Non-Invasive Spectroscopic Characterization of Blood Glucose Level using Machine Learning Approach

By

Ali Newaz Bahar

Mohammad Habibullah

Mohammad Abdul Moin Oninda

University of Saskatchewan

57 Campus Dr, Saskatoon, SK S7N 5A9, Canada

6th December, 2018

Declaration

This is to certify that the project work entitled **"Non-Invasive Spectroscopic Characterization of Blood Glucose Level using Machine Learning Approach"** has been carried out by *Ali Newaz Bahar*, *Mohammad Habibullah* and *Mohammad Abdul Moin Oninda* in the Department of Electrical and Computer Engineering, School of Engineering, University of Saskatchewan, Saskatoon, SK, Canada. The above project work or any part of this project work has not been submitted any where for the award of any degree or diploma.

Signature of Candidate	Signature of Candidate	Signature of Candidate
Ali Newaz Bahar	Mohammad Habibullah	Mohammad Abdul Moin Oninda
PhD Student	M.Sc. Student	M.Sc. Student

Acknowledgments

All the praise and appreciation go to the almighty God for giving us patience in work, ability and intelligent to complete, as well as, to submit our project report successfully without any major problem. We would like to express our sincere appreciation to our course instructor Dr. Anh Dinh, Professor, Department of Electrical and Computer Engineering, University of Saskatchewan, for giving me his valuable guidance and insight, simulating suggestion and advises, continuous encouragement and support and reliance over the time to produce this report within a very short time. His guidance, punctuality, discipline and simplicity have been indelible memory for us.

Finally, we would like to thank all other honorable teachers and friends who have helped, inspired and given us mental support at different stages during completion of our project work.

Ali Newaz Bahar
Mohammad Habibullah
Mohammad Abdul Moin Oninda

December 2018

Abstract

Diabetes Mellitus (DM) commonly referred to as diabetes belongs to a group of metabolic disorders which exhibits high blood sugar levels over a prolonged period of time. It can be controlled by appropriate regimen: diet therapy, weight reduction, exercises and insulin injection or oral drugs to lower blood glucose. In the management of diabetes, glucose monitoring technology has been used for decades. Current the most widely used self-monitoring method involves the *finger pricking* approach, an invasive method of measuring blood glucose causing pain and discomfort to patients. Researchers have been working various approaches of developing a non-invasive glucose monitoring devices, which will drastically improve the quality of life for people suffering from diabetes and facilitate their compliance for glucose monitoring. This project emphasizes on a non-invasive spectroscopic characterization and prediction of blood glucose level using machine learning approach. A brief idea about diabetes mellitus along with several measurement techniques have also been discussed. $NeoSpectraMicto^{TM}$ development kit and platform was used for data acquisition.

Keywords: Diabetes Mellitus (DM), Non-invasive, NIR Spectroscopy$NeoSpectraMicto^{TM}$, Machine Learning, PCA, SVM, PLSR,

Preface

This course project is outlined based on the results obtained from the laboratory experiment. This is carried out in the Department of Electrical and Computer Engineering, School of Engineering, University of Saskatchewan, Saskatoon, SK, Canada.

This report includes four chapters which are briefed as follows:

Chapter-1

Chapter 1 provides a detailed discussion on the importance of the work that has been done and why the current topic is selected as our project. Also, it gives the information about our project motivation, aim and objectives, and challenges.

Chapter-2

Chapter 2 discusses the technical background related to the non-invasive blood glucose monitoring.

Chapter-3

Chapter 3 discusses about the data collection procedure and data analysis.

Chapter-4

Chapter 4 provides the summary discussions based on the results.

Contents

List of Figures

List of Tables

List of Abbreviations

NIDDM	Non-Insulin-Dependent Diabetes Mellitus
CGM	Continuous Glucose Monitoring
SNR	Signal-to-Noise Ratio
NIR	Near-Infrared
MEMS	Micro Electro Mechanical Systems
IoT	Internet of Things
GUI	Graphical User Interface
API	Application Program Interface
ML	Machine Learning
PLSR	Partial Least Squares Regression
SVM	Support Vector Machine
PCA	Principal Component Analysis
MSE	Mean Squared Error

Chapter 1

Introduction

One of the principle healthcare epidemics of the modern era is Diabetes. The total number of people with diabetes worldwide is expected to rise from 171 million in 2000 to 366 million in 2030 [1]. Diabetes was the sixth leading cause of death listed on US death certificates in 2002 with an estimates cost of diabetes in US of $132 billion consisting of both direct and indirect expenses (disability, work loss, premature mortality)[2].

Diabetes mellitus is an incurable disease [3] resulting from an insufficiency of insulin in the body [4], causing elevated blood-glucose levels, known as hyperglycemia, or reduces glucose concentration, known as hypoglycemia [5,6]. Insulin is a hormone which is secreted from the pancreas to handle metabolic reactions involving glucose [7,8]. It initiates glucose uptake by cell types in the body, hence reducing glucose concentrations in blood [4, 9]. Diabetes can lead to several medical conditions such as coeliac disease, cystic fibrosis, tuberculosis and heart disease. These can lead to blindness, renal failure, peripheral nerve damage, amputation, cardiovascular disease or even cancer [4, 6, 8]. There are three main types of diabetes:

- **Type 1 Diabetes Mellitus:** Occurs then the pancreas fails to produce enough insulin due to loss of beta cells also referred to as *Insulin-Dependent Diabetes Mellitus (IDDM)* or *Juvenile Diabetes.*

- **Type 2 Diabetes Mellitus:** This starts with insulin resistance, a condition in which the cell does not respond to insulin properly. A lack of insulin may also develop as the disease is progressive. This is also referred to as *Non-Insulin-Dependent Diabetes Mellitus (NIDDM)* or *Adult-onset Diabetes.*

- **Gestational Diabetes:** It occurs in pregnant women without previous history of diabetes who develop high blood sugar levels.

1.1 Motivation

Glucose monitoring represents an exciting frontier in diabetes research, holding the potential to improve the lives of over 400 million people worldwide which is expected to rise by approximately 55% within the next 25 years [10]. The conventional blood glucose testing and monitoring system (Finger pricking method) is painful, inconvenient and introduces discomfort in patient's life. Figure 1 show several methods of blood glucose monitoring system. Researchers have been working on to develop a blood glucose monitoring system having the following qualities [11]:

- Non-Invasive
- Non-contact
- Fast measurement capability
- Painless measurement
- Convenience for continuous real-time monitoring
- Cost effective
- Adequate control
- Reduction in complication from usage

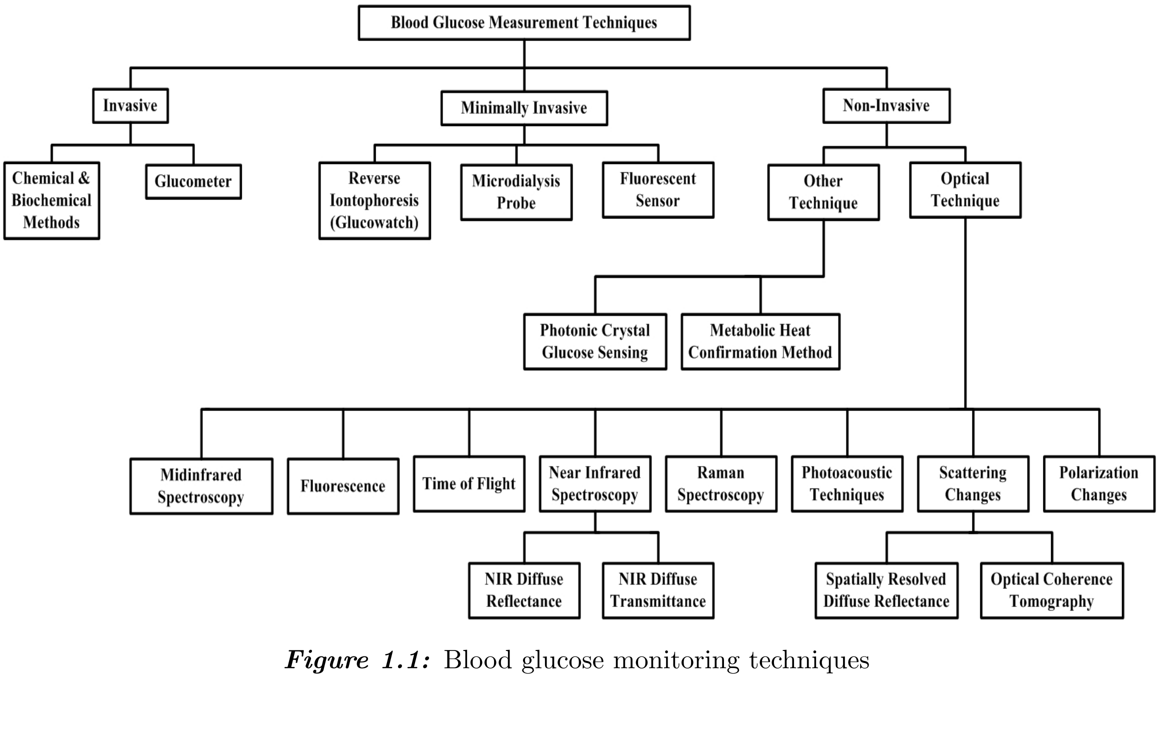

Blood Glucose Measurement Techniques
Invasive
Minimally Invasive
Non-Invasive
Chemical & Biochemical Methods
Glucometer
Reverse Iontophoresis (Glucowatch)
Microdialysis Probe
Fluorescent Sensor
Other Technique
Optical Technique
Photonic Crystal Glucose Sensing
Metabolic Heat Confirmation Method
Midinfrared Spectroscopy
Fluorescence
Time of Flight
Near Infrared Spectroscopy
Raman Spectroscopy
Photoacoustic Techniques
Scattering Changes
Polarization Changes
NIR Diffuse Reflectance
NIR Diffuse Transmittance
Spatially Resolved Diffuse Reflectance
Optical Coherence Tomography

Figure 1.1: Blood glucose monitoring techniques

Continuous Glucose Monitoring (CGM) provides additional temporal information such as trends, magnitude, duration, and frequency of glucose level fluctuations which can aid in the identification and prevention of unwanted hypo and hyperglycemic episodes, activate alarm signals for extreme glucose levels and facilitate in automatic feedback-controlled insulin delivery systems such as artificial pancreas [12]. Although blood remains the most studied body fluid for glucose measurement, other more accessible biological fluids such as intestinal fluid, ocular fluid, sweat, saliva, breath or urine have been investigates as alternative sample media for non-invasive continuous monitoring [3] as shown in Table 1. Fig 1 shows a summary of widely used monitoring techniques of blood glucose concentration.

Table 1.1: Mean blood glucose concentration

Level	**mg/dL**	**mmol/L**	**Risk**
Dangerously High	315+	17.4	Very High
High	280	15.6	High
High	250	13.7	High
High	215	11.0	High
Borderline	180	10.0	Medium
Borderline	150	8.2	Medium
Borderline	120	7	Medium
Normal	108	6	No risk
Normal	72	4	No risk
Low	70	3.9	Medium
Dangerously Low	50	2.8	High

1.2 Aims and Objectives

The basic component of QCA circuits is majority voter gates. Hence, efficient design of QCA circuits using majority gates has attracted a considerable attention. Since majority gates do not function as a universal gate, majority logic can be used to implement any logical function instead of Boolean logic operators.

1.3 Challenges

1.3.1 Non-invasive glucose monitoring

The major difficulties in the development of non-invasive glucose sensors are associated with the indirect nature of the measurement and the inevitable calibration process. This may respectively result in reduced accuracy, low usability and diminished applicability for home use that require much effort to overcome [10].

1.3.2 Accuracy related challenges

Poor glucose specificity and sensitivity

The indirect nature of non-invasive approaches subjects them to suffer from a relatively low signal-to-noise ratio SNR, since the measured parameters may be affected both from physiological factors other than glucose as well as from external elements. Most non-invasive devices rely on optical technologies which suffer from low sensitivity and specificity to glucose, due to very low signal produced by glucose molecules [13].

Physiological time lag: The physiological time lag between blood and tissue glucose decreases the accuracy of indirect glucose monitoring. As non-invasive technologies are based on indirect estimations of glucose levels, a time lag may occur between measurements of blood glucose content from different parts of the body. The lag time of glucose dynamics varies depending on the properties of skin layer, therefore estimating glucose from the whole tissue subjects the accuracy to suffer from a time lag between the glucose concentration in blood and *ISF* of all skin layers, *i.e.* the epidermis, dermis and subcutaneous layers.

1.3.3 Usability and applicability challenges

Device calibration: The indirect nature of non-invasive measurement requires calibration against concurrent blood glucose values, which provides an estimation of the glucose concentration. The calibration process is conducted prior to using the device, in order to minimize the impact of individual quasi stable factors, such as tissue thickness and structure. Typically, the process consists of several paired invasive-non-invasive measurements in a varying frequency, depending on the device and the technology employed. Another desired goal in non-invasive glucose sensing iis reducing the calibration frequency and even neglecting it completely.

Suitability for various people: In order to reach high efficacy, non-invasive devices should be suitable for a variety of users. This is challenging since most of the technologies used to indirectly estimate glucose suffer from interfering human factors such as skin characteristics [14]. This is indeed a major limitation of optical techniques, since the transmission of light at each wavelength is a function of thickness, color and structure of the skin, bone, blood and other material through which the light passes [15].

1.3.4 Experimental challenges

While conducting the experiment for developing a method of predicting glucose level in blood non-invasively several challenges needed to be overcome:

- Obtaining a database of glucose levels in order to train and test the selected algorithm.
- Obtaining diabetic patients having a wide range of glucose concentration in blood.
- Making up an experimental setup in order to obtain data accurately.
- Making blood and blood + glucose solutions for testing the proposed device as it would take a lot of time to get approval for testing on live samples or patients.
- Selecting the specific part of the human body from which the data will be collected and tested.

1.3.5 Limitations

The absorption coefficient of glucose in the NIR band is low and is much smaller than that of water by virtue of the large disparity in their respective concentrations. Thus, in the NIR the weak glucose spectral bands only overlap with the stronger bands of water, but also of hemoglobin, proteins and fats. As regards the scattering coefficient, the effect of a solute (like glucose) on the refractive index of a medium is non-specific, and hence it is common to other soluble analytes. Furthermore, physical and chemical parameters, such as variation in blood pressure, body temperature, skin hydration, and triglyceride and albumin concentrations may interfere with glucose measurement [16]. Errors can also occur due to environmental variations, such as changes in temperature, humidity,carbon dioxide and atmospheric pressure. Changes in glucose by themselves can introduce other confounding factors: for instance, it has been proved that hyperglycemia, as well as hyper *insulinemia* (often connected to the former in obese patients) can induce *vasodilatation*,which results in increased perfusion [17, 18]. This phenomenon increases light absorption, and hence it can lead to errors in the estimation of the blood glucose concentration if not taken into account. It has also been reported that hyperglycemia can have effects on skin structural properties. Diabetic subjects in fact can exhibit *thickskin* and *yellowskin*, probably due to accelerated collagen aging and elastic fiber fraying [19,20]. Thus, light reflected from the skin of diabetic subjects may have different intensity than in healthy subjects at equal level of glycaemia. Also thermal properties of the skin were found to be different in subjects with hyperglycemia, thus affecting the localized reflectance of light [21]. Hyperglycemia can also cause differences in the refractive index of red blood cells, thus leading to different light scattering [22, 23]. Another confounding factor is due to the fact that NIR measurements often reflect glucose concentration indifferent body compartments, that is, not only blood but also the interstitial fluid in different body tissues can contribute to the measured signal.

Chapter 2

Technical Background

We utilizes the concept of near-infrared spectroscopy for the detection of blood glucose concentration in the range of *1300 nm–2300 nm* range.This chapter provides an overview of the *Near Infrared Spectroscopy (NIR)*, $NeoSpectraMicro^{TM}$ development kit which is essential for understanding the project work. Various machine learning algorithms have been employed to build the model. We also discussed the different concentration glucose and blood sample preparation procedure.

2.1 Methodology

The aim of this project is to develop a device that can non-invasively predict the blood glucose concentration using machine learning approach. The project can be briefly described using the Function (F), Context (C), Behavior (B), Principle (P), State (S) and Structure (S) FCBPSS design methodology. The function refers to the purpose of designing the object or a system. The function of the project is to non-invasively detect blood glucose concentration in human body. The effect of the environment as well as the pre-condition and post-condition related to or surrounding the structure can be referred to as the context. Behavior refers to the attributes from the object structure. Spectroscopic characterization of the the blood glucose concentration over a wide range of wavelength and the data obtained from the spectrum which was used to predict the blood glucose concentration using machine learning algorithm can be term as the behavior of the device. The principle of the system is the theory based on which the behavior of the system can be explained – in this case Near-Infrared Spectroscopy.States can be of either physical or chemical form. State variables are used to express the dynamics of the system and can be either dependent or independent state variables. The independent state variables (skin texture, thickness etc.) affects the system from outside whereas the dependent state variables (haemoglobin and other contents of the blood other than glucose) affect the system from the inside.Structure defines the components needed to design the object and how they are linked together. The overall project was divided into six phases as shown in Figure 2.1.

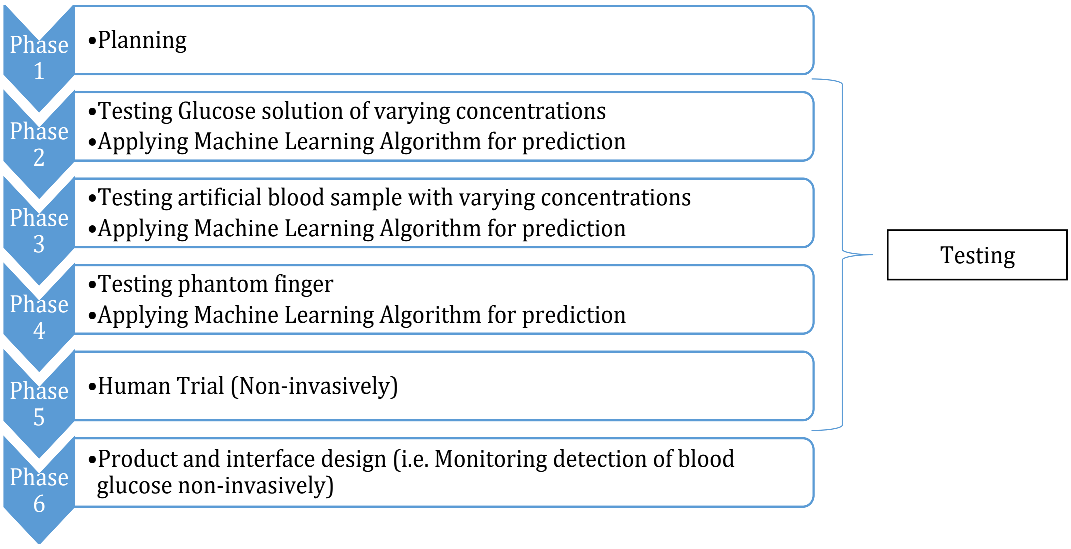

Phase 1
•Planning
Phase 2
•Testing Glucose solution of varying concentrations
•Applying Machine Learning Algorithm for prediction
Phase 3
•Testing artificial blood sample with varying concentrations
•Applying Machine Learning Algorithm for prediction
Phase 4
•Testing phantom finger
•Applying Machine Learning Algorithm for prediction
Phase 5
•Human Trial (Non-invasively)
Phase 6
•Product and interface design (i.e. Monitoring detection of blood glucose non-invasively)
Testing

Figure 2.1: Project development phases

Phase 1 of the project involved research of various techniques of monitoring blood glucose concentration and selecting suitable technique and algorithm for proper prediction of blood glucose concentration for the project. This project utilizes Near-Infrared Spectroscopy (NIR) from $1300nm - 2300nm$ wavelength for characterization of glucose concentration in blood using the $NeoSpectraMicro^{TM}$ Development kit and platform as the sensing device.

2.1.1 Near Infrared Spectroscopy (NIR)

Near infrared (NIR) spectroscopy is based on focusing on the body a beam of light in the 750–2500 *nm* spectrum [24]. NIR spectroscopy allows glucose measurement in tissues in the range of *1–100 mm* of depths, with a decrease in penetration depth for increasing wavelength values. The light focused on the body is partially absorbed and scattered, due to its interaction with the chemical components within the tissue. Attenuation of light in tissue is described,according to light transport theory, by the equation [25].

$$I = I_0 e^{-\mu_{eff} d} \tag{2.1}$$

Where, I is the reflected light intensity

I_0 is the incident light intensity

μ_{eff} is the effective attenuation coefficient

d is the optical path length in tissue.

On the other hand, μ_{eff} can be expressed as a function,

$$\mu_{eff} = f(\mu_a, \mu_s) \tag{2.2}$$

Where, μ_a is the absorption coefficient

μ_b is the scattering coefficient.

Changes in glucose concentration can influence μ_a of a tissue through changes of absorption corresponding to water displacement or changes in its intrinsic absorption. Changes in glucose concentration also affect the intensity of light scattered by the tissue, i.e. μ_s.This coefficient is a function of the density of scattering centers in the tissue observation volume, the mean diameter of scattering centers, their refractive index and the refractive index of the surrounding fluid. For the case of cutaneous tissue,connective tissue fibers are the scattering centers. Erythrocytes are the scattering centers for blood. In summary, glucose concentration could be estimated by variations of light intensity both transmitted through a glucose containing tissue and reflected by the tissue itself. Transmission or reflectance (localized or diffuse)of the light can be measured by proper detectors.

2.1.2 NeoSpectraMicro^{TM} Development kit

The NeoSpectra Micro, shown in Figure 2.2 is a chip-sized, Near-InfraRed (NIR) spectral sensor that delivers the spectral response of the light absorbed by materials for quantification, qualification and identification. It is designed to be used in different systems such as OEM module for applications that can be enabled by the spectral range *1350 – 2500 nm*. NeoSpectra Micro's core technology is based on semiconductor Micro Electro Mechanical Systems (MEMS) micro fabrication techniques, Figure 2.3 shows the block diagram of NeoSpectra Micro, promising unprecedented economies of scale.

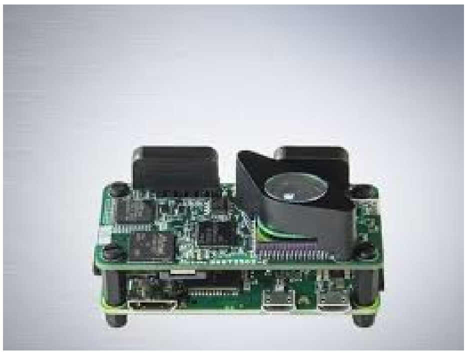


Figure 2.2: $NeoSpectraMicro^TM$ development kit

The NeoSpectra Micro sensor determines the spectral content of the input light in NIR range between *1350 – 2500 nm.* NeoSpectra Micro's unique feature enable the creation of new usage models ranging from IoT sensors, to deployment in handheld devices in various application areas, including: food analysis, agriculture, pharmaceutical, oil and gas, polymers, healthcare, industrial and chemicals analysis. Features:

- Low cost embedded NIR spectral sensor solution
- Smallest FT-IR solution with a single photo detector
- Wide spectral range at the higher wavelengths end of NIR (*1350 – 2500 nm*)
- Designed for high volume production with economies of scale
- Fast, on-chip, data processing
- Alignment free optics
- Low power consumption

NeoSpectra module comes with a suite of software editions (SpectroMost) which provides an easy to use GUI that can be utilized to perform measurements, as well as a set of APIs that can be utilized to build user defines software. SpectroMost as shown in Figure 2.4, is the software GUI that will be used to evaluate NeoSpectra Modules. It enables plotting, saving, and loading measured spectra, as well as setting parameters for the NeoSpectra modules. The device works on the principle of plug and play so all that is required is a computer with windows or Linux operating system.

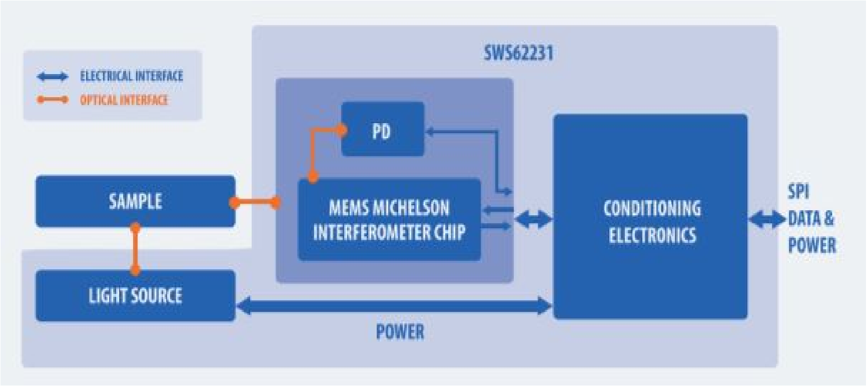
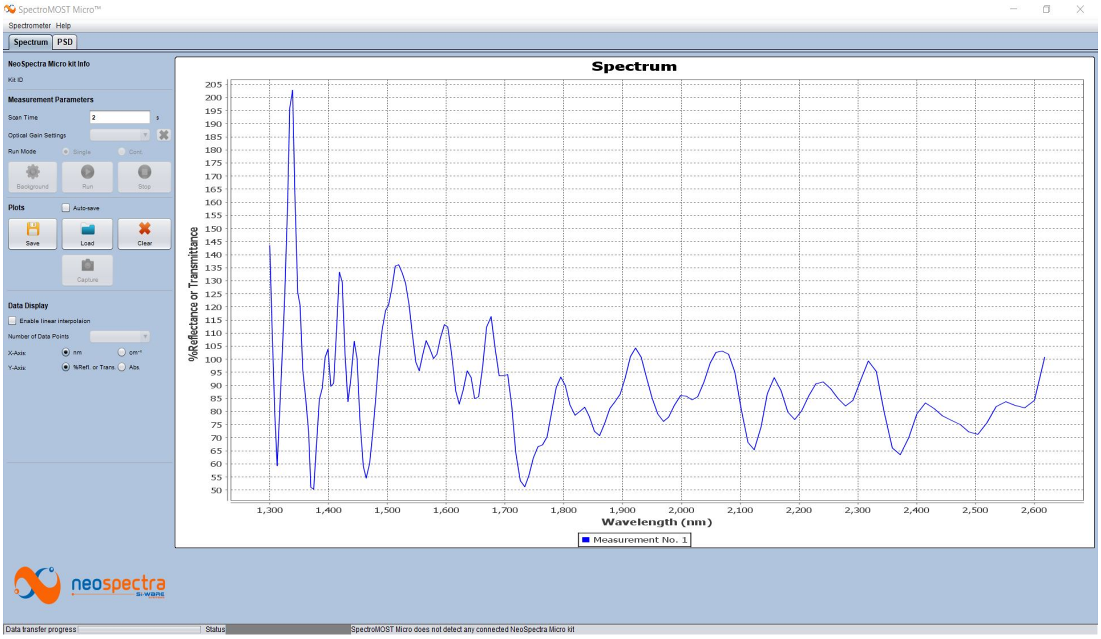


Figure 2.3: Block diagram of $NeoSpectraMicro^{TM}$ development kit



Kit ID

Scan Time 2 s

Optical Gain Settings

Run Mode ○ Single ○ Cont.

□ Auto-save

□ Enable linear interpolaion

Number of Data Points

X-Axis: ● nm ○ cm^{-1}

Y-Axis: ● %Refl. or Trans. ○ Abs.

Data transfer progress Status SpectroMOST Micro does not detect any connected NeoSpectra Micro kit

Figure 2.4: Software GUI for NeoSpectra Modules (SpectroMost)

2.1.3 Machine Learning (ML)

Machine learning (ML) is the study of algorithms and mathematical models that computer systems use to progressively improve their performance on a specific task. Machine learning algorithms build a mathematical model of sample data, known as ”training data”, in order to make predictions or decisions without being explicitly programmed to perform the task [26]. Machine learning algorithms are often categorized as supervised or unsupervised.

Supervised machine learning algorithms can apply what has been learned in the past to new data using labeled examples to predict future events. Starting from the analysis of a known training dataset, the learning algorithm produces an inferred function to make predictions about the output values. The system is able to provide targets for any new input after sufficient training. The learning algorithm can also compare its output with the correct, intended output and find errors in order to modify the model accordingly.

In contrast, **unsupervised machine learning** algorithms are used when the information used to train is neither classified nor labeled. Unsupervised learning studies how systems can infer a function to describe a hidden structure from unlabeled data. The system doesn't figure out the right output, but it explores the data and can draw inferences from datasets to describe hidden structures from unlabeled data.

Semi-supervised machine learning algorithms fall somewhere in between supervised and unsupervised learning, since they use both labeled and unlabeled data for training – typically a small amount of labeled data and a large amount of unlabeled data. The systems that use this method are able to considerably improve learning accuracy. Usually, semi-supervised learning is chosen when the acquired labeled data requires skilled and relevant resources in order to train it learn from it. Otherwise, acquiringun labeled data generally doesn't require additional resources.

Reinforcement machine learning algorithms is a learning method that interacts with its environment by producing actions and discovers errors or rewards. Trial and error search and delayed reward are the most relevant characteristics of reinforcement learning. This method allows machines and software agents to automatically determine the ideal behavior within a specific context in order to maximize its performance. Simple reward feedback is required for the agent to learn which action is best; this is known as the reinforcement signal. As the aim of the project is to estimate the blood glucose level from

NIR-spectrum data, so finding the best model that suits the data set with highest accuracy is the major part of the project. This part is discussed step by step below:

1. First, Partial least square regression model (PLSR) was used splitting the samples into 60% for training and 40% for validation. And optimum number of components for least MSE (mean square error) was calculated. But the accuracy and coefficient of determination

was not good.

2. As, PLSR was not able to predict accurately,Support Vector Machine (SVM), a supervised machine learning modeling, was applied on the NIR- data set by classifying them in ten classes each signifying defined glucose levels. The data set was split by 60% for training and 40% for testing. In the data set, total no of samples taken is 100 which includes ten samples for each class and total number of features was 156 which are the wavelengths from 1200 nm – 2400 nm.But the result was not up to the mark.

3. So, to overcoming high dimensionality of the model, Principal component analysis (PCA) was performed on the data set prior to using SVM modeling. As PCA is affected by scale, so scaling the features in the data was done before applying PCA.

4. After performing scaling and PCA modeling, linear kernel SVM classification was applied on the dataset by splitting 60% for training and 40% for testing. And the score, accuracy, precision was observed.

2.1.4 Measurement sites

NIR light transmission or reflectance has been studied through an ear lobe, finger web and finger cuticle, skin of the forearm, lip mucosa, oral mucosa,tongue, nasal septum, cheek and arm. NIR diffuse reflectance measurements performed on the finger showed a correlation with blood glucose but predictions were often not sufficiently accurate to be clinically acceptable [16]. Diffuse reflectance studies of the inner lip also showed good correlation with blood glucose and indicated a time lag of some minutes between blood glucose and the measurement signal [27]. Salivary glucose levels (a component of lip measurements) did not reflect blood glucose levels. These locations were selected for some advantages, such as high vascularization, little fatty tissue, homogeneous composition,limited temperature variations. It must be noted that,depending on the considered site, some specific intervals in the NIR band have been considered and the choice of one specific site also influenced the type of studied light, i.e. transmitted, or reflected (localized or diffuse), or both.

2.2 Experimental Setup

2.2.1 Glucose Solutions

Glucose solutions of varying concentration was prepared for Phase 2 testing according to the amount of glucose present in blood for diabetic and non-diabetic patients. Table 2.1 shows glucose concentration in blood for diabetic and non-diabetic patients.

Table 2.1: Standard human blood glucose content

Level	**mg/dL**	**mmol/L**	**Risk**
Dangerously High	315+	17.4	Very High
High	280	15.6	High
High	250	13.7	High
High	215	11.0	High
Borderline	180	10.0	Medium
Borderline	150	8.2	Medium
Borderline	120	7	Medium
Normal	108	6	No risk
Normal	72	4	No risk
Low	70	3.9	Medium
Dangerously Low	50	2.8	High

A stock solution can be prepared by weighing out appropriate portion of a solid or by measuring out an appropriate volume of a pure liquid and diluting to a known volume. The amount of solute (glucose) required to make the the glucose solution can be obtained using,

Mass of solute(g)=Concentration $\frac{mol}{L}\times$ *Relative Formula Mass*$\times Volume(L)$

The chemical formula for glucose, $C_6H_{12}O_6$ contains six carbons, twelve hydrogen and six oxygen. The relative atomic weight of C, H and O was obtained from the periodic table giving the relative formula mass as,

Relative Formula Mass=180.156

Ten samples of glucose solutions were prepared as shown in Table 2.2 using the glucose concentration data in blood taken from Table 2.1. The solute was measured using a measuring scale and was prepared in test tubes as in Figure 2.5.

Table 2.2: Sample preparation for varying glucose concentration

No	**Sample**	**Concentration (mol/L)**	**Relative formula mass (g)**	**Volume (L)**	**Mass of solute (g)**
1	S1	0.0040	180.156	0.250	0.180156
2	S2	0.0060	180.156	0.250	0.270230
3	S3	0.0070	180.156	0.250	0.315270
4	S4	0.0082	180.156	0.250	0.369300
5	S5	0.0100	180.156	0.250	0.450390
6	S6	0.0110	180.156	0.250	0.495400
7	S7	0.0137	180.156	0.250	0.617030
8	S8	0.0156	180.156	0.250	0.702600
9	S9	0.0174	180.156	0.250	0.783600
10	S10	0.0200	180.156	0.250	0.900780

Figure 2.5: Glucose solution of varying concentration according to Table 2.2.

Spectroscopic characterization of the prepared glucose solution was done using the $NeoSpectraMicro^{TM}$ Development kit and GUI interface with Windows operating system. The experiment was conducted in a controlled environment by pouring the prepared solution in a black container as shown in Figure 2.6 and obtaining the respective spectrum, and the experimental setup and data collection procedure is shown in Figure 2.7.

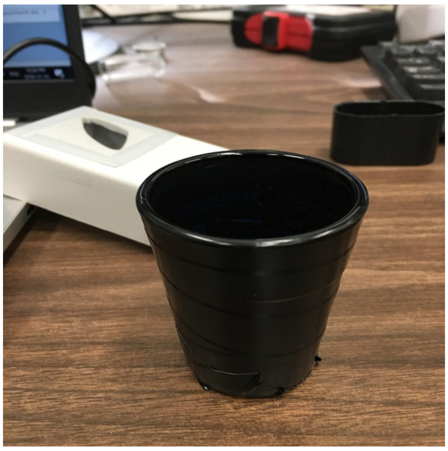


Figure 2.6: Sample container used during testing

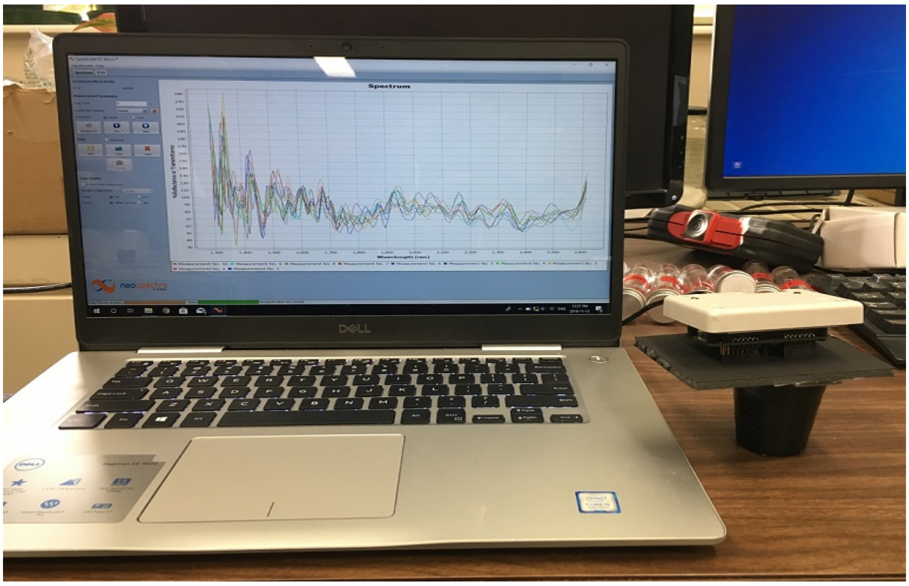
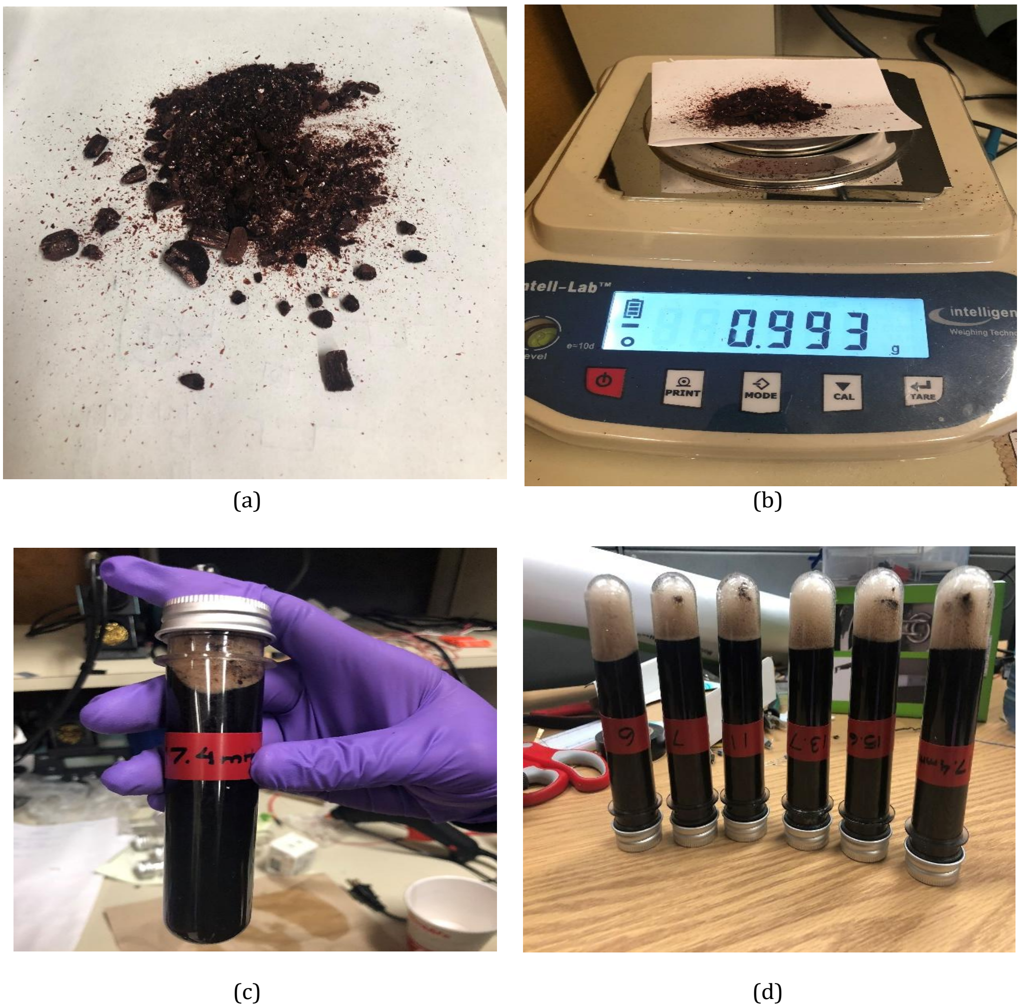


Figure 2.7: Experimental setup and data collection

Phase 3 of the project was conducted by utilizing the solution prepared in Phase 2. The only modification in phase 3 was the mixing of 1g of Hemoglobin power shown in Figure 2.8 with the previously prepared glucose solution according to Table 2.2.



(a) (b)

(c) (d)

Figure 2.8: Artificial blood sample preparation *(a)* artificial hemoglobin power, *(b)* measuring 1g hemoglobin power for each solution, *(c)* making artificial blood sample, *(d)* different concentration artificial blood sample

The experimental setup for Phase 3 was similar to the one used in Phase 2 as shown in Figure 2.9. Phase 4 of the project is aimed toward making and testing the methodology on a phantom finger [12] (i.e. an artificial finger) in order to mimic the original finger. The phantom finger as shown in Figure 2.10 is designed to mimic the scattering and absorption properties of the finger in near-infrared spectral region. Tissue scattering is modeled by the use of an aqueous suspension of lipid droplets ($Intralipid^{TM}$), whereas absorption is accounted for by direct use of red blood cells (RBC). Both components are mixed to mimic the blood-tissue compound and are inserted into a rectangular cuvette [12].

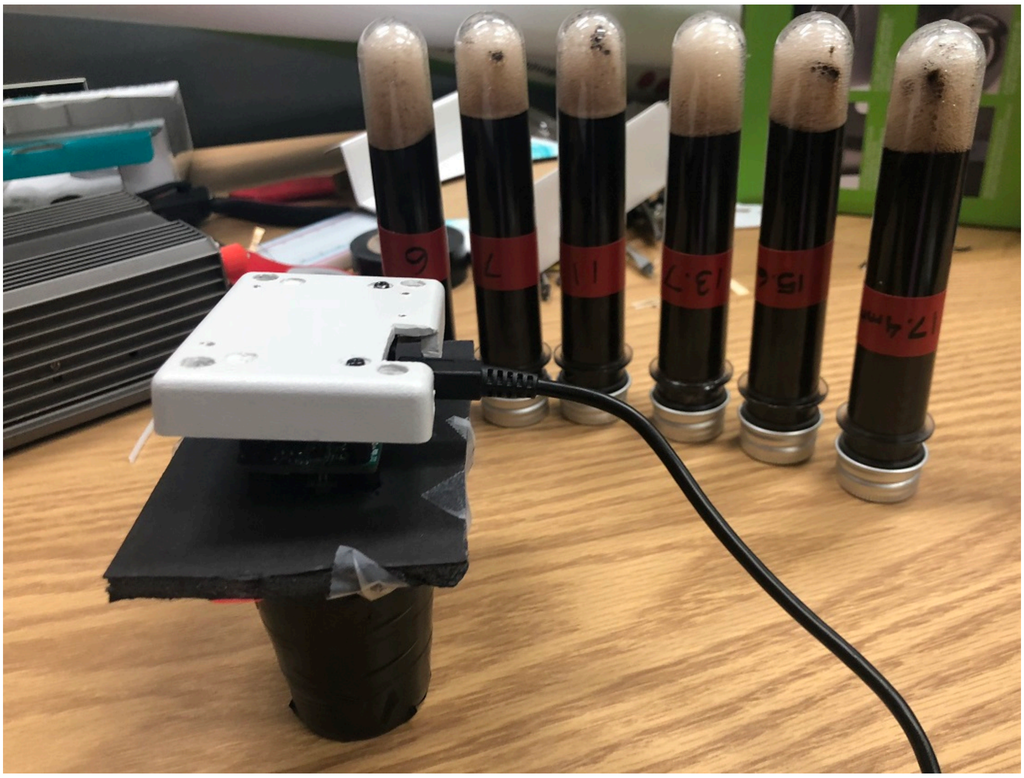


Figure 2.9: Experimental setup for phase 3: data collection of blood sample

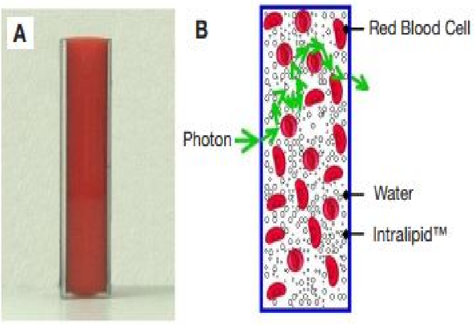


Figure 2.10: Phantom finger

Phase 5 of the project aims to test the selected methodology non-invasively on humans. Different parts of human body specially, ear lobe, finger web and finger cuticle, skin of the forearm, lip mucosa, oral mucosa, tongue, nasal septum, cheek and arm will subjected to testing using the $NeoSpectraMicro^{TM}$ Development kit and data will be collected for analysis using proper machine learning algorithm in order to accurately predict the blood glucose level in patients. The obtained result will be matched with the result obtained from the device currently used in market for measuring blood glucose level (i.e. finger pricking method). The final phase of this project will be the development of the sensing device and the interface so that it offers the features of portability, affordability, accuracy and high performance.

A new scale of blood glucose level has been developed in this project with the aim of properly predicting the concentration of glucose in blood as shown in Table 2.3.

Table 2.3: New proposed scale for measuring blood glucose level

No	Proposed Scale	mmol/L	Risk
1	Class 9	20,0	Very High
2	Class 8	17.4	Very High
3	Class 7	15.6	High
4	Class 6	13.7	High
5	Class 5	11.0	High
6	Class 4	10.0	Medium
7	Class 3	8.2	Medium
8	Class 2	7	Medium
9	Class 1	6	No risk
10	Class 0	4	No risk

Chapter 3

Result and Analysis

NeoSpectra Micro is a chip-sized, OEM Near InfraRed (NIR) spectral sensor that delivers the spectral response of the light absorbed by materials for quantification, qualification or identification. This chapter discussed the data collection procedure as well as data analysis by employing the various machine learning techniques.

3.1 Data Collection

NeoSpectra Micro development kit was utilized to obtained spectrum in Near-Infrared region ($1300 - 2500nm$) and the data was visualized using the SpectroMost GUI. Data for both glucose solution and glucose solution infused with blood sample was collected over a wide range (i.e. 10 samples). Data from the plots of reflectance/transmittance vs wavelength and absorbance vs wavelength was utilized to accurately predict the glucose concentration in respective samples using proper machine learning algorithms. Figure 3.1 shows a plot of reflectance vs wavelength of glucose solution whereas Figure 3.2 shows a plot of absorbance vs wavelength of glucose solution. It was observed that glucose shows significant effect for wavelength in the region of 1700nm – 2200nm.

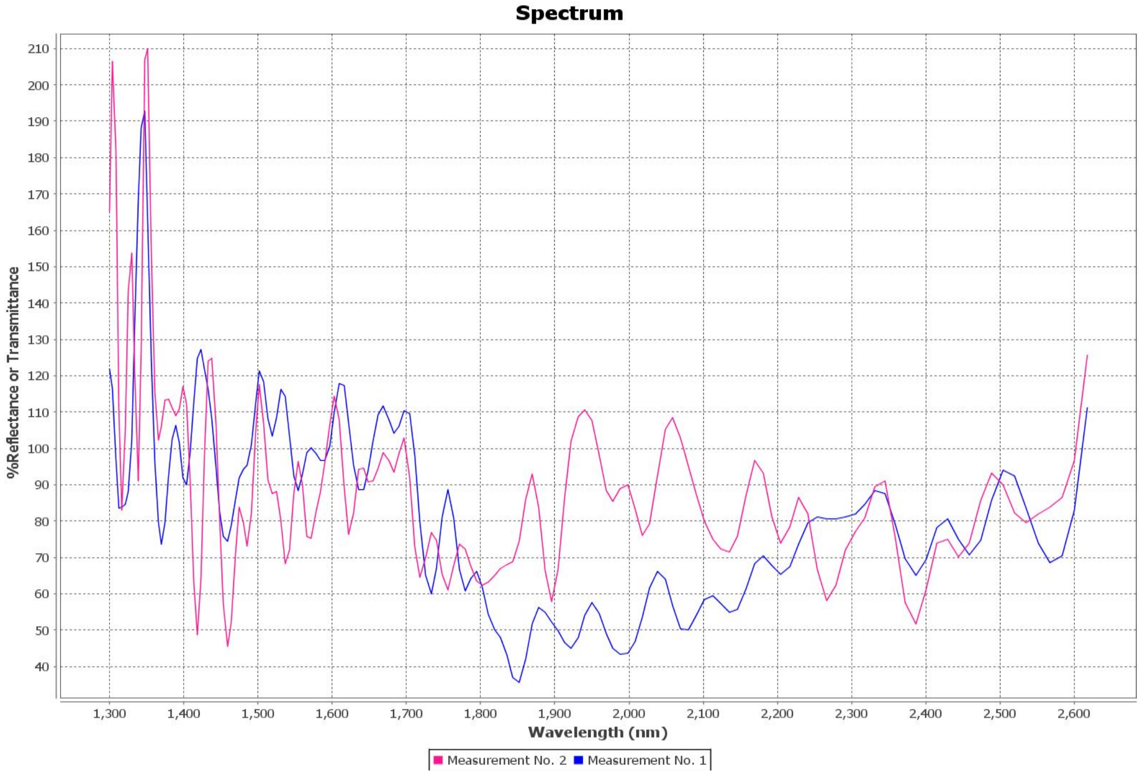


Figure 3.1: Plot of reflectance vs wavelength of glucose solution

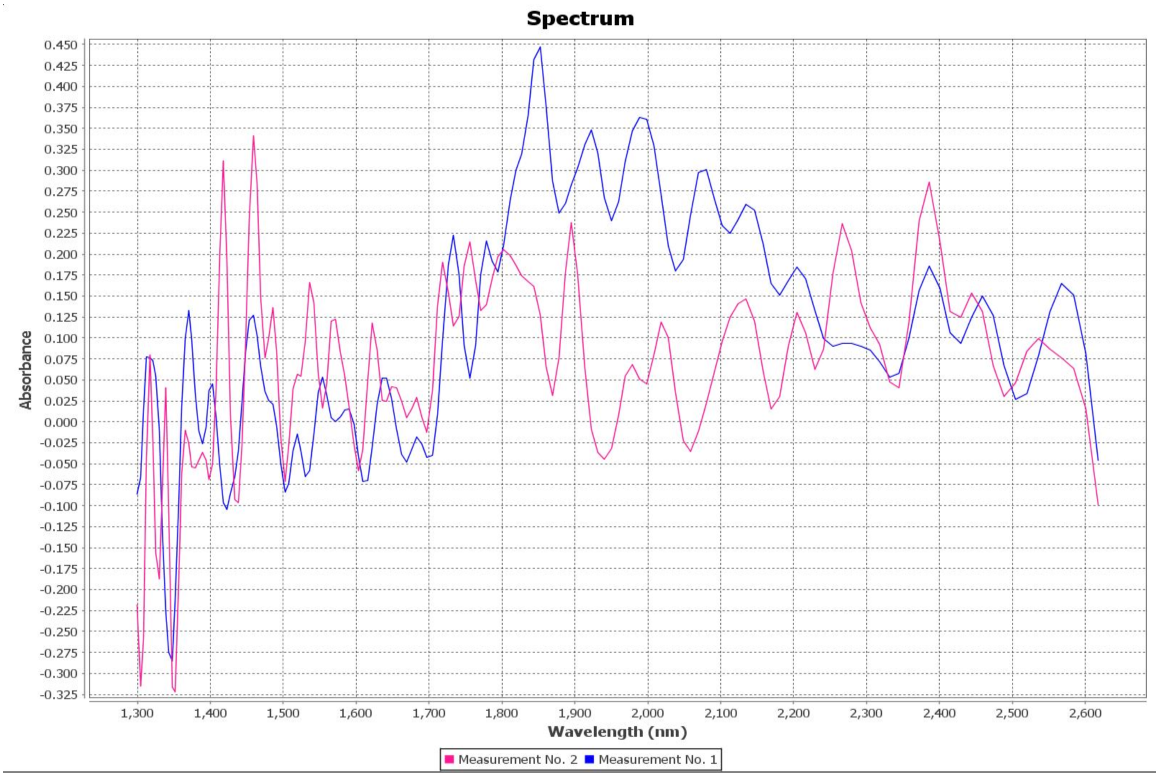


Figure 3.2: Plot of absorbance vs wavelength of glucose solution

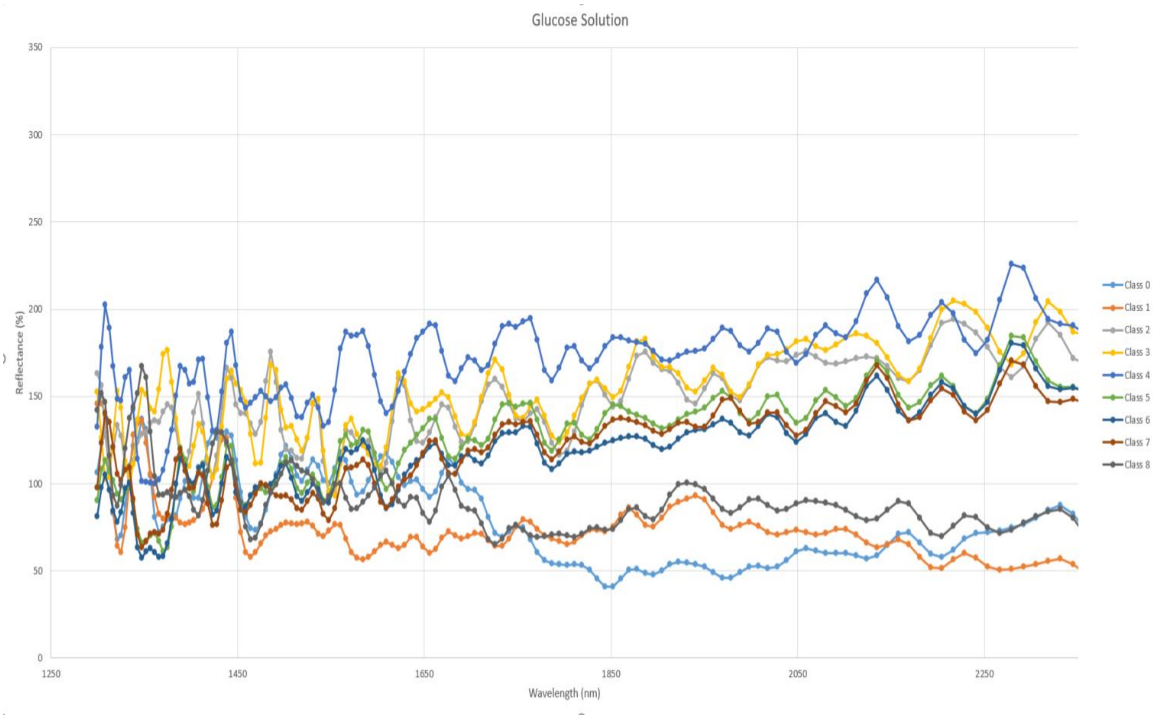


Figure 3.3: Plot of reflectance vs wavelength of glucose solution for the proposed scale

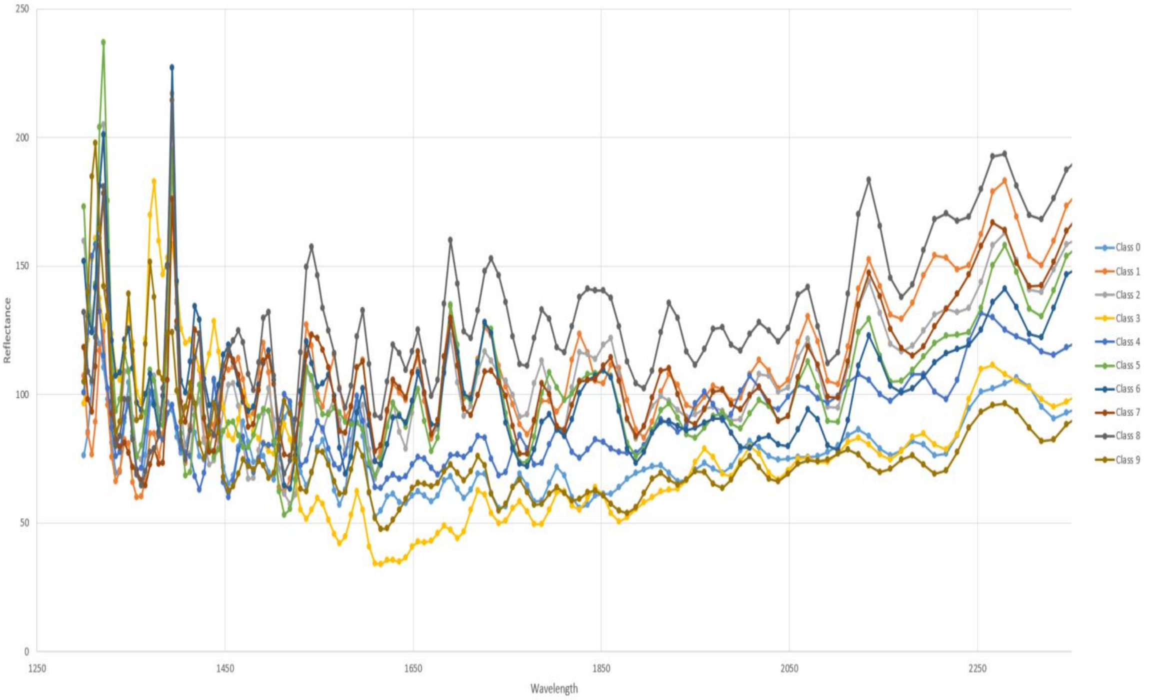


Figure 3.4: Plot of reflectance vs wavelength of blood glucose solution for the proposed scale

A comparison between glucose solution and glucose solution infused with blood was obtained in Figure 3.5 showing that the reflectance decreases as blood was introduced into the solution.

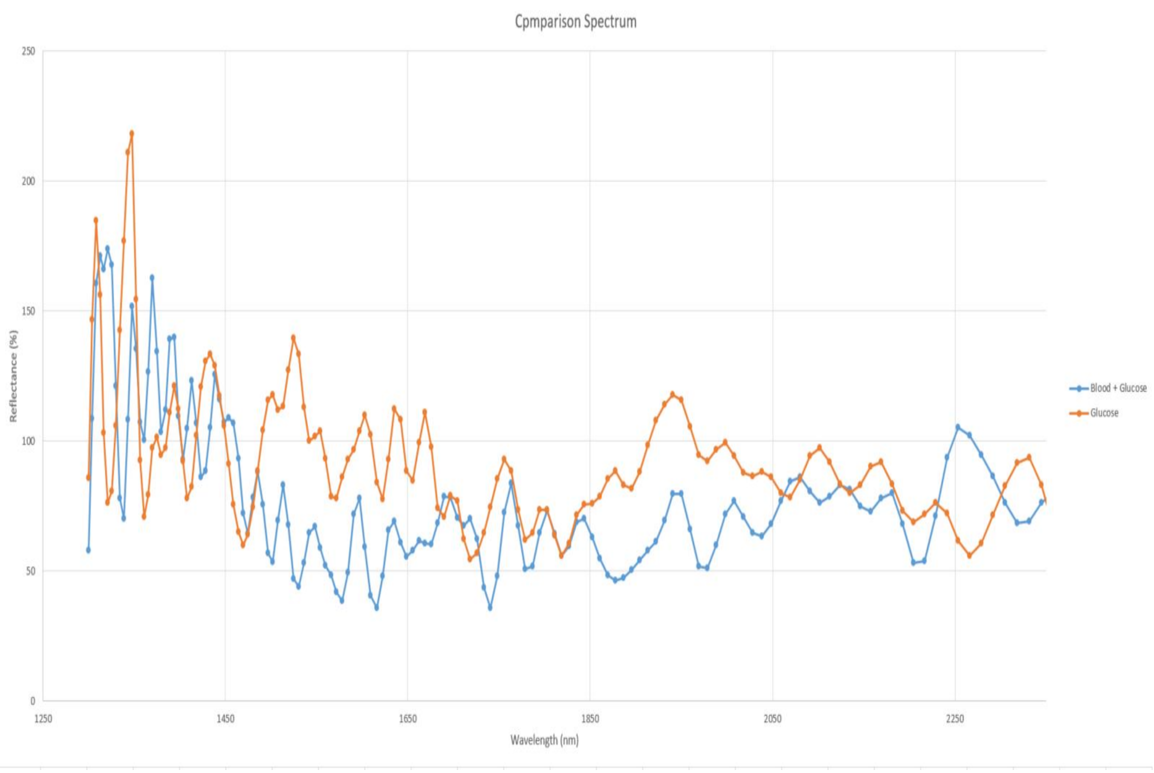


Figure 3.5: Comparison plot of reflectance vs wavelength of glucose and blood glucose solution for the proposed scale

3.2 Data Modeling and Prediction

3.2.1 Partial Least Squares Regression method

To get the best accuracy from this method applying in the data set, the optimum number of Partial Least Squares (PLS) components had to be measured. So, it was found that the number of PLS component is one for the least value of the Mean Squared Error (MSE). Figure 3.6 shows the Plot of finding the required number of PLS components for least value of MSE, and Figure 3.7 shows the relation between measured and predicted by PLS regression method.

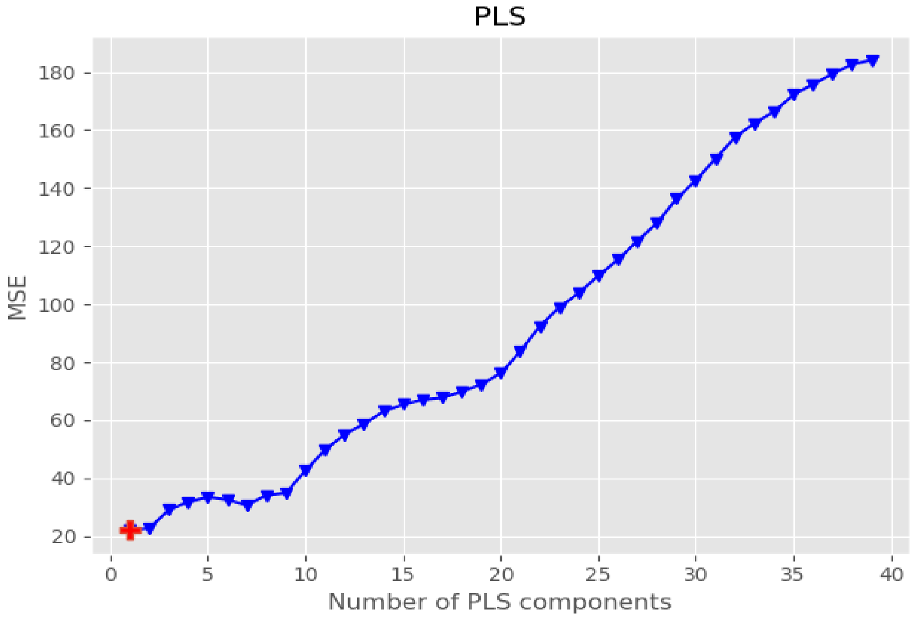


Figure 3.6: Plot of finding the required number of Partial Least Squares components for least value of Mean Squared Error

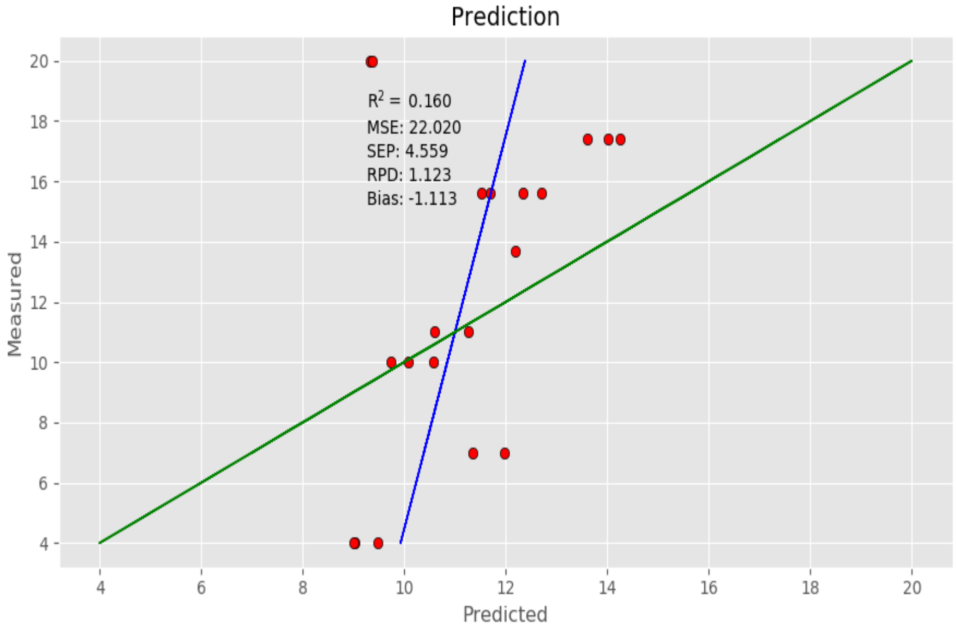


Figure 3.7: Relation between measured and predicted by Partial Least Squares Regression method

Table 3.1: Performance statistics of Partial Least Squares Regression method

$\mathbf{R_2}$	**MSE**	**SEP**	**RPD**	**Bias**
0.160	22.020	4,559	1.123	-1.113

Comments: *The coefficient of determination (R_2) of this method is 0.16 with MSE 22.02 which is not so good for prediction.*

3.2.2 SVM (kernel-linear)

Linear kernel supervised support vector machine method was used to model the data set. The 156 different wavelengths had been considered as features and different predefined glucose levels were labeled as classes for building the model.

Confusion matrix:

$$\begin{pmatrix}
5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 3 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 4 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 2 & 2 & 0 & 0 & 0 \\
0 & 4 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 4
\end{pmatrix}$$

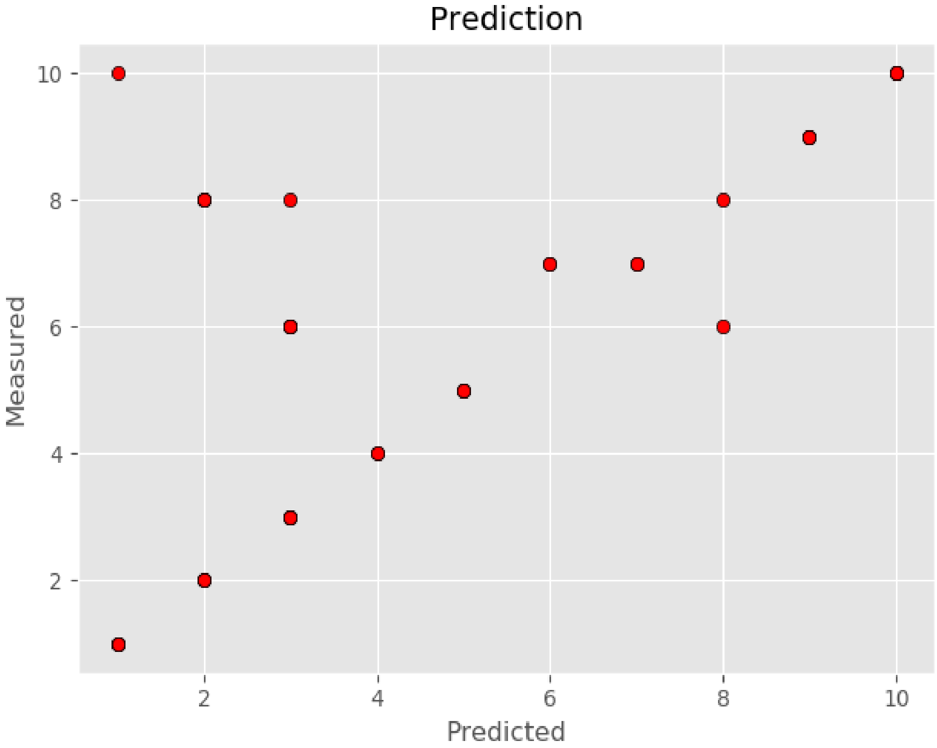


Figure 3.8: Relation between measured and predicted by using Support Vector Machine (SVM) method

Table 3.2: Performance statistics of Support Vector Machine (SVM) method

Class	**Precision**	**Recall**	**F1-score**	**Support**
1	0.83	1.00	0.91	5
2	0.43	1.00	0.60	3
3	0.44	1.00	0.62	4
4	1.00	1.00	1.00	2
5	1.00	1.00	1.00	3
6	0.00	0.00	0.00	5
7	1.00	0.50	0.67	4
8	0.50	0.17	0.25	6
9	1.00	1.00	1.00	3
10	1.00	0.80	0.89	5
Weighted avg	**0.68**	**0.68**	**0.64**	**40**

Comments: *The overall prediction accuracy of Support Vector Machine (SVM) is 67.5% . But, the prediction of Class 6 is zero percent, which is not expected. Moreover, prediction accuracy of class 2 and class 3 also less than 50%.*

3.2.3 Principal Component Analysis (PCA)

Support vector machine method showed better estimation than partial least squares regression method in comparison. The accuracy of the SVM method was improved by introducing principal component analysis (PCA). Feature scaling through standardization is an important preprocessing step for many machine learning algorithms. Standardization involves re-scaling the features such that they have the properties of a standard normal distribution with a mean of zero and a standard deviation of one. Principal component analysis is affected by scale so it was needed to scale the features in the data before applying PCA.

Confusion matrix:

$$\begin{pmatrix}
5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 3 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 3 & 0 & 0 & 2 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 2 & 2 & 0 & 0 & 0 \\
0 & 2 & 1 & 0 & 0 & 0 & 1 & 2 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 5
\end{pmatrix}$$

Choosing the of Principal Components: While choosing the number of principal components, 95% variance was used that means, it chose the minimum number of principal components such that 95% of the variance was retained. In this case, 95% of the variance amounts to 35 principal components. Now, Support vector machine was applied with 35 variables and 100 samples and the result was observed.

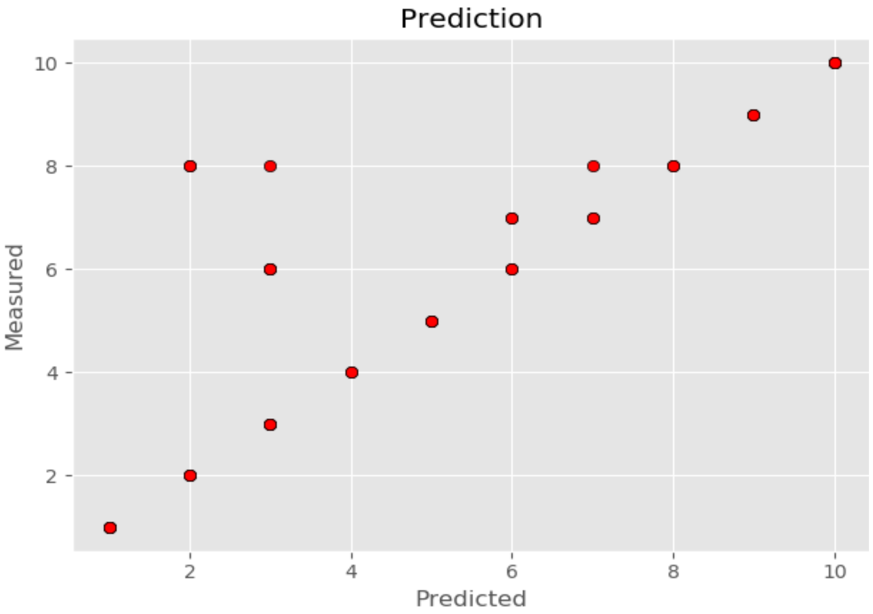


Figure 3.9: Relation between measured and predicted by using improved Support Vector Machine (SVM) method

Table 3.3: Performance statistics of improved SVM model by PCA analysis

Class	**Precision**	**Recall**	**F1-score**	**Support**
1	1.00	1.00	1.00	5
2	0.60	1.00	0.75	3
3	0.50	1.00	0.67	4
4	1.00	1.00	1.00	2
5	1.00	1.00	1.00	3
6	0.50	0.40	0.44	5
7	0.67	0.50	0.57	4
8	1.00	0.33	0.50	6
9	1.00	1.00	1.00	3
10	1.00	1.00	1.00	5
Weighted avg	**0.82**	**0.78**	**0.76**	**40**

Comments: *The overall prediction accuracy of improved Support Vector Machine (SVM) model by PCA analysis is 77.5% . Class 2, 3, 6 were predicted better than previous. The overall procession has also improved to 82 %.*

Appendix

APPENDIX A

PLSR:

```
from sys import stdout
importnumpy as np
import pandas as pd
importmatplotlib.pyplot as plt
fromsklearn.cross_decomposition import PLSRegression
fromsklearn.metrics import mean_squared_error, r2_score
fromsklearn.model_selection import train_test_split
fromsklearn.model_selection import train_test_split
# Run PLS including a variable number of components, up to 40, and calculate MSE
mse = []
component = np.arange(1,40)
for i in component:
pls = PLSRegression(n_components=i)
```

# Fit:

```
pls.fit(X_calib,Y_calib)
```

# Prediction

```
Y_pred = pls.predict(X_valid)
mse_p = mean_squared_error(Y_valid, Y_pred)
mse.append(mse_p)
comp = 100 * (i + 1) / 40
stdout.write(“%d%% completed” % comp)
stdout.flush()
stdout.write(“n”)
```

```
ifplot_components is True:
withplt.style.context((‘ggplot’)):
plt.plot(component, np.array(mse), ‘-v’, color= ‘blue’, mfc= ‘blue’)
plt.plot(component[msemin], np.array(mse)[msemin], ‘P’, ms=10, mfc= ‘red’)
plt.xlabel(‘Number of PLS components’)
plt.ylabel(‘MSE’)
plt.title(‘PLS’)
plt.xlim(xmin=-1)
plt.show()

# Run PLS with suggested number of components
pls = PLSRegression(n_components=msemin + 1)
pls.fit(X_calib, Y_calib)
Y_pred = pls.predict(X_valid)

# Calculate and print scores
score_p = r2_score(Y_valid, Y_pred)
mse_p = mean_squared_error(Y_valid, Y_pred)
sep = np.std(Y_pred[:, 0] - Y_valid)
rpd = np.std(Y_valid) / sep
bias = np.mean(Y_pred[:, 0] - Y_valid)
print(‘R2: %5.3f’ % score_p)
print(‘MSE: %5.3f’ % mse_p)
print(‘SEP: %5.3f’ % sep)
print(‘RPD: %5.3f’ % rpd)
print(‘Bias: %5.3f’ % bias)

# Plot regression and figures of merit
rangey = max(Y_valid) - min(Y_valid)
rangex = max(Y_pred) - min(Y_pred)
 z = np.polyfit(Y_valid, Y_pred, 1)
withplt.style.context((‘ggplot’)):
fig, ax = plt.subplots(figsize=(9, 5))
ax.scatter(Y_pred, Y_valid, c= ‘red’, edgecolors= ‘k’)
ax.plot(z[1] + z[0] * Y_valid, Y_valid, c= ‘blue’, linewidth=1)
ax.plot(Y_valid, Y_valid, color= ‘green’, linewidth=1)
plt.xlabel(‘Predicted’)
plt.ylabel(‘Measured’)
plt.title(‘Prediction’)

# Print the scores on the plot
plt.text(min(Y_pred) + 0.05 * rangex, max(Y_valid) - 0.1 * rangey, ‘R^2 = %5.3f’ %
score_p)
plt.text(min(Y_pred) + 0.05 * rangex, max(Y_valid) - 0.15 * rangey, ‘MSE: %5.3f’ %
mse_p)
plt.text(min(Y_pred) + 0.05 * rangex, max(Y_valid) - 0.2 * rangey, ‘SEP: %5.3f’ % sep)
plt.text(min(Y_pred) + 0.05 * rangex, max(Y_valid) - 0.25 * rangey, ‘RPD: %5.3f’ % rpd)
plt.text(min(Y_pred) + 0.05 * rangex, max(Y_valid) - 0.3 * rangey, ‘Bias: %5.3f’ % bias)
plt.show()
data = pd.read_csv(‘C:/Courses/bio project/Blood Glucose Dataset.csv’)
X = data.drop(‘ref’, axis=1)
Y = data[‘ref’]
X_calib = X[:80]
X_valid = X[81:]
Y_calib=Y[:80]
Y_valid=Y[81:]
X_valid, X_calib, Y_valid, Y_calib = train_test_split(X, Y, test_size = 0.80)
prediction(X_calib, Y_calib, X_valid, Y_valid, plot_components=True)
```

SVM

```
import pandas as pd
importnumpy as np
importmatplotlib.pyplot as plt
blooddata = pd.read_csv(“C:/Courses/bio project/Blood Glucose Dataset_final.csv”)
blooddata.head()
X = blooddata.drop(‘Class’, axis=1)
y = blooddata[‘Class’]
fromsklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.40, random_state=0)
fromsklearn.svm import SVC
```

```
svclassifier = SVC(kernel=‘linear’)
svclassifier.fit(X_train, y_train)
y_pred = svclassifier.predict(X_test)
fromsklearn.metrics import classification_report, confusion_matrix, accuracy_score
print(confusion_matrix(y_test,y_pred))
print(classification_report(y_test,y_pred))
print(accuracy_score(y_test,y_pred))
withplt.style.context((‘ggplot’)):
fig, ax = plt.subplots()
ax.scatter(y_pred, y_test, c=‘red’, edgecolors=‘k’)
plt.xlabel(‘Predicted’)
plt.ylabel(‘Measured’)
plt.title(‘Prediction’)
plt.show()
```

PCA+SVM:

```
import pandas as pd
importmatplotlib.pyplot as plt
fromsklearn.preprocessing import StandardScaler
fromsklearn.model_selection import train_test_split
fromsklearn.decomposition import PCA
fromsklearn.svm import SVC
fromsklearn.metrics import classification_report, confusion_matrix, accuracy_score
data = pd.read_csv(‘C:/Courses/bio project/Blood Glucose Dataset_final.csv’)
X = data.drop(‘Class’, axis=1)
y = data[‘Class’]
X = StandardScaler().fit_transform(X)
pca = PCA(.95)
X = pca.fit_transform(X)
print(X.shape)
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.40,random_state=0)
svclassifier = SVC(kernel=‘linear’)
svclassifier.fit(X_train, y_train)
y_pred = svclassifier.predict(X_test)
print(y_pred.shape)
print(confusion_matrix(y_test,y_pred))
print(classification_report(y_test,y_pred))
print(accuracy_score(y_test,y_pred))
print(y_test,y_pred)
withplt.style.context((‘ggplot’)):
fig, ax = plt.subplots()
ax.scatter(y_pred, y_test, c=‘red’, edgecolors=‘k’)
plt.xlabel(‘Predicted’)
plt.ylabel(‘Measured’)
plt.title(‘Prediction’)
plt.show()
```

Bibliography

[1] M. W. Steffes and D. B. Sacks, "Measurement of circulating glucose concentrations: the time is now for consistency among methods and types of samples," 2005.

[2] N. D. I. Clearinghouse, "National diabetes statistics," 2011.

[3] A. D. Association *et al.*, "Accuracy of the glucowatch g2 biographer and the continuous glucose monitoring system during hypoglycemia: experience of the diabetes research in children network," *Diabetes care*, vol. 27, no. 3, pp. 722–726, 2004.

[4] A. D. Association *et al.*, "Diagnosis and classification of diabetes mellitus," *Diabetes care*, vol. 37, no. Supplement 1, pp. S81–S90, 2014.

[5] M. Sabokdast, M. Habibi-Rezaei, A. A. Moosavi-Movahedi, M. Ferdousi, E. Azimzadeh-Irani, and N. Poursasan, "Protection by beta-hydroxybutyric acid against insulin glycation, lipid peroxidation and microglial cell apoptosis," *DARU Journal of Pharmaceutical Sciences*, vol. 23, no. 1, p. 42, 2015.

[6] S. Coster, M. Gulliford, P. Seed, J. Powrie, and R. Swaminathan, "Monitoring blood glucose control in diabetes mellitus: a systematic review," *BRITISH JOURNAL OF CLINICAL GOVERNANCE-BRADFORD-*, vol. 5, no. 4, pp. 225–227, 2000.

[7] C. for Disease Control, Prevention, *et al.*, "National diabetes statistics report: estimates of diabetes and its burden in the united states, 2014," *Atlanta, GA: US Department of Health and Human Services*, vol. 2014, 2014.

[8] Y. Date, M. Nakazato, S. Hashiguchi, K. Dezaki, M. S. Mondal, H. Hosoda, M. Kojima, K. Kangawa, T. Arima, H. Matsuo, *et al.*, "Ghrelin is present in pancreatic α-cells of humans and rats and stimulates insulin secretion," *Diabetes*, vol. 51, no. 1, pp. 124–129, 2002.

[9] K. M. Bratlie, R. L. York, M. A. Invernale, R. Langer, and D. G. Anderson, "Materials for diabetes therapeutics," *Advanced healthcare materials*, vol. 1, no. 3, pp. 267–284, 2012.

[10] L. Guariguata, T. Nolan, J. Beagley, *et al.*, “International diabetes federation. idf diabetes atlas,” 2014.

[11] R. Poddar, J. T. Andrews, P. Shukla, and P. Sen, “Non-invasive glucose monitoring techniques: A review and current trends,” *arXiv preprint arXiv:0810.5755*, 2008.

[12] O. Amir, D. Weinstein, S. Zilberman, M. Less, D. Perl-Treves, H. Primack, A. Weinstein, E. Gabis, B. Fikhte, and A. Karasik, “Continuous noninvasive glucose monitoring technology based on “occlusion spectroscopy”,” 2007.

[13] A. Srivastava, M. K. Chowdhury, S. Sharma, and N. Sharma, “Blood glucose monitoring using non invasive optical method: Design limitations and challenges,” *International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering*, vol. 2, no. 1, 2013.

[14] N. A. B. A. Salam, W. H. bin Mohd Saad, Z. B. Manap, and F. Salehuddin, “The evolution of non-invasive blood glucose monitoring system for personal application,” *Journal of Telecommunication, Electronic and Computer Engineering (JTEC)*, vol. 8, no. 1, pp. 59–65, 2016.

[15] K.-U. Jagemann, C. Fischbacher, K. Danzer, U. A. Mueller, and B. Mertes, “Application of near-infrared spectroscopy for non-invasive determination of blood/tissue glucose using neural networks,” *Zeitschrift für Physikalische Chemie*, vol. 191, no. 2, pp. 179–190, 1995.

[16] R. Waynant and V. Chenault, “Overview of non-invasive fluid glucose measurement using optical techniques to maintain glucose control in diabetes mellitus,” *LEOS newsletter*, vol. 12, no. 2, pp. 3–6, 1998.

[17] H. Yki-Järvinen and T. Utriainen, “Insulin-induced vasodilatation: physiology or pharmacology?,” *Diabetologia*, vol. 41, no. 4, pp. 369–379, 1998.

[18] P. Oomen, G. Kant, R. Dullaart, W. Reitsma, and A. Smit, “Acute hyperglycemia and hyperinsulinemia enhance vasodilatation in type 1 diabetes mellitus without increasing capillary permeability and inducing endothelial dysfunction,” *Microvascular research*, vol. 63, no. 1, pp. 1–9, 2002.

[19] R. G. Sibbald, S. J. Landolt, and D. Toth, “Skin and diabetes.,” *Endocrinology and metabolism clinics of North America*, vol. 25, no. 2, pp. 463–472, 1996.

[20] V. M. Monnier, O. Bautista, D. Kenny, D. R. Sell, J. Fogarty, W. Dahms, P. A. Cleary, J. Lachin, and S. Genuth, "Skin collagen glycation, glycoxidation, and crosslinking are lower in subjects with long-term intensive versus conventional therapy of type 1 diabetes: relevance of glycated collagen products versus hba1c as markers of diabetic complications. dcct skin collagen ancillary study group. diabetes control and complications trial.," *Diabetes*, vol. 48, no. 4, pp. 870–880, 1999.

[21] S.-j. Yeh, O. S. Khalil, C. F. Hanna, and S. Kantor, "Near-infrared thermo-optical response of the localized reflectance of intact diabetic and nondiabetic human skin," *Journal of biomedical optics*, vol. 8, no. 3, pp. 534–545, 2003.

[22] G. Mazarevica, T. Freivalds, and A. Jurka, "Properties of erythrocyte light refraction in diabetic patients," *Journal of biomedical optics*, vol. 7, no. 2, pp. 244–248, 2002.

[23] R. Fusman, R. Rotstein, K. Elishkewich, D. Zeltser, S. Cohen, K. Kofler, D. Avitzour, N. Arber, S. Berliner, and I. Shapira, "Image analysis for the detection of increased erythrocyte, leukocyte and platelet adhesiveness/aggregation in the peripheral blood of patients with diabetes mellitus," *Acta diabetologica*, vol. 38, no. 3, pp. 129–134, 2001.

[24] S. F. Malin, T. L. Ruchti, T. B. Blank, S. N. Thennadil, and S. L. Monfre, "Noninvasive prediction of glucose by near-infrared diffuse reflectance spectroscopy," *Clinical chemistry*, vol. 45, no. 9, pp. 1651–1658, 1999.

[25] O. S. Khalil, "Non-invasive glucose measurement technologies: an update from 1999 to the dawn of the new millennium," *Diabetes technology & therapeutics*, vol. 6, no. 5, pp. 660–697, 2004.

[26] J. R. Koza, F. H. Bennett, D. Andre, and M. A. Keane, "Automated design of both the topology and sizing of analog electrical circuits using genetic programming," in *Artificial Intelligence in Design'96*, pp. 151–170, Springer, 1996.

[27] R. Marbach, T. Koschinsky, F. Gries, and H. Heise, "Noninvasive blood glucose assay by near-infrared diffuse reflectance spectroscopy of the human inner lip," *Applied Spectroscopy*, vol. 47, no. 7, pp. 875–881, 1993.