM	78.14	69.95	71.99	83.92	0.91	70.66
			EBoT	95.81	85.01	87.13	89.18	0.98	85.80
			EBaT	94.61	83.46	86.09	69.09	0.99	84.51
			SKNN	83.50	81.93	84.07	80.06	0.99	82.61

Where; LR- Logistic Regression, ANN- Artificial Neural Network, DT Decision Tree, SVM- Support Vector Machine, CT- Classification Tree, KNN- K Nearest Neighbors, BN- Bayesian Network, NB-Naive Bayes, RF-Random Forest, AB- Ada Boost, GL- Generalized Linear Models

Table 4.11 depicts that the combination of physiological features and features of specific areas of iris show a better performance as compared to their individual performance. Heydari et al. [47] achieved a higher accuracy (max 97.44%) for prediction of the diabetes, but the classification has been performed over highly imbalanced data set (ratio of diabetic to non-diabetic subjects is equals to 2305:231). Using considering physiological parameters, Dwivedi [46] obtained maximum classification accuracy of 78% using CT classifier, whereas Meng et al. [143] achieved maximum accuracy of 77.83% using DT classifier. Bansal et al. [63] and Samant et al. [149] have attained maximum accuracy of 87.5 % and 89.97% by using features of specific areas of the iris, respectively. However, in the presented work, the maximum accuracy of 95.81% has been achieved using both kinds of features combined.

Medical imaging and CAD are used as low cost and non-invasive techniques for disease diagnosis. Duration of diabetic state along with its diagnosis is also crucial as it provides the information of possible complications due to diabetes. In the presented study, with the help of the supervised learning technique, a framework is designed not only to diagnosis diabetes but also to its duration. Total 334 subjects were screened and features from the specific areas of iris along with physiological features were combined together to make a complete FV. PCA and modified t-test feature selection techniques were compared with the results of 9 different classifiers. Modified t-test with EBoT classifier have shown best accuracy of 95.81%. The proposed framework classified the non-diabetic and diabetic patients (having different duration of diabetic state) with good accuracy. Further, a portable and hand-held device can be developed in future using the proposed methodology. The device can diagnose the diabetes and its complications by analyzing the images and physiological parameters of the subject. In addition, non-medical practitioners can easily operate it. Therefore, it might help the people of rural areas where basic facilities to diagnose diabetes are not available.

Chapter 5

Assessment of Kidney problem in the diabetic patients using iris

Diabetic Kidney Disease (DKD) is a major cause of kidney failure and it also increases the risk of having heart and blood vessel disease. It is a chronic disease that can develop in the patients having diabetes and high blood pressure over the period of time. The currently available pathology to diagnose DKD involves the blood test which is a complex and agitate procedure for the patents. In this chapter, a non-invasive procedure is presented for the diagnosis of the DKD using iris images. The main aim of this study is to assess the concept of iridology for the diagnosis of DKD. To prove the hypothesis, first a total of 130 subjects were evaluated and labeled according to their health condition as (1) not diagnosed with diabetes, (2) diagnosed with diabetes but not diagnosed with DKD, and (3) diagnosed with diabetes and DKD both. Pre-image processing techniques are implemented to get the ROI from the iris. First Order Statistical (FOS) and second-order textural features are extracted from the cropped ROI. Finally, with a different sets of features, different classification models are compared to find the best suitable model for this type of problem.

5.1 Introduction

Diabetes is a major cause of kidney disease worldwide. Around 40% of diabetic patients also diagnose to have kidney disease. The primary activity of kidneys is to filter out waste materials and extra water from the blood and make urine. Kidneys also assist to maintain the blood pressure and generate hormones that are necessary to stay healthy. Improper working of kidneys can building up the waste in the body and it can cause other health problems. Kidney damage due to diabetes is a chronic diseases and it is also called Diabetic Kidney Disease (DKD) or Diabetic Nephropathy (DN) [150]. A high blood glucose level due to diabetes can damage the blood vessels of the kidneys. Most of the patients with diabetes also develop the problem of high blood pressure, which is the prime cause kidney damage [50]. Having diabetes for a longer time increases the chances that one will have damaged kidneys. Indians, Hispanics/Latinos, African, and, American develop DKD at a higher rate as compared to Caucasians.

DKD does not show any specific symptoms to most of the patients. The only method to know DKD is to do a regular kidney checkup. Health-care professionals perform blood and urine tests to check DKD. The best way to prevent DKD is to maintain blood glucose and blood pressure level all together. Healthy lifestyle, habits and proper medication as prescribed can help to attain these goals and improve overall health. Early detection of the DKD is important,

as it is associated with an increased risk of death in general, mainly owing to cardiovascular diseases. At present, the diagnosis of DKD is based on the abnormal concentrations of urinary albumin caused mainly by glomerular hyperfiltration, a condition known as microalbuminuria or incipient nephropathy. Further progression of the disease is defined by the presence of macroalbuminuria (proteinuria) or overt nephropathy.

CAM therapies are quite popular for the diagnosis of the chronic diseases like diabetes *etc.* [149, 151]. Iridology is an alternative medicine science which correlates iris patterns, colors, tissue weakness, breakage and other characteristics, that can acquire evidence about patient's systemic health [10]. It reveals weakness or brakeage in tissues long before the symptoms appear. Iridology practitioners match their interpretations to iris chart in which iris is divided into several zones corresponding to specific body organs [49]. Traditionally, iridologists perform eye analysis manually using a magnifying glass, slit lamp, and one plastic iridology chart. As technology evolves, a manual annotation software is developed where the chart is drawn onto eye images. The advanced modern computer vision techniques are used to propose automatic iridology based diagnosis approach [152].

Diabetes diagnosis techniques that utilize the concept of iridology with supervised machine learning techniques are popular and have achieved high accuracy and reliability [160]. Samant *et al.* [149] proposed a machine learning based technique to diagnose the diabetes using iris images based on iridology. Hussein *et al.* [57] presented a method to diagnose the kidney disease which uses the impression of the iridology and modern machine learning techniques. The results show classification accuracy of 93% for the healthy subjects and 82% for the subjects with kidney problems using Neural Network (NN) classifier. Alam *et al.* [154] presented an iridology based study to detect the liver disease. Authors used GLCM based textural features to train the ANN classifier and achieved the accuracy of 91.42%.

Simon *et al.* [155] analyzed the condition of subjects suffering from kidney disease by a creatinine intensity and compared with the subjects which are free from any kind of kidney disease. Total 146 subjects were considered and the images of both the iris were captured. The coded images were shown to 3 different iridology practitioners and 3 ophthalmologists. Young *et al.* [51] studied the relationship of Angiotensin Converting Enzyme (ACE) and iris constitution. ACE is genetic marker of the vascular disease. Sulistiyo *et al.* [156] proposed technique for the early detection of dyspepsia from iris images. Authors have used Linear Discriminant Analysis (LDA) feature extraction and Cascade Correlation Neural Network (CC-NN) classifier to satisfy their hypothesis.

The studies for the diagnosis of diabetes and related diseases with the help of iris images motivates researchers and medical practitioner to offer new directions in healthcare technologies. These unorthodox technologies may play a major role in the diagnosis and prevention of the chronic diseases. With this point of view, iridology can show the direction

for an accurate and user friendly healthcare model. In this case study, assessment of the diagnostic validity of DKD using iris images is described. A fully automatic framework based on the combination of the image processing and machine learning techniques is presented. The specific defined areas correspond to the kidney in the iris are cropped out with the help of image processing techniques. Then, from the ROI, FOS (First order Statistical), GLCM and GLRLM features are extracted. In order to select the most significant features, correlation based feature reduction technique PCA is used. Finally, the classification models are applied on the labeled features according to the DKD conditions of the subjects.

5.2 Methodology

The main aim of this research work is to perform a pilot study for the assessment of the diagnosis of DKD using the concept of the iridology. On the same time, to develop a procedure as a standard for iridology based diagnosis system using advanced machine learning methods. Doing this, we can accompany the conventional diagnosis methods and consider the treatment progress. The procedure adapted in this chapter is summarized in Figure 5.1 and is explained as follows:

5.2.1 Data-set Collection

DKD is a chronic disease and worldwide public health problem for the diabetic population. DKD is often associated with the higher risk of chronic renal failure and cardiovascular disease. For this study, total 43 non-diabetic and 83 diabetic patients ware considered. Out of those 41 were suffering with DKD. To diagnose the diabetes, fasting and 2-hour postprandial tests were performed. To check the kidney damage degradation, Glomerular Filtration Rate (GFR) was calculated. If calculated GFR is less than 90 mL/min/1.73 m^2 then diabetic patient is said to be diagnosed as DKD. Images of both the eyes were captured using iris scanner I-SCAN-2 by Cross-match Technologies, USA.

5.2.2 Pre-Image Processing Techniques

Iris segmentation and normalization techniques are used in pre-image processing. The main parameters of iris are center and radius of inner as well as outer circular parts. Although, a number of segmentation algorithms have already been addressed in [59, 144, 145], however, the most popular and acknowledged iris segmentation algorithm based on Circular Hough Transformation (CHT) [69] is used in this work.

Rubber-sheet normalization algorithm converts the circular iris into a fixed rectangular section [59]. The output of the rubber-sheet normalization model is invariant to camera direction, alignment, and position. It also removes the external factors like uneven illumination and dimensional irregularities [80]. In this research work, Daugman's rubber-

sheet model [59] is used to convert segmented iris into a rectangle of a constant size of 360×720 (pixels). After that from the normalized iris, ROI has been cropped according to the iridology chart given by Jensen [49]. The kidney is shown between 6.30 to 7 o'clock in the left eye and in between 5 to 5.30 o'clock in the right eye. Therefore, the ROIs have been cropped corresponding to these specific areas of the iris as shown in figure 5.2.

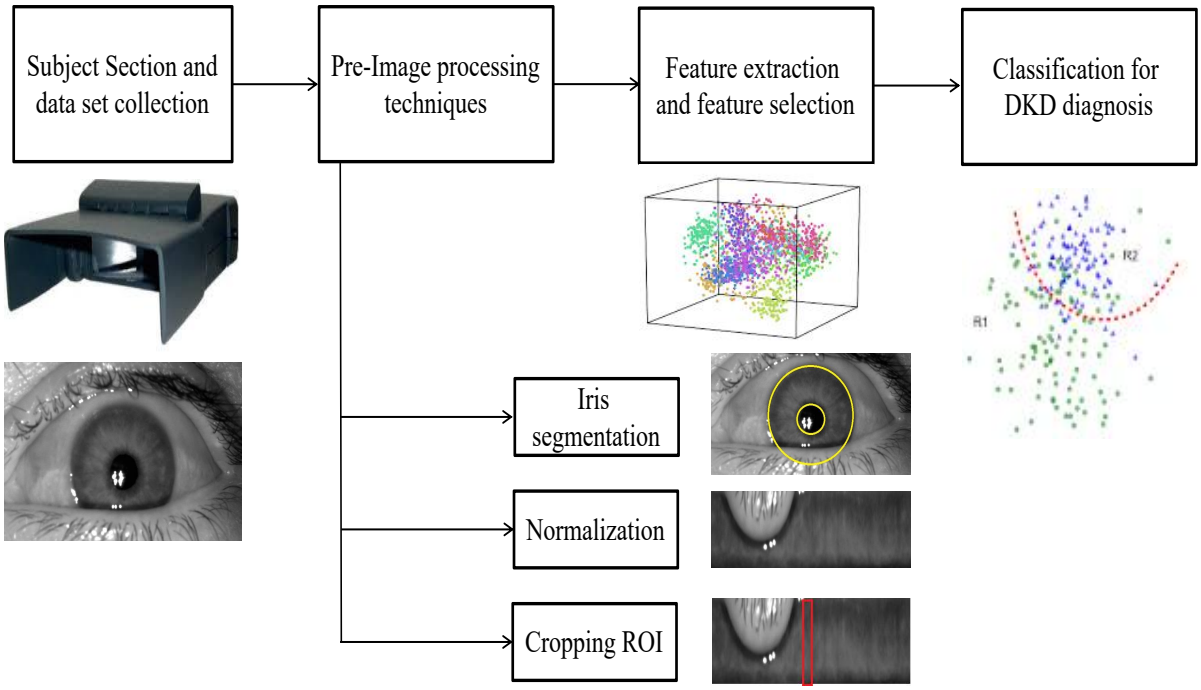


Figure 5.1 Block diagram of the adapted methodology

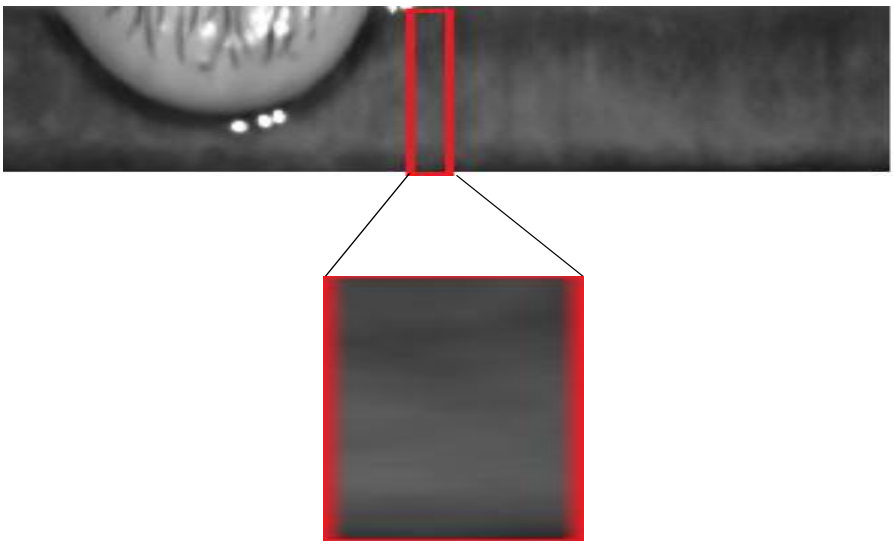


Figure 5.2 Extraction of Region of Interest (ROI)

5.2.4 Feature Extraction and Experiment Design

The main aim of feature extraction is to obtain important and indistinct information from the ROI. Total 5 First Order Statistical (FOS), 19 GLCM and 7 GLRLM based features have been extracted from cropped ROI for both the iris. The experiment is designed to evaluate the contribution of the feature sets for diagnosing the DKD. In different experiments effect of only FOS, GLCM and GLRM features are evaluated for the diagnosis of DKD. Also, combined feature set is also used for the diagnosis so that the role of individual features and dominating feature sets can also be computed.

5.2.5 Classification

The classification is carried out with the help of different kernels of two different classifiers, k-Nearest Neighbour (KNN) and SVM. Two kernels named fine Gaussian and cubic kernels of SVM are considered while three kernels of KNN named subspace, weighted and fine are considered for further evaluation. SVM is one of the most popular classifier used in biomedical applications. It is accurate and easy to understand at the same time it has a higher learning capability. While KNN is the most frequently used classifier in many benchmark applications, its simplicity and high performance make it preferable over other classifiers.

5.3 Result and Discussion

For the experiment, total 62 features are extracted from the group of FOS, GLCM and GLRLM features. All the experiment results are implemented over the normalized feature set to minimize the influence of the subsequent steps over the features. The experiments were performed to design an efficient classification system for the prediction of the DKD.

5.3.1 Classifier Analysis by FOS Features

This experiment has been performed to check the effect and the contribution of only the FOS features on the prediction of the DKD. To evaluate the performance of the classifiers: accuracy, sensitivity, and specificity have been calculated with the help of 10-fold cross validation technique. Table 5.1 shows the classification parameters for all the selected classifiers. SVM Fine Gaussian classifier outperforms among all the classifiers where average accuracy calculated as 94.0%. While average sensitivity and specificity calculated as 94.3% and 94.1% respectively. Sensitivity for non-diabetic class is minimum as compare to other classes.

5.3.2 Classifiers Performance Analysis by GLCM Features

GLCM are second order statistical features and calculated according to the correlation between the pairs of pixels in a particular direction. In this work, first degree of correlation is

considered to reflect the correlation between adjacent pixels. This experiment has been designed to check the performance of only GLCM based features in the diagnosis of the DKD. Table 5.2 shows the classifier performance analysis of GLCM features only.

5.3.3 Classifiers Performance Analysis by GLRLM Features

GLRLM is computed according to the number of runs of a particular pixel in the specified direction. Features extracted from the GLRLM are also second-order features. For this work, total seven GLRLM features are extracted and in this experiment, classifier performance is evaluated only for the second-order GLRLM based textural features. Table 5.3 shows the performance according to the GLRLM features.

5.3.4 Classifiers Performance Analysis Using Combined Feature Vector

In this experiment a compete FV is designed as the combination of all the extracted features *i.e.* FOS, GLCM and GLRLM features as shown in figure 5.3. In a nutshell total 62 features are extracted for the analysis. Table 5.4 shows the classifier performance according to the complete FV. Figure 5.3 shows the flow diagram of this designed experiment.

Table 5.1 Average accuracy, sensitivity and specificity for FOS features

Classifier	Kernel	Accuracy (%)	Sensitivity (%)				Specificity (%)			
			S_1	S_2	S_3	S_{AV}	SP_1	SP_2	SP_3	SP_{AV}
SVM	Fine Gaussian	94.0	87	99	97	94.3	98.1	91.8	92.5	94.1
	Cubic	72.2	67	77	70	71.3	80.3	69.2	73.5	74.3
KNN	Subspace	85.3	88	85	84	85.6	84.5	85.4	86.2	85.4
	Weighted	76.2	77	81	72	76.6	75.9	59.9	79.7	71.8
	Fine	79.0	88	85	73	82.0	78.6	75.3	83.4	79.1

Table 5.2 Average accuracy, sensitivity and specificity for GLCM features

Classifier	Kernel	Accuracy (%)	Sensitivity (%)				Specificity (%)			
			S_1	S_2	S_3	S_{AV}	SP_1	SP_2	SP_3	SP_{AV}
SVM	Fine Gaussian	91.7	82	97	99	92.7	98.0	89.2	88.5	91.9
	Cubic	73.0	72	72	74	72.7	73.3	73.4	72.6	73.1
KNN	Subspace	78.6	78	86	73	79.0	78.7	75.0	82.4	78.7
	Weighted	76.6	82	73	78	77.7	75.3	79.0	75.8	76.7
	Fine	77.4	80	75	78	77.7	76.6	79.1	76.7	77.5

Table 5.3 Average accuracy, sensitivity and specificity for GLRLM features

Classifier	Kernel	Accuracy (%)	Sensitivity (%)				Specificity (%)			
			S_1	S_2	S_3	S_{AV}	SP_1	SP_2	SP_3	SP_{AV}
SVM	Fine Gaussian	94.4	89	98	98	95.0	97.5	92.9	93.0	94.4
	Cubic	71.4	72	73	70	71.7	81.7	70.6	73.3	75.2
KNN	Subspace	84.1	84	86	83	84.3	84.0	83.1	85.3	84.1
	Weighted	77.4	85	78	74	79.0	75.8	76.8	80.1	77.7
	Fine	80.6	91	73	84	82.7	78.0	86.4	78.6	81.0

Table 5.4 Average accuracy, sensitivity and specificity using combined feature vector

Classifier	Kernel	Accuracy (%)	Sensitivity (%)				Specificity (%)			
			S_1	S_2	S_3	S_{AV}	SP_1	SP_2	SP_3	SP_{AV}
SVM	Fine Gaussian	94.5	86	100	100	95.3	100	91.7	94.2	95.3
	Cubic	81.7	81	83	81	81.7	81.1	81.1	82.3	81.5
KNN	Subspace	84.9	90	86	80	85.3	83.1	84.6	87.9	85.2
	Weighted	73.8	83	73	72	76.0	72.2	74.6	75.3	74.0
	Fine	83.3	87	83	81	83.7	82.3	83.3	84.6	83.4

where, S_1= Sensitivity for non-diabetic class, S_2= Sensitivity for class having diabetic but non-DKD, S_3= Sensitivity for DKD class, S_{OV}= Average Sensitivity
SP_1= Specificity for non-diabetic class, SP_2 =Sensitivity for class having diabetic but non-DKD, SP_3= Specificity for DKD class, SP_{OV}= Average Specificity

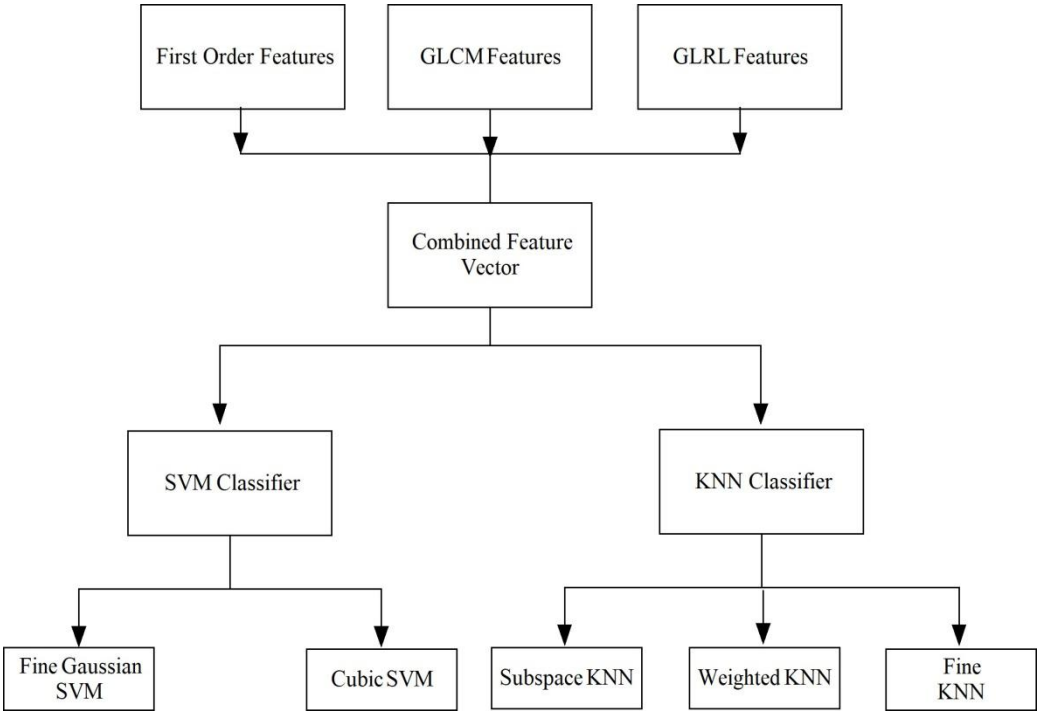


Figure 5.3 Block diagram of the experiment designed for combined feature vector

In this chapter, a pilot study is presented to diagnosis the DKD using advanced machine learning techniques. The objective is to perform an unbiased primary study to attract more and more researchers form biomedical and machine learning backgrounds to develop an outperforming model for the fast, easy, and early diagnosis of DKD. In this chapter, a comparison of the contribution of statistical and texture features (GLCM and GLRLM) is presented. Classification is performed using the SVM, and KNN classifiers and results are compared. The proposed model shows excellent results and encourage the more exhaustive and detailed study. These results confirm that the FOS features are best appropriate for designing the DKD diagnosing system while, at the same time, a combination of different textural features gives the added advantage and good classification accuracy. As this is a pilot study so there is a need of more regress study with more number of subjects can be performed. These findings indicate that analysis of the features of the specific areas of the iris are used to predict the DKD condition but performed on a closed group population. If replicated in a larger population, this novel approach may assist the clinician in determining treatment-efficacy.

Chapter 6

Sclera Segmentation and Recognition

Sclera recognition is relatively new, and it needs to explore more. The sclera of the human eye contains the unique blood vessel patterns that make it a potential tool for personal identification. In this chapter, a fast and robust sclera segmentation algorithm is presented, and unique features for sclera based recognition system are proposed. The most versatile data-set in terms of image quality (UBIRIS V1) is selected for segmentation and recognition purposes. For sclera segmentation, and unsupervised algorithm based on pixel mapping and cauterized grayscale is presented. Thereafter, Local Binary Patterns (LBP) features and Gabor Wavelet Transform (GWT) of segmented sclera are further used for recognition. To find out the optimized feature set of LBP, Principal Component Analysis (PCA) is used. The classification was performed using the Multiclass Support Vector Machine (MSVM). The results of the classifier for the recognition task are presented in terms of overall system accuracy and sensitivity. It is notionally proved that the proposed technique is highly reliable and accurate for the execution of the sclera recognition.

6.1 Introduction

With the advancement of computer technology, new moralities of human identification are getting popular. Sclera identification is a comparatively new biometric technique that needs to be explored. The sclera is the white part of the eye, which acts as the outermost protective layer for the inner parts of the eye. The sclera of the human eye contains unique blood vessel patterns. Sclera consists of four different tissue layers, which make it slightly opaque. A high degree of randomness and stability throughout the lifetime of blood vessels make it a reliable and ideal choice for personal identification [157]. Das et al. [91] used Daugman'sIntego Differential Operator (IDO) for the selection of seed points. The method shows high accuracy; however, the computational complexity is high due to the calculation of the center and the radius of the iris.

Zhou et al. [158] proposed a quality fusion of iris and sclera based features for biometric. Tankasala et al. [159] presented a K-means, clustering-based sclera segmentation technique. Authors have used GLCM features for person recognition. A similar kind of study was presented by K. Oh et al. [89] in which the Local Binary Patterns (LBP) based features have been used for identification and matching. A number of research articles have already been proposed for sclera segmentation and features based matching in eye-based biometric. However, there is still space for improvement in the sclera segmentation and feature extraction techniques. There is a need for accurate and robust sclera segmentation algorithm

so that the precise sclera can be extracted from the image of the eye.Moreover, an optimized set of features is also required to attain high accuracy for matching. Therefore, in this chapter, an unsupervised learning based sclera segmentation technique is presented. For the matching, a comparison of LBP-PCA based optimized feature set and GWT features are presented by classification with MSVM classifier.

3.2 Proposed Method

The whole framework of sclera based person identification is performed into two stages first is sclera segmentation from the image of eye and second is feature extraction and matching. Figure 6.1 shows the block diagram of the proposed model.

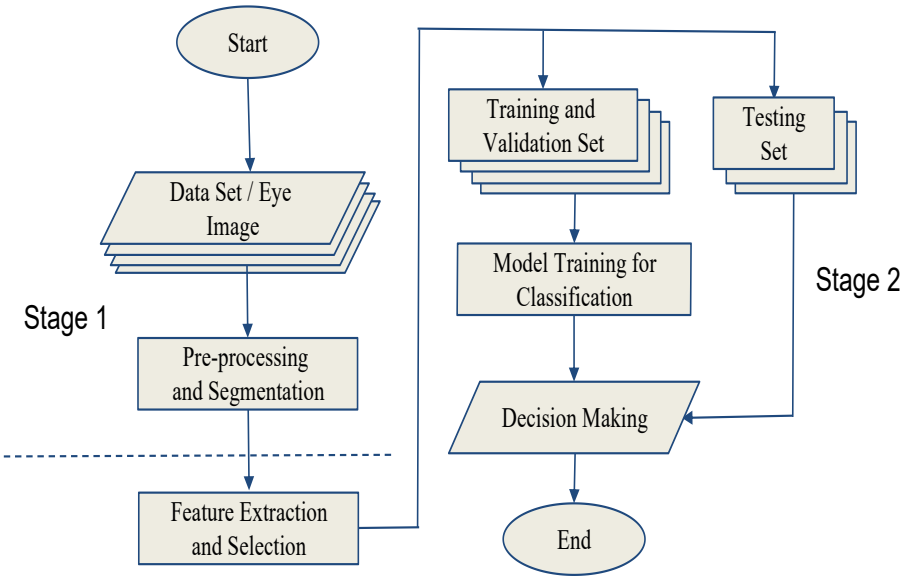


Figure 6.1 Flowchart of proposed methodology

6.2.1 Data Set

In this study to evaluate the performance of the proposed model UBIRIS V1 [160] data set has been used. UBIRIS V1 data set is the collection of total 1877 images from total 241 subjects. The images were captured in two scenarios. For this study, we have selected 24-bit color images of size 200x150. The images of both the scenarios are merged for the experiment design in order to make system more robust.

6.2.2 Pre-Processing and Sclera Segmentation

Figure 6.2 shows the block diagram of the pre-image processing and the sclera segmentation. The region of the sclera is almost white and have maximum intensity in the image of the eye. The algorithm is performed in 3 different parts.

In the first part, the color image is converted into the gray level by using an adaptive histogram technique. Each channel R, G, and B of the color image is mapped into a fixed intensity range using 2 stages sigmoid function. Since the red channel has higher intensity values for human skin as compared to the green and blue channel in addition to this blue and green channel have higher intensity differences between skin and sclera. Therefore, the red channel is subtracted from a combination of the blue and green channel in order to create a grayscale image having intensity range [0-1]. Finally, the output gray image is mapped between the grayscale range of [0-255].

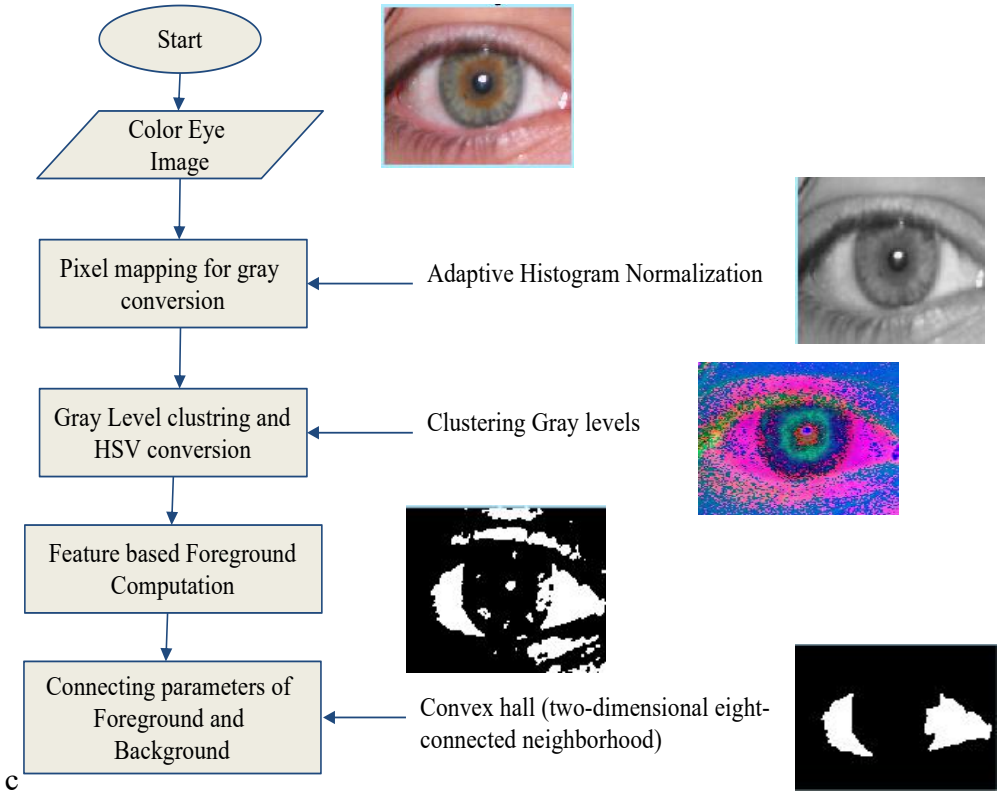


Figure 6.2 Flow chart of the sclera segmentation algorithm

In the second part feature based adaptive threshold technique is used for gray level clustering and background and foreground are estimated. Finally, in the third part components of the foreground are connected using convex hull enclosing of the foreground region.

6.2.3 Feature Extraction

The segmented sclera is highly reflective in nature; hence, sclera vascular patterns appear blurry, and overall the ROI comes out as low contrast. It is very important and necessary to enhance the patterns of vessels in the segmented region before extracting the features. The Gabor filters [161] are excellent in an approximation of the revelation processes of the principal visual cortex. Since the vascular patterns have multiple orientations, hence we have calculated directional Gabor Wavelet Transformation (GWT) for vascular pattern

enhancement and ultimately used it as a feature set of sclera region. GWT works as a band-pass filter, which enhances the bars and edges of the localized areas. GWT also gives a good estimation of the orientation and selectivity.

The extraction of the features involves building an unswerving mathematical model of the vessel pattern from segmented sclera for person identification. The traditional statistical, textural, and GLCM [149, 151] based features produce occlusions which are sometimes very difficult to manage. Also, these features decrease the versatility of the overall system. In the present work, LBP features are extracted from the segmented sclera region from the image of the eye. LBP is a simple and effective way for textural and tissue-related studies. LBP textural operator labels to each and every pixel of the sclera region by thresholding the 8 neighbors and present the results in the form of a mathematical number. Simultaneously, to reduce the complexity, the PCA is applied for the selection of the optimal size of features. In addition to this, GWT features are also extracted from the segmented sclera. Hence, in this study, two different types of features are extracted, and matching results are compared for both types of features.

6.2.4 Classification and Matching

Biometric algorithms generally aim to provide a reasonable answer for all possible inputs; therefore, classification plays an important role in feature-based matching. The MSVM classifier was used for classification in this research work. MSVM is a popular supervised classification technique that performs an embedded mapping onto the feature space of a higher dimension. After the mapping, MSVM finds a hyperplane having maximal edges to divide features from its higher-dimensional space. Kernel plays an important role in the performance of the MSVM classifier. Although many kernels of SVM are available, however, the frequently used kernel functions are Radial Basis Function (REF), polynomial, linear, etc. In the present work, we have used the RBF kernel of the MSVM classifier. The classifier was trained through 10 fold cross-validation. The overall accuracy was considered as the average of all accuracies obtained by the classifier to improve the performance.

6.3 Result and Discussion

The images of both the scenarios are combined of the UBIRIS V1 data-set. Both scenario contain 5 images of a each subject. For training and testing purpose ratio of 4:1 is maintained. Moreover, the training is done through 10 fold cross-validation technique. The analysis is performed on Matlab R 2019b version. For the best visualization of segmentation and recognition results, a graphical user interface has been created as shown in figure 6.3.

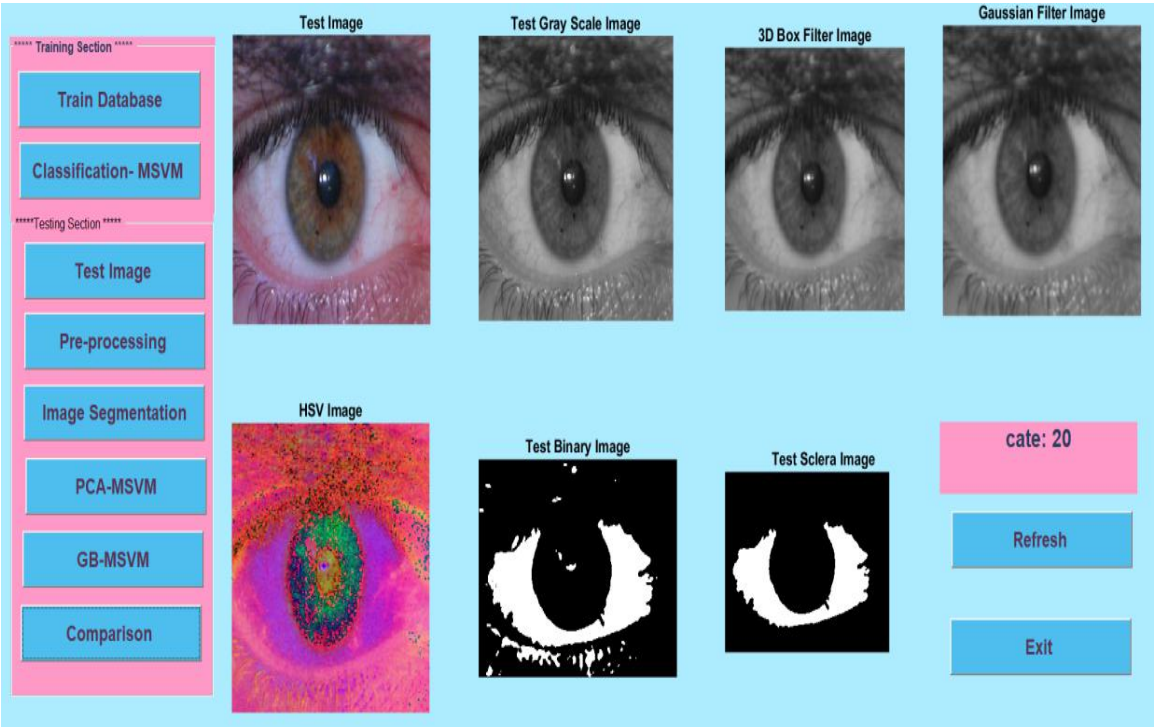
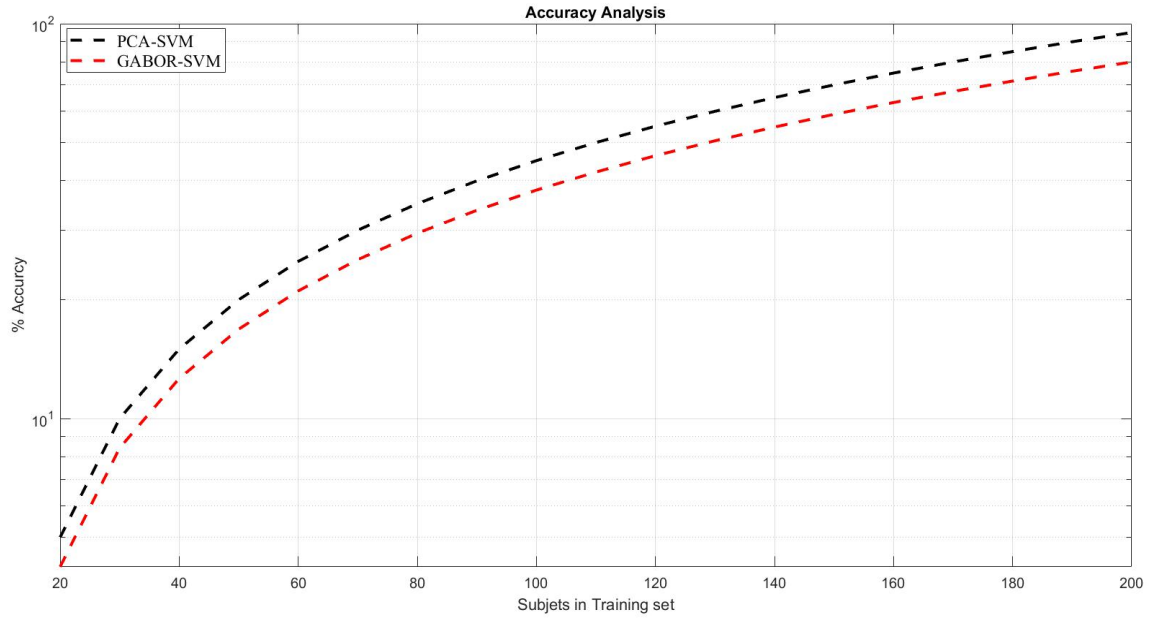


Figure 6.3 Graphical user interface designed for visualization of segmentation and recognition results

A comparison of the recognition results obtained from the LBP-PCA and GWT features is presented in terms of accuracy and sensitivity. The MSVM classifier is trained with the features obtained from sclera region of 241 subjects. Figure 6.4 shows the overall recognition accuracy and sensitivity comparisons for both types of the features.



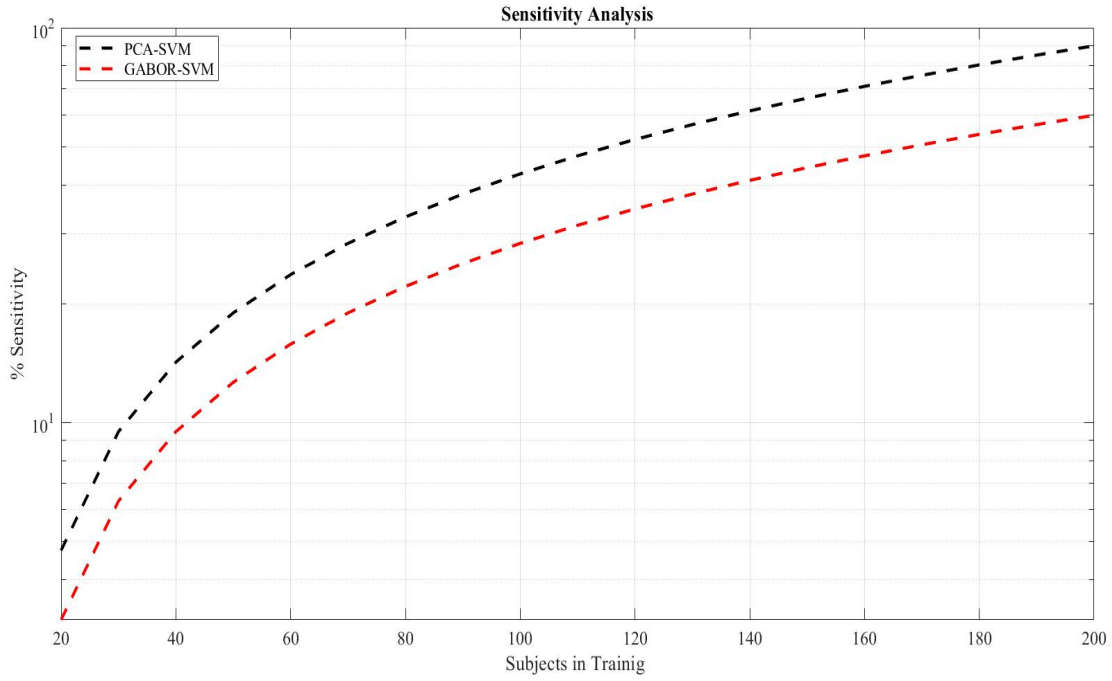


Figure 6.4 Comparisons of Accuracy and Sensitivity of the LBP-PCA and GWT features

Figure 6.4 clearly shows that, LBP-PCA feature set gives the better results as compared to the GWT features. Accuracy for LBP-PCA features is achieved as 95.3% while for GWT features attained maximum accuracy of 80.8%. While sensitivity is calculated as 90.7% and 60.5% for LBP-PCA features and GWT features respectively.

This chapter consists of simple pixel intensity and HSV model based sclera segmentation technique and a statistical approach for the feature based identification model. Weighted average gray conversion method and histogram equalizer technique were used for pre-image processing. Thereafter, LBP and GWT features were extracted from the segmented sclera. Finally, RBF kernel of the MSVM classifier was used for the classification and matching. The work on UBIRIS V1 database shows promising results. In future, the sclera based recognition system will be used identification and surveillance in parallel with established biometric techniques.

Chapter 7

Conclusion and Future Work

7.1 Conclusion

Diabetes leads to severe complications that can affect different organs of the body. As diabetes progresses, it causes numerous complications like amputation, kidney failure, stroke, heart failure, and heart attack. Timely and accurate diagnosis of diabetes can slow down the progress of the disease, and similarly, the complications due to diabetes can also be controlled up to some extent. In this thesis, image-based Computer Aided Diagnosis (CAD) techniques for the diagnosis of diabetes and other related diseases are explored. The use of CAD systems are easy and effective means to assist in diagnosis to the medical experts. Different reasons can hamper the accurate diagnosis of a iridologists and resulting loss in faith in the techniques are identified. The possible reason is the lack of standard procedure and an automated methodology for the diagnosis. This research work gives the solution to both problems. A fully automatic standard procedure is proposed for the diagnosis based on the concept of iridology. Machine learning techniques and theoretical concepts of iridology have been integrated to present a non-invasive, accurate, and reliable model for the diagnosis of diabetes, its duration, and related disease. The physiological parameters related to genetics and/or lifestyle are often associated with the diabetic health condition. Therefore in this research work, the framework was designed in two ways, *i.e.,* by considering iris based features, and thereafter by combining the physiological features with the features of specific regions of iris to achieve better results.

Changes in the features of the specific areas of iris were mapped with the diabetic health condition of an individual. The methodology was designed and employed to improve system accuracy, sensitivity, specificity, and other parameters. A strong relationship between the features of the specific areas of the iris and chronic diseases like diabetes, its duration, and Diabetes Kidney Disease (DKD) was established. It was found that, by observing the changes in statistical and Discrete Wavelet Transformation (DWT) features of the specific areas of the iris, the diabetic state could be predicted with better accuracy (89.97%). The complications due to diabetes also increase as diabetes progresses; therefore, the methodology was also designed to diagnose the duration of diabetes. It was also found that these changes in the specific areas of the iris become more prominent as diabetes progresses.

The physiological features and features of ROI of IRIS images were fused for a more accurate screening of diabetes and its duration. Statistical features, GLCM features, and GLRL features were extracted from ROI. These features were combined with features like BMI, family history, sex, genetics, another disease history, *etc.* The proposed framework classifies the non-diabetic and diabetic subjects (having a different duration of diabetic state) with adequate accuracy (95.6%). Therefore, it was concluded that a hybrid non-invasive

CAD system can be used as a diagnostic tool for diabetes and its duration diagnosis. Since the statistical, GLCM, and GLRL features have shown remarkable results; therefore, the DKD framework structure was designed with the help of these features. The individual and combined performance of these features were evaluated. The results showed the accuracy and reliability of the proposed model. The objective of this work was to perform unbiased research to develop an outperforming model for the user friendly, and early diagnosis of diabetes and related diseases like DKD.

The clinical finding of this research work can also perform as the routine clinical examination by the medical practitioners. The model based on data mining techniques saves time, efforts, and invasive blood tests. The results obtained in this research work are very promising and facilitate the medical community as a primary opinion in decision making before recommending conventional, invasive blood tests.

7.2 Future work

Although the presented study has a significant contribution to the medical domain for the disease diagnosis, still there are possible extensions that can be undertaken as possible future work. Few of the future work are listed below:

1. The data-set thus should be strengthened to cover the wide possibilities of other chronic diseases also and the classification algorithm can also be more explored to achieve better results. To improve the diagnosis results further modern deep learning algorithms can also be applied, as an extension to this research work.

2. The proposed diagnostic model is tested on diabetes, duration of diabetes state and DKD disease only however, the study can be explored to more diseases according to the different ROI.

3. The pigmentation of the iris also depends on the genetics and other geographical conditions too. Therefore, the study can be extended for the population of different areas and countries. Type 1 and type 2 diabetes also depend on genetics therefore, the data can also be collected genetics-wise to explore the work for type 1 diabetes also.

4. Disease diagnosis experiment can be tested on the iris images acquired in visible light source also.The results with the visible light will encourage the researchers to develop a mobile-based application for diagnosis purposes.

5. A portable hand-held device can be developed so that it can be used in remote locations as a non-invasive tool for the screening of chronic diseases.

.

List of Publications

SCI/SCIE Journals:

- Piyush Samant and Ravinder Agarwal, "Machine learning techniques for medical diagnosis of diabetes using iris images", *Computer Methods and Programs in Biomedicine*, vol. 157, pp. 121–128, 2018.
  Publisher: Elsevier, Category: SCIE, Impact factor: 3.424

- Piyush Samant and Ravinder Agarwal, "Analysis of computational techniques for diabetes diagnosis using the combination of iris-based features and physiological parameters", *Neural Computing and Applications*, vol. 31, no. 12, pp. 8441-8453, 2019.
  Publisher: Springer, Impact factor: 4.664

- Piyush Samant and Ravinder Agarwal, "Assessment of Kidney problem in the diabetic patients using iris", *Iiranian Journal of Kidney Diseases,* (**Submitted**)
  Publisher: Iranian Society of Nephrology

- Piyush Samant and Ravinder Agarwal, "An advanced automatic region growing algorithm for sclera segmentation and its application in Biometric"*, Biocybernetics and Biomedical Engineering,* (**Submitted**)
  Publisher: Elsevier

Scopus Journals

- Piyush Samant and Ravinder Agarwal, "Comparative analysis of classification based algorithms for diabetes diagnosis using iris images", *Journal of Medical Engineering and Technology*, vol. 42, pp. 35-42, 2018.
  Publisher: Taylor and Francis

Conference Publication

- Piyush Samant, Ravinder Agarwal and Atul Bansal, "Enhanced Discrete Cosine Transformation feature based iris recognition using various scanning techniques", *4th IEEE Uttar Pradesh Section UPCON*, pp. 685-690, 2017. (Scopus Indexed).

- Piyush Samant and Ravinder Agarwal, "Diagnosis of diabetes by computer methods using iris", *4th National conference AdMeT*, February 2017,NPL, New-Delhi

Reference

[1] L. Chen, D. J. Magliano, and P. Z. Zimmet, “The Worldwide Epidemiology of Type 2 Diabetes Mellitus - present and future perspectives,” *Nat. Publ. Gr.*, vol. 8, no. 4, pp. 228–236, 2011.

[2] K. Ogurtsova, J. D. Rocha, Y. Huang, U. Linnenkamp, and L. Guariguata, “IDF Diabetes Atlas : Global estimates for the prevalence of diabetes for 2015 and 2040,” *Diabetes Res. Clin. Pract.*, vol. 128, pp. 40–50, 2017.

[3] L. J. Baier and R. L. Hanson, “Genetic Studies of the Etiology of Type 2 Diabetes in Pima Indians,” *Perspect. Diabetes*, vol. 53, no. 13, pp. 1181–1186, 2004.

[4] L. A. Levin, S. F. Nilsson, J. V Hoeve, and S. M. Wu, *ADLER’S Physiology of the Eye*, 11th ed. 2011.

[5] D. M. N.- Than, P. A. Cleary, and M. S. Backlund, “Intensive Diabetes Treatment and Cardiovascular Disease in Patients with Type 1 Diabetes,” *New englend J. Res.*, vol. 353, no. 25, pp. 2643–2653, 2005.

[6] V A FONSECA, “Defining and Characterizing the Progression of Type 2 Diabetes,” *Diabetes Care*, vol. 32, no. pp.151–156, 2009.

[7] U. Linnenkamp, L. Guariguata, J. Beagley, D. R. Whiting, and N. H. Cho, “The IDF Diabetes Atlas Methodology for Estimating Global Prevalence of Hyperglycaemia in Pregnancy,” *Diabetes Res. Clin. Pract.*, vol. 103, no. 2, pp. 186–196, 2013.

[8] 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.

[9] Bernard Jensen, “Bernard Jensen International,” 2011. [Online]. Available: http://www.bernardjensen.com/.

[10] B. Soediono, “Study of Eye: Iridology,” *J. Chem. Inf. Model.*, vol. 53, pp. 1-11, 1989.

[11] M. Sonka, H. Vaclav, and B. Roger, “Image Pre-processing,” in *Image Processing, Analysis and Machine Vision*, 1993, pp. 56–111.

[12] Z. He, T. Tan, Z. Sun, and X. Qiu, “Towards Accurate and Fast Iris Segmentation for iris Biometrics,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 9, pp. 1670–1684, 2008.

[13] F. Jan and I. Usman, “Iris Segmentation for Visible Wavelength and Near Infrared Eye Images,” *Opt. - Int. J. Light Electron Opt.*, 2014.

[14] S. Banerjee and D. Mery, “Iris Segmentation Using Geodesic Active Contours and GrabCut Iris Segmentation using Geodesic Active Contours and GrabCut,” in *Pacific-Rim Symposium on Image and Video Technology*, February 2016, pp. 1–12.

[15] U. R. Acharya, E. Y. K. Ng, and J. Tan, “Thermography Based Breast Cancer Detection Using Texture Features and Support Vector Machine,” *J. Med. Syst.*, vol. 36, pp. 1503–1510, 2010.

[16] P. Tschandl, C. Rosendahl, and H. Kittler, “Data Descriptor : The HAM 10000 dataset , a Large Collection of Multi-source Dermatoscopic Images of Common Pigmented Skin Lesions,” *Nat. Publ. Gr.*, vol. 5, pp. 1–9, 2018.

[17] 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.

[18] S. Kumar *et al.*, “Diabetes in India: A Long Way to Go,” *Int. J. Sci. Reports*, vol. 1, no. 1, pp. 92–98, 2011.

[19] L. Guariguata, D. R. Whiting, I. Hambleton, and J. Beagley, “Global Estimates of Diabetes Prevalence for 2013 and Projections for 2035,” *Diabetes Res. Clin. Pract.*, vol. 103, no. 2, pp. 137–149, 2013.

[20] N. C. D. Risk and F. Collaboration, “Worldwide Trends in Diabetes Since 1980 : a Pooled Analysis of 751 Population-based Studies with 4·4 Million Participants,” *Lancet*, vol. 387, pp. 1513–1530, 2008.

[21] R. Fernandes, K. Ogurtsova, and U. Linnenkamp, “IDF Diabetes Atlas Estimates of 2014 Global Health Expenditures on Diabetes,” *DIABETES Res. Clin. Pract.*, 2016.

[22] D. R. Whiting, L. Guariguata, C. Weil, and J. Shaw, “IDF Diabetes Atlas : Global Estimates of the Prevalence of Diabetes for 2011 and 2030,” *Diabetes Res. Clin. Pract.*, vol. 94, no. 3, pp. 311–321, 2011.

[23] J. E. Shaw, R. A. Sicree, and P. Z. Zimmet, “Global Estimates of the Prevalence of Diabetes for 2010 and 2030,” *Diabetes Res. Clin. Pract.*, vol. 87, pp. 4–14, 2010.

[24] L. Guariguata, D. Whiting, C. Weil, and N. Unwin, “The International Diabetes Federation Diabetes Atlas methodology for Estimating Global and National Prevalence of Diabetes in Adults,” *Diabetes Res. Clin. Pract.*, vol. 94, no. 3, pp. 322–332, 2011.

[25] N. Unwin, L. Guariguata, D. Whiting, and C. Weil, “Complementary Approaches to Estimation of the Global Burden,” *Lancet*, vol. 379, pp. 1487–1488, 2012.

[26] L. Guariguata, “By the Numbers : New Estimates from the IDF Diabetes Atlas Update for 2012,” *Diabetes Res. Clin. Pract.*, vol. 98, no. 3, pp. 522–525, 2012.

[27] C. Patterson, L. Guariguata, G. Dahlquist, G. Ogle, and M. Silink, “Diabetes in the young – A Global View and Worldwide Estimates of Numbers of Children with Type 1 Diabetes,” *Diabetes Res. Clin. Pract.*, vol. 1033, pp. 161–175, 2014.

[28] 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.

[29] L. Guariguata, U. Linnenkamp, and J. Beagley, “Global Estimates of the Prevalence of Hyperglycaemia in Pregnancy,” *Diabetes Res. Clin. Pract.*, vol. 103, no. 2, pp. 176–185, 2013.

[30] IDF Diabetes Atlas Group, “Update of Mortality Attributable to Diabetes for the IDF Diabetes Atlas : Estimates for the year 2011,” *Diabetes Res. Clin. Pract.*, vol. 103, pp. 277–279, 2013.

[31] I. D. F. Diabetes and A. Group, “Update of Mortality Attributable to Diabetes for the IDF Diabetes Atlas: Estimates for the year 2013,” *Diabetes Res. Clin. Pract.*, vol. 109, no. 3, pp. 461–474, 2015.

[32] N. Peer, A. Kengne, A. A. Motala, and J. Claude, “Diabetes in the Africa region : An update,” *Diabetes Res. Clin. Pract.*, vol. 103, no. 2, pp. 197–205, 2013.

[33] T. Tamayo *et al.*, “Diabetes in Europe: An update,” *Diabetes Res. Clin. Pract.*, vol. 103, no. 2, pp. 206–217, 2014.

[34] A. Majeed, A. A. El-sayed, T. Khoja, R. Alshamsan, C. Millett, and S. Rawaf, “Diabetes in the Middle-East and North Africa : An update,” *Diabetes Res. Clin. Pract.*, vol. 103, no. 2, pp. 218–222, 2013.

[35] S. F. Yisahak, J. Beagley, I. R. Hambleton, and K. M. V. Narayan, “Diabetes in North America and The Caribbean : An update,” *Diabetes Res. Clin. Pract.*, vol. 103, no. 2, pp. 223–230, 2013.

[36] P. Aschner *et al.*, “Diabetes in South and Central America: An update,” *Diabetes Res. Clin. Pract.*, vol. 103, no. 2, pp. 238–243, 2014.

[37] A. Ramachandran, C. Snehalatha, R. Ching, and W. Ma, “Diabetes in South-East Asia : An update,” *Diabetes Res. Clin. Pract.*, vol. 103, no. 2, pp. 231–237, 2013.

[38] J. C. N. Chan, N. H. Cho, N. Tajima, and J. Shaw, “Diabetes in the Western Pacific Region — Past , Present and Future,” *Diabetes Res. Clin. Pract.*, vol. 103, no. 2, pp. 244–255, 2013.

[39] A. Kumar, M. K. Goel, R. B. Jain, P. Khanna, and V. Chaudhary, “India towards diabetes control: Key issues,” *Australas. Med. J.*, vol. 6, no. 10, pp. 524–531, 2013.

[40] R. M. Anjana *et al.*, “The need for obtaining accurate nationwide estimates of diabetes prevalence in India - Rationale for a national study on diabetes,” *Indian J. Med. Res.*, vol. 133, no. 4, pp. 369–380, 2011.

[41] S. A. Kaveeshwar and J. Cornwall, “The current state of diabetes mellitus in India,” *Australas. Med. J.*, vol. 7, no. 1, pp. 45–48, 2014.

[42] H. Temurtas, N. Yumusak, and F. Temurtas, “A comparative study on diabetes disease diagnosis using neural networks,” *Expert Syst. Appl.*, vol. 36, no. 4, pp. 8610–8615, 2009.

[43] K. Polat and S. Güneş, “An expert system approach based on principal component analysis and adaptive neuro-fuzzy inference system to diagnosis of diabetes disease,” *Digit. Signal Process. A Rev. J.*, vol. 17, no. 4, pp. 702–710, 2007.

[44] K. Polat, S. Güneş, and A. Arslan, “A cascade learning system for classification of diabetes disease: Generalized Discriminant Analysis and Least Square Support Vector Machine,” *Expert Syst. Appl.*, vol. 34, no. 1, pp. 482–487, 2008.

[45] K. Kayaer and T. Yildirim, “Medical diagnosis on Pima Indian diabetes using general regression neural networks,” *Proc. Int. Conf. Artif. neural networks neural Inf. Process.*, pp. 181–184, 2003.

[46] A. K. Dwivedi, “Analysis of computational intelligence techniques for diabetes mellitus prediction,” *Neural Comput. Appl.*, pp. 1–9, 2017.

[47] M. Heydari, M. Teimouri, and Z. Heshmati, “Comparison of various classification algorithms in the diagnosis of type 2 diabetes in Iran,” *Int. J. Diabetes Dev. Ctries.*, vol. 36, no. 2, pp. 167–173, 2015.

[48] N. Liljequist, “The diagnosis from the eye, ,” 1916.

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

[50] Mjp. S. L. F. Salles, “Iridology: A systematic review,” *Rev. Da Esc. Enferm. Da USP*, vol. 42, no. 3, pp. 585–589, 2006.

[51] J.-Y. Um *et al.*, “Novel approach of molecular genetic understanding of iridology: relationship between iris constitution and angiotensin converting enzyme gene polymorphism.,” *Am. J. Chin. Med.*, vol. 33, no. 3, pp. 501–505, 2005.

[52] X. Qiu, Z. Sun, and T. Tan, “Global Texture Analysis of Iris Images for Ethnic Classification,” in *International Conference on Advances in Biometrics*, 2006, vol. 6, pp. 411–418.

[53] 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.

[54] 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.

[55] L. Ma, D. Zhang, N. Li, Y. Cai, W. Zuo, and K. Wang, “Iris-based medical analysis by geometric deformation features,” *IEEE J. Biomed. Heal. Informatics*, vol. 17, no. 1, pp. 223–231, 2013.

[56] R. Passarella and M. Fachrurrozi, “Development of iridology system database forcolon disorders identification using Image processing,” *Indian J. Bioinforma. Biotechnol.*, vol. 2, no. 6, pp. 100–103, 2013.

[57] 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.

[58] R. A. Ramlee, K. Azha, R. Singh, and S. Singh, “Detecting cholesterol presence with iris recognition algorithm,” in *Biometric Systems, Design and Applications*, 2011, pp. 129–148.

[59] J. Daugman, “How iris recognition works,” *IEEE teansactions circuits Syst. video Technol.*, vol. 14, no. 1, pp. 715–739, 2004.

[60] L. Masek, “Recognition of human iris patterns for biometric identification,” 2003.

[61] R. A. Ramlee and S. Ranjit, “Using iris recognition algorithm, detecting cholesterol presence,” in *International Conference on Information Management and Engineering, ICIME*, 2009, pp. 714–717.

[62] R. A. Ramlee, K. A. Aziz, M. E. S.Ranjit, S. Ranjit, M. Esro, and U. Teknikal, “Automated detecting arcus senilis , symptom for cholesterol presence using iris recognition algorithm,” *J. Telecommun. Electron. Comput. Eng.*, vol. 3, no. 2, pp. 29–39, 2011.

[63] 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.

[64] P. D. Lesmana, E. K. Purnama, and M. H. Purnomo, “Abnormal condition detection of pancreatic beta-cells as the cause of diabetes mellitus based on iris image,” in *International Conference on Instrumentation, Communication, Information Technology and Biomedical Engineering*, 2011, pp. 150–155.

[65] E. Arvacheh, “A study of segmentation and normalization for iris recognition systems,” 2006.

[66] J. 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.

[67] T. T. and Y. W. Zhu, Young, “Biometric personal identification system based on iris analysis,” 1994.

[68] H. Proenc, “Toward Noncooperative Iris Recognition : A Classification Approach Using Multiple Signatures,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 4, pp. 607–612, 2007.

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

[70] Wildes *et al.*, “Automated, Non-Invasive IRIS Recognition system and method (II),” 1998.

[71] Wildes*et al.*, “Automated, Non-Invasive IRIS Recognition system and method (Iss),” 1996.

[72] R. P. Wildes *et al.*, “A system for automated iris recognition,” *Proc. 1994 IEEE Work. Appl. Comput. Vis.*, pp. 121–128, 1994.

[73] L. Ma, Y. Wang, and T. Tan, “Iris Recognition Using Circular Symmetric Filters,” in *16th Int. Conf. Pattern Recognition*, 2002, pp. 4–7.

[74] L. Ma, T. Tan, Y. Wang, and D. Zhang, “Efficient Iris Recognition by Characterizing Key Local Variations,” *IEEE Trans. Image Process.*, vol. 13, no. 6, pp. 739–750, 2004.

[75] N. Abdul Jalil, R. Sahak, and A. Saparon, “Iris localization using colour segmentation and circular hough transform,” in *EEE-EMBS Conference on Biomedical Engineering and Sciences,* 2012, pp. 784–788.

[76] Z.-C. Li, J.-P. Qiao, B.-S. Li, and H.-L. Wan, “Non-ideal iris segmentation using anisotropic diffusion,” *IET Image Process.*, vol. 7, no. 2, pp. 111–120, 2013.

[77] N. I. Images, J. Zuo, S. Member, and N. A. Schmid, “On a Methodology for Robust Segmentation of nonideal iris images,” *IEEE Trans. Syst. man, Cybern.*, vol. 40, no. 3, pp. 703–718, 2010.

[78] A. F. Abate, M. Frucci, C. Galdi, and D. Riccio, “BIRD : Watershed Based IRis Detection for mobile devices ☆,” *Pattern Recognit. Lett.*, vol. 000, pp. 1–9, 2014.

[79] M. Frucci, M. Nappi, D. Riccio, and G. S. di Baja, “WIRE: Watershed based iris recognition,” *Pattern Recognit.*, vol. 52, pp. 148–159, 2015.

[80] J. Daugman, “New methods in iris recognition.,” *IEEE Trans. Syst. man, Cybern.*, vol. 37, no. 5, pp. 1167–1175, 2007.

[81] M. Vatsa, S. Member, R. Singh, S. Member, and A. Noore, “Improving Iris Recognition Performance Using Segmentation , Quality Enhancement , Match Score Fusion , and Indexing,” *IEEE Trans. Syst. man, Cybern.*, vol. 38, no. 4, pp. 1021–1035, 2008.

[82] S. M. Talebi, A. Ayatollahi, and S. M. S. Moosavi, “A Novel Iris Segmentation Method based on Balloon Active Contour,” in *6th Iranian Mach. Vis. Image Process*, 2016, pp. 1–5.

[83] K. Roy, P. Bhattacharya, and C. Y. Suen, “Unideal Iris Segmentation Using Region-Based Active Contour Model Kaushik,” in *7th Int. Conf. Image Anal. Recognit. (ICIAR)*, 2010, pp. 256–265.

[84] N. Susitha and R. Subban, “Reliable Pupil Detection and Iris Segmentation Algorithm based on SPS,” *Cogn. Syst. Res.*, vol. 57, pp. 78-84, 2019.

[85] R. Derakhshani and A. Ross, “A Texture-Based Neural Network Classifier for Biometric Identification using Ocular Surface Vasculature,” in *International Joint Conference on Neural Networks*, 2007, pp. 12–17.

[86] R. Derakhshani, A. Ross, and S. Crihalmeanu, “A new biometric modality based on conjunctival vasculature,” in *Proc. Artificial Neural Networks in Engineering*, 2006, pp. 1–8.

[87] S. Crihalmeanu, A. Ross, and R. Derakhshani, “Enhancement and registration schemes for matching conjunctival vasculature,” in *3rd IEEE Int. Conf. Biometrics (ICB), Alghero, Italy*, 2009, pp. 1240–1249.

[88] Z. Zhou, E. Y. Du, N. L. Thomas, and E. J. Delp, “A comprehensive approach for sclera image quality measure,” *Int. J. Biom.*, vol. 5, no. 2, pp. 181-198, 2013.

[89] K. Oh and K. Toh, “Extracting Sclera Features for Cancelable Identity Veri fi cation,” in *5th IAPR Int. Conf. Biometrics (ICB), New Delhi, India,*2012, pp. 245–250.

[90] M. H. Khosravi and R. Safabakhsh, “Human eye sclera detection and tracking using a modified time-adaptive self-organizing map,” *Pattern Recognit.*, vol. 41, no. 8, pp. 2571–2593, 2008.

[91] A. Das, U. Pal, M. A. F. Ballester, and M. Blumenstein, “Sclera recognition using dense-SIFT,” *13th Int. Conf. Intell. Syst. Des. Appl. ISDA*, pp. 74–79, 2014.

[92] Z. Zhou, S. M. Ieee, E. Y. Du, S. M. Ieee, N. L. Thomas, and S. M. Ieee, “A Comprehensive Sclera Image Quality Measure,” in *11th Int. Conf. Control, Automation, Robotics and Vision*, December 2010, pp. 7–10.

[93] A. Laddi *et al.*, “Non-invasive Jaundice Detection using Machine Vision Non - invasive Jaundice Detection using Machine Vision,” *IETE J. Res. ISSN*, vol. 59, no. 5, pp. 591–595, 2014.

[94] E. Friis-jensen, “Modeling and Simulation of Glucose-Insulin Metabolism,” *Congr. Lyngby*, 2007.

[95] S. Crihalmeanu and A. Ross, “Multispectral scleral patterns for ocular biometric recognition,” *Pattern Recognit. Lett.*, vol. 32, pp. 1860–1869, 2012.

[96] Z.-Y. Wu, M.-H. Wang, H.-M. Qi, M.-H. Wu, Y.-Z. Ge, and H.-T. Li, “Relationship between hOGG1 Ser326Cys gene polymorphism and coronary artery lesions in patients with diabetes mellitus.,” *Int. J. Clin. Exp. Med.*, vol. 8, no. 10, pp. 18629–18637, 2015.

[97] M. Ueyama *et al.*, “The impact of PNPLA3 and JAZF1 on hepatocellular carcinoma in non-viral hepatitis patients with type 2 diabetes mellitus,” *J. Gastroenterol.*, vol. 51, no. 4, pp. 370–379, 2016.

[98] F. Capone *et al.*, “The cytokinome profile in patients with hepatocellular carcinoma and type 2 diabetes,” *PLoS One*, vol. 10, no. 7, pp. 1–34, 2015.

[99] B. Jensen, “Iridology Chart.” [Online]. Available: http://www.bernardjensen.com/.

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

[101] L. F. Salles, M. Júlia, and E. A. C. De, “The prevalence of iridologic signs in individuals with Diabetes Mellitus *,” *Acta paul enferm*, vol. 21, no. 3, pp. 474–480, 2008.

[102] L. F. Salles and M. J. Silva, “The sign of the Cross of Andreas in the iris and Diabetes Mellitus: a longitudinal study,” *Rev Esc Enferm USP*, vol. 49, no. 4, pp. 626–631, 2015.

[103] U. M. Chaskar and M. S. Sutaone, “On a Methodology for Detecting Diabetic Presence from Iris Image Analysis,” in *International Conference on Power, Signals, Controls and Computation (EPSCICON)*, 2012, pp. 1–6.

[104] J. Daugman, “Iris recognition,” in *Tutorial, on International confrence on Biometrics*, 2012, pp. 1–118.

[105] W. Sankowski, K. Grabowski, M. Napieralska, M. Zubert, and A. Napieralski, “Reliable algorithm for iris segmentation in eye image,” *Image Vis. Comput.*, vol. 28, no. 2, pp. 231–237, 2010.

[106] P. E. Hart, “How the Hough transform was invented [DSP History],” *IEEE Signal Process. Mag.*, vol. 26, no. 6, pp. 18–22, 2009.

[107] C. Li, C. Xu, C. Gui, and M. D. Fox, “Level Set Evolution Without Re-initialization : A New Variational Formulation,” in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2005, pp. 430–436.

[108] A. Radman, N. Zainal, and K. Jumari, “Fast and reliable iris segmentation algorithm,” *IET Image Process.*, vol. 7, no. 1, pp. 42–49, 2013.

[109] J. Daugman, “The importance of being random: Statistical principles of iris recognition,” *Pattern Recognit.*, vol. 36, no. 2, pp. 279–291, 2003.

[110] Y. H. Li and M. Savvides, “An automatic iris occlusion estimation method based on high-dimensional density estimation,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 4, pp. 784–796, 2013.

[111] J. Daugman and C. Downing, “Epigenetic randomness, complexity and singularity of human iris patterns,” *Proc. Biol. Sci.*, vol. 268, pp. 1737–1740, 2001.

[112] S. Dey and D. Samanta, “Iris data indexing method using Gabor energy features,” *IEEE Trans. Inf. Forensics Secur.*, vol. 7, no. 4, pp. 1192–1203, 2012.

[113] A. Deshpande, S. Dubey, H. Shaligram, A. Potnis, and S. Chavan, “Iris Recognition System using Block Based Approach with DWT and DCT,” in *IEEE India Conference (INDICON)*, 2014, pp. 1–5.

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

[115] A. L. Boulesteix, S. Janitza, J. Kruppa, and I. Konig, “Overview of Random Forest Methodology and Practical Guidance with Emphasis on Computational Biology and Overview of Random Forest Methodology and Practical Guidance with Emphasis on

Computational Biology and Bioinformatics," *WIRE's Data Mining and Knowledge Discovery*, 2012.

[116] 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.

[117] K. Sivasankar, "FCM based iris image analysis for tissue imbalance stage identification," in *International Conference on Emerging Trends in Science, Engineering and Technology*, 2012, pp. 210–215.

[118] 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.

[119] D. H. Hareva *et al.*, "Implementation of iridology application on smartphone," in *7th ICTS,*2013, pp. 33–38.

[120] E. Bengtsson, H. Danielsen, D. Treanor, M. N. Gurcan, C. MacAulay, and B. Molnár, "Computer-aided diagnostics in digital pathology," *J. Int. socirty Adv. Cytom.*, vol. 91, no. 6, pp. 551–554, 2017.

[121] F. P. M. Oliveira and J. M. R. S. Tavares, "Medical image registration: A review," *Computer Methods in Biomechanics and Biomedical Engineering*, vol. 17, no. 2. Taylor & Francis, pp. 73–93, 2014.

[122] R. Kanawong, T. Obafemi-Ajayi, D. Liu, M. Zhang, and D. Xu, "Tongue Image Analysis and Its Mobile App Development for Health Diagnosis," in *Adv Exp Med Biol*, 2017, pp. 99–121.

[123] B. Pang, D. Zhang, and K. Wang, "Tongue image analysis for appendicitis diagnosis," *Inf. Sci. (Ny).*, vol. 175, no. 3, pp. 160–176, 2005.

[124] K. Goyal and R. Agarwal, "Pulse based sensor design for wrist pulse signal analysis and health diagnosis .," *Biomed. Res.*, vol. 28, no. 12, pp. 5187–5195, 2017.

[125] D. Dayal, "Non-invasive blood glucose monitoring is an elusive goose," *Int. J. Diabetes Dev. Ctries.*, vol. 36, pp. 399-400, 2016.

[126] T. S. Leung *et al.*, "Screening neonatal jaundice based on the sclera color of the eye using digital photography," *Biomed. Opt. Express*, vol. 6, no. 11, pp. 132–140, 2015.

[127] T. Xiong, Y. Qu, S. Cambier, and D. Mu, "The side effects of phototherapy for neonatal jaundice: What do we know? What should we do?," *Eur. J. Pediatr.*, vol. 170, no. 10, pp. 1247–1255, 2011.

[128] F. Sharan, *Iridology: A complete guide to diagnosing through the iris and to related forms of treatment*. 1992.

[129] K. W. Bowyer, K. Hollingsworth, and P. J. Flynn, “Image understanding for iris biometrics: A survey,” *Comput. Vis. Image Underst.*, vol. 110, no. 2, pp. 281–307, 2008.

[130] K. Hollingsworth, K. W. Bowyer, and P. J. Flynn, “Pupil dilation degrades iris biometric performance,” *Comput. Vis. Image Underst.*, vol. 113, no. 1, pp. 150–157, 2009.

[131] T. J. Buchanan, C. J. Sutherland, R. J. Strettle, T. J. Terrell, and A. Pewsey, “An investigation of the relationship between anatomical features in the iris and systemic disease, with reference to iridology,” *Complement. Ther. Med.*, vol. 4, no. 2, pp. 98–102, Apr. 1996.

[132] Z. Othman and A. Satria Prabuwono, “Preliminary study on iris recognition system: Tissues of body organs in iridology,” *Proc. 2010 IEEE EMBS Conf. Biomed. Eng. Sci. IECBES 2010*, no. December, pp. 115–119, 2010.

[133] 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.

[134] P. S. K. Bhatia, P. Atole, S. Kamble, and P. Telang, “Methodology for detecting diabeticpresence from iris image analysis,” *Int. J. Adv. Res. Comput. Eng. Technol.*, vol. 4, no. 3, pp. 776–779, 2015.

[135] N.D.Pergad and S. B. More, “Detection of diabeticpresence from iris by using support vector machine,” *Int. J. Eng. Sci. Res.*, vol. 4, no. 7, pp. 562–565, 2015.

[136] K. Zahirnia, M. Teimouri, R. Rahmani, and A. Salaq, “Diagnosis of Type 2 Diabetes Using Cost-Sensitive Learning,” *Int. Conf. Comput. Knowl. Eng. Diagnosis, IEEE*, pp. 58–63, 2015.

[137] M. Sudha, “Evolutionary and Neural Computing Based Decision Support System for Disease Diagnosis from Clinical Data Sets in Medical Practice,” *J. Med. Syst.*, vol. 41, no. 11, 2017.

[138] B. A. Tama and K. H. Rhee, “Tree-based classifier ensembles for early detection method of diabetes: an exploratory study,” *Artif. Intell. Rev.*, no. 48513, pp. 1–16, 2017.

[139] A. K. Dwivedi, “Performance evaluation of different machine learning techniques for prediction of heart disease,” *Neural Comput. Appl.*, pp. 1–9, 2016.

[140] W. Cannan and M. Eastwood, “Iridology: Diagnostic Validity in Orthopedic Trauma,” *Sci. Rev. Altern. Med.*, vol. 6, no. 2, pp. 63–67, 2002.

[141] T. Ibrikci, D. Ustun, and I. E. Kaya, “Diagnosis of Several Diseases by Using Combined Kernels with Support Vector Machine,” pp. 1831–1840, 2012.

[142] L. Frank, “The reliability of iridology in the diagnosis of previous acute appendicitis , as evidenced by appendectomy A dissertation submitted to the Faculty of Health Sciences , University of Johannesburg , as partial fulfilment of the requirement for the Degree :,” 2012.

[143] X. H. Meng, Y. X. Huang, D. P. Rao, Q. Zhang, and Q. Liu, “Comparison of three data mining models for predicting diabetes or prediabetes by risk factors,” *Kaohsiung J. Med. Sci.*, vol. 29, no. 2, pp. 93–99, 2013.

[144] N. Kaur and M. Juneja, “A review on Iris Recognition,” in *Recent Advances in Engineering and Computational Sciences, RAECS 2014*, 2014, pp. 6–8.

[145] Y. Alvarez-betancourt, and M. Garcia-silvente, “A keypoints-based feature extraction method for iris recognition under variable image quality conditions,” *Knowledge-Based Syst.*, vol. 92, pp. 169–182, 2016.

[146] D. J. Pesek, *HOLISTIC IRIDOLOGY – AN OVERVIEW*. 2013.

[147] N. Zhou and L. Wang, “A Modified T-test feature selection method and its application on the hapmap genotype data,” *Genomics, Proteomics Bioinforma.*, vol. 5, no. 3–4, pp. 242–249, 2007.

[148] S. Karamizadeh, S. M. Abdullah, A. A. Manaf, M. Zamani, and A. Hooman, “An overview of principal component analysis,” *J. Signal Inf. Process.*, vol. 04, no. 03, pp. 173–175, 2013.

[149] 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.

[150] G. V Kapoula, P. I. Kontou, and P. G. Bagos, “Diagnostic Accuracy of Neutrophil Gelatinase- Associated Lipocalin for Predicting Early Diabetic Nephropathy in Patients with Type 1 and Type 2 Diabetes Mellitus : A Systematic Review and Meta-analysis,” *J. Appl. Lab. Med.*, vol. 4, pp. 78–94, 2019.

[151] P. Samant and R. Agarwal, “Comparative analysis of classification based algorithms for diabetes diagnosis using iris images,” *J. Med. Eng. Technol.*, vol. 42, no. 1, pp. 35–42, 2018.

[152] C. M. Gold, “Iridology diabetes mellitus Compiled,” 2008.

[153] A. Bansal, “Iris Recognition System for Some Clinical Applications,” 2015.

[154] R. G. A. Nusantara, P. Herlambang, R. R. Isnanto, and A. A. Z, “Application of Liver Disease Detection Using Iridology with Back-Propagation Neural Network,” in *Int. Conference on Information Technology, Computer and Electrical Engineering (ICITACEE)*, 2015, pp. 123–127.

[155] A. Simon, D. M. Worthen, L. John, and A. M. Ii, “An Evaluation of Iridology,” *J. Amer-ican Med. Assoc.*, vol. 242, pp. 1385–1387, 2015.

[156] M. D. Sulistiyo, R. N. Dayawati, and P. M. Pahirawan, “Iridology-based dyspepsia early detection using linear discriminant analysis and Cascade Correlation Neural Network,” *2014 2nd Int. Conf. Inf. Commun. Technol.*, pp. 139–144, 2014.

[157] Paul Tower, “The Fundus Oculi in Monozygotic Twins: Report of Six Pairs of Identical Twins,” *Arch. Ophthalmol.*, vol. 54, pp. 225–239, 1955.

[158] Z. Zhou *et al.*, “Quality Fusion Based Multimodal Eye Recognition,” in *IEEE International Conference on Systems, Man, and Cybernetics*, 2012, pp. 1297–1302.

[159] S. P. Tankasala, P. Doynov, R. Derakhshani, A. Ross, and S. Crihalmeanu, “Biometric Recognition of Conjunctival Vasculature using GLCM Features,” in *International Conference on Image Information Processing*, 2011, pp. 1-6.ss

[160] H. Proenc and A. Alexandre, “UBIRIS : A Noisy Iris Image Database,” in *13th International Conference on Image Analysis and Processing - ICIAP*, 2005, pp. 970–977.

[161] S. Innovation, “Local Gabor Wavelet-Based Feature Extraction and Evaluation,” in *Smart Innovation, Systems and Technologies*, vol. 43, pp. 181–189, 2015.

[162] E. Ernst, “Iridology Not Useful and Potentially Harmful, ” *Arch. Ophthalmol.*, vol. 118, pp. 120-121, 2000.

[163] D. M. Cockburn, “A Study of the Validity of Iris Diagnosis, ” *The Australian Journal of Optometry*, vol. 64, pp. 154-157, 1981.

[164] P. Knipschild, “Looking for Gall Bladder Disease in the Patient's Iris, ” *BMJ*, vol. 297, pp. 1578-1581, 1988.