2021-08-01

Non-Invasive In-Vitro Glucose Monitoring Using Optical Sensor And Machine Learning Techniques For Diabetes Applications

Maryamsadat Shokrekhodaei
University of Texas at El Paso

NON-INVASIVE IN-VITRO GLUCOSE MONITORING USING OPTICAL SENSOR AND MACHINE LEARNING TECHNIQUES FOR DIABETES APPLICATIONS

MARYAMSADAT SHOKREKHODAEI

Doctoral Program in Electrical and Computer Engineering

APPROVED:

Stella Quinones, Ph.D., Chair

Robert C. Roberts, Ph.D., Co-Chair

David Cistola, M.D., Ph.D.

Mahesh Narayan, Ph.D.

Stephen L. Crites, Jr., Ph.D.
Dean of the Graduate School

NON-INVASIVE IN-VITRO GLUCOSE MONITORING USING OPTICAL SENSOR AND MACHINE LEARNING TECHNIQUES FOR DIABETES APPLICATIONS

by

MARYAMSADAT SHOKREKHODAEI, MS

DISSERTATION

Presented to the Faculty of the Graduate School of

The University of Texas at El Paso

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

Department of Electrical and Computer Engineering

THE UNIVERSITY OF TEXAS AT EL PASO

August 2021

Acknowledgements

For most, I would like to express my sincere gratitude to my PhD advisor Dr. Stella Quinones who gives a lot of her time to support me throughout my research period. Her patient, enthusiasm, kindness and timely advice have helped me to accomplish my career goals.

I would like to extend my gratitude to my PhD co-advisor, Dr. Robert C. Roberts for his major contribution to my research, insightful comments, and support, and the rest of my PhD committee members, Prof. David Cistola and Prof. Mahesh Narayan for their invaluable help, guidance and advices for accomplishment of this task.

I sincerely thank to electrical and computer engineering department chair, Prof. Miguel Velez-Reyes for all of his support during my PhD study. I would also like to thank UTEP College of Engineering and Graduate School.

I owe a deep sense of gratitude to my husband Mahdi, for all his support, love, patience, and constant encouragement. I would also like to thank to my wonderful parents for supporting me throughout my life.

Abstract

Diabetes is a major public health challenge affecting more than 451 million people. Physiological and experimental factors influence the accuracy of non-invasive glucose monitoring, and these need to be addressed before replacing the finger prick method with a non-invasive glucose measurement technique. Also, the suitable employment of machine learning techniques on experimental data can significantly improve the accuracy of glucose predictions.

This work includes the design, development, testing and data analysis of an optical based sensor for glucose measurements. The feasibility of non-invasive measurement of glucose within aqueous solutions that assimilate the composition of human blood plasma is investigated. The laboratory testing of the sensor with controlled solutions helps to make valid conclusions about the performance of the sensor and the accuracy of the predictions. There are several goals associated with this study.

The first goal is to use light sources with multiple wavelengths to enhance the sensitivity and selectivity of glucose detection in the aqueous solution. Multiple wavelength measurements have the potential to compensate for errors associated with inter- and intra-individual differences in blood and tissue components. In this study, the transmission measurements of the custom built optical sensor are examined using 18 different wavelengths between 410 and 940 nm. Results show a high correlation value (0.98) between glucose concentration and transmission intensity for 4 of the 18 wavelengths (485, 645 and 860 and 940 nm).

The second goal of this study is to extract the intensity data using these 4 wavelengths and to collectively analyze the accuracy of glucose concentration predictions based on machine learning techniques. The intensity data measured using the four optimal wavelengths is first input into a support vector machine (SVM) classifier to discriminate hypoglycemic, normal and

hyperglycemic ranges with 99% accuracy based on F1-scores. Then, for each of the three ranges, a feed-forward neural network model is developed and applied to a test dataset to predict the glucose concentration within each range separately. The combination of both the classification and regression models to predict glucose concentrations for a given dataset results in more accurate and reliable predictions compared to using a single regression model for the entire glucose range. The use of hybrid models improves the root mean square error (RMSE) from 12.7 mg/dL (in the case of a single regression model) to 9.3 mg/dL (in the case of hybrid models as was done in this study).

Using this hybrid approach also results in 100% of the glucose readings falling within zones A and B of the Clarke error grid. This is an important step towards critical diagnosis during an emergency patient situation. Future work should include an in-vivo study to test the sensor and data analysis approach used here on human blood samples.

Table of Contents

List of Tables

List of Figures

1. Introduction

Diabetes mellitus is a metabolic disorder that affects the body's ability to process blood glucose. Diabetic patients monitor their blood glucose levels in an effort to keep them in the normal range (approximately 70 to 180 mg/dL, depending on food intake) by medication, exercise, proper diet, etc. Blood glucose levels below 70 mg/dL serves as an alert for possible life-threatening hypoglycemia, and glucose levels higher than 180 mg/dL could indicate clinically significant hyperglycemia. Patients with diabetes are unable to produce (Type 1) or properly use the insulin hormone (Type 2). Insulin is a glucose regulatory hormone that interacts with insulin receptors in a process that allows cells to absorb glucose as a source of energy. The number of diabetics in the world is substantial and is increasing. The World Health Organization (WHO) estimates that there will be 693 million diabetics (age 18-99 years) worldwide by 2045, compared to 451 million in 2017 [1]. In the US, the number of diabetics is estimated to increase by 54%, from 35.6 million in 2015 to more than 54.9 million by 2030 [2]. This is estimated to increase the total annual cost associated with diabetes (medical and non-medical) by 53%, from $407.6 billion in 2015, to more than $622.3 billion by 2030 [2]. Long-term diabetes leads to chronic complications such as heart disease, kidney disease, stroke, vision loss and nervous system damage.

Diabetes is classified as: Type 1 diabetes (T1D), Type 2 diabetes (T2D), gestational diabetes due to pregnancy, and other types of diabetes (caused by monogenic diabetes syndromes, diseases of the exocrine pancreas or drug induced diabetes) [3]. In Type 1 diabetes, beta (β) cells in the pancreas are destroyed due to an autoimmune response and without β cells to detect glucose, insulin is not released into the bloodstream. Without insulin, cells cannot absorb glucose and thus glucose in the body rises to dangerous levels. As a result, Type 1 diabetics needs to be controlled with insulin supply to maintain constant blood glucose levels. Patients with Type 2 diabetes

produce insulin, but the body does not respond properly to the insulin hormone. The insulin receptors that allow glucose to enter cells may be damaged or desensitized to insulin. Type 2 diabetics may have normal or elevated insulin levels; however, this may be insufficient to compensate for the body's insulin resistance [3]. Type 2 diabetics need to enhance their body's sensitivity to insulin and this can be accomplished by exercise, diet, insulin therapy and weight loss.

1.1 MOTIVATION

There are chronic complications associated with long term diabetes such as heart disease, kidney disease, stroke, vision loss and nervous system damage, and patients with long term complications may lead to premature death. Frequent monitoring, ease of blood glucose measurement, real time measurement and accuracy is instrumental for better control and management of diabetes. Self-monitoring of blood glucose (SMBG) by the conventional finger prick method is the most accurate glucose detection method to date. However, this method is painful, inconvenient and carries a risk of infection, especially for patients who are required to check their blood glucose levels several times a day. Non-invasive glucose monitoring technology has the potential to ease glucose detection and can lead to more tailored treatment options [4].

Research groups have been trying for decades to release a reliable and long-term accurate commercial non-invasive glucose device [4]–[6]. However, none of the current non-invasive devices exhibit the required accuracy and long-term consistency required to replace the more accurate finger prick methods. Therefore, additional work is required to produce a non-invasive glucose sensor applicable for public use under regular home conditions. Several non-invasive glucose measurement techniques have been thoroughly investigated and include:

- Infrared spectroscopy [7]–[9]

- Raman spectroscopy [10], [11]
- Polarimetry [12]
- Photo-acoustic spectroscopy [13]
- Millimeter wave/microwaves spectroscopy [14]–[16]
- Optical coherence tomography [17]
- Other hybrid techniques [18], [19]

The technology behind non-invasive glucose measurement techniques can be categorized as: (1) optical based methods or (2) electrical based methods (See Figure 1.1). The focus of this work is on the design, test and evaluation of an optical based glucose sensor.

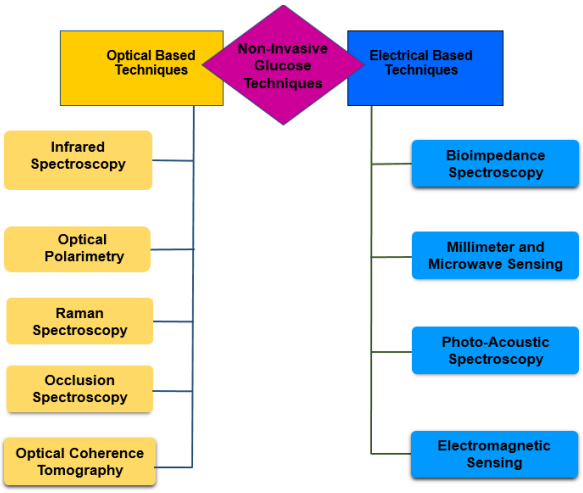


Figure 1.1: Categorization of non-invasive glucose measurement techniques.

The technology to support optical based sensing of blood glucose is developing at a fast pace. The optical based methods are promising and function using the properties of light that interact with glucose molecules in a manner that is concentration-dependent over a wide range of frequencies [4]. Optical frequencies can be strategically selected based on historical data to consciously eliminate signals from molecules that produce errors in glucose readings. Physiological processes and components influence glucose sensor inputs and can be considered in

the sensor design in order to improve the accuracy of measurement data. Optical sensing methods are fast, low cost, portable, easy to implement, and do not require consumable reagents.

Among the optical based sensing methods, the visible/infrared spectroscopy methods demonstrate a higher potential for accurate glucose sensing [4], [20], [21]. This type of sensor is analyzed using in vivo experiments and in vitro experiments, and yields acceptable accuracy. However, additional work is required to ensure that these glucose sensors can provide accurate and consistent glucose data for a wide range of diabetic patients. Sensors need to be able to accurately distinguish between the hypoglycemic, normal and hyperglycemic ranges. Experimental factors and physiological factors that contribute to errors need to be investigated further in order to improve the sensor design and the elimination of unwanted signals.

This work attempts to examine the effect of 2 main tissue and blood plasma components and their influence on the accuracy of glucose measurements within a phosphate buffered saline (PBS) solution:

1. water (the main constituent in blood and tissue), and
2. albumin (the dominant protein within the blood plasma)

Since the concentration of glucose in the blood is much lower than any of the other blood components, the role that these “other” components have on the absorption and scattering of light may be more significant compared to that of glucose molecules. This makes it challenging to measure or detect fluctuations in glucose concentration. Thus, the main purpose of this research is to design a sensor with a sensitivity and selectivity to glucose molecules (in comparison to water and albumin) using a methodological approach that supports multi wavelength analysis. In addition, there are many factors that affect the repeatability of glucose readings which needs to be investigated, including motion artifacts. This research will examine a sensor design and the data

collection, data analysis and data processing methods that have an effect on the sensitivity, selectivity and repeatability challenges.

Data collection, data analysis and glucose prediction models (data processing) contribute to the accuracy of glucose monitoring techniques. Various machine learning approaches are used in glucose prediction models:

- multiple linear regression [22]
- partial least square regression [23], [24]
- principal component regression [23]
- feed-forward neural networks [25], [26]
- deep neural networks [27], [28]
- support vector machines [28], [29]
- random forest regression [19]

These regressing-based models predict continuous values of glucose concentration based on one or more sets of input data. The goal of regression techniques is to train a model to minimize the error between an actual and predicted glucose concentration value. This process is not directly involved in reducing the error associated with the detection of hypo- or hyperglycemia. Thus, predictions made by regression models may be limited in their ability to accurately predict hyperglycemia and hypoglycemia, especially in real life applications where various factors affect the measured signal.

A suitable employment of machine learning techniques can significantly improve the accuracy of glucose predictions. The second aim of this research is to improve glucose prediction accuracy, and to reduce error in identification of hypoglycemia and hyperglycemia in patients by

creating a glucose prediction model using a combination of machine learning classification and regression techniques. The specific goals associated with this research include:

Goal 1: Increase the sensitivity and selectivity of an optical sensor to detect changes in glucose concentration within an aqueous solution by using multiple light sources and detectors, along with multiple wavelength analysis.

Goal 2: Increase the accuracy of glucose predictions using a combination of machine learning classification and regression techniques.

1.2 CONTRIBUTION

The non-invasive measurement of blood glucose in the body is affected by many physiological, patient-specific and technological factors that influence the measurement and cause false readings. It is important to identify components in the body, motion artifacts, and other measurement design parameters that contribute to inaccurate measurements. The literature and carefully designed in vitro experiments help to identify obstacles to accurate glucose readings. In vitro experiments make it possible to measure the influence of confounding factors, and to identify methodological or theoretical approaches to improve the accuracy of glucose measurements.

In this work, an optical based non-invasive sensor for measuring glucose levels within an aqueous solution is designed, implemented and calibrated. The optical based sensor is comprised of visible (VIS) and near infrared (NIR) light sources illuminating a sample of glucose, and a detector that measures the intensity of light passing through the sample in the range from 410 nm to 940 nm. The measured light intensity correlates with the optical property of light (absorption and scattering) which itself depends on the glucose concentration [4]. The use of VIS-NIR over mid infrared (MIR) light sources for the detection of glucose is preferred due to the lower absorption of VIS-NIR light by tissue and water content in the blood compared to MIR light. The

strong absorption of MIR light by water decreases the amount of light that is able to penetrate the blood and tissue samples and the corresponding intensity of light that is picked up by the detector. The stronger absorption of MIR compared to VIS-NIR light means that MIR light can only penetrate the interstitial fluid of tissue whereas VIS-NIR light can penetrate deeper into the tissue and interact with glucose molecules within both the interstitial fluid and the blood vessels [4]. It takes time for glucose to diffuse from blood vessels into the interstitial fluid, and thus there is a difference between the glucose concentration levels within the blood vs the interstitial fluid at any given time [4]. VIS-NIR light can interact with glucose molecules within the blood in real time, thus increasing the sensitivity of the sensor device to read real time changes in the glucose concentration.

In vitro glucose measurement experiments consist of glucose in a phosphate buffered saline (PBS) solution that is supplemented with albumin. Since the glucose concentration in blood is much lower than other blood and tissue components, the role of these components in the measured signal may be more dominant compared to that of the glucose molecules. These components include water which is the main content in blood and tissue, and protein such as albumin which is a dominant protein in blood plasma. Furthermore, the variation of these components between individuals creates errors in glucose readings. This makes it challenging to detect and measure fluctuations in glucose concentrations among individuals. In this research, the effort is to reduce the influence of these confounding factors by developing a glucose sensor that measures light intensity using multiple light wavelength sources and detectors. The light attenuation coefficient is unique for each species and depends on the wavelength of the incident light interacting with each molecule. This will help to compensate the error associated with only one light source. The hypothesis for this work is that a sensor with multiple light sources and detectors can extract more

accurate quantitative information as a result of the combined light interaction with the different species in glucose aqueous solutions. A greater sensitivity and selectivity to glucose concentration is achieved by measuring and analyzing the intensity of light with four distinct wavelengths (485, 645 and 860 and 940 nm) passing through an aqueous glucose sample. These four wavelengths will be identified as "optimal wavelengths" for glucose measurements due to a high correlation coefficient between glucose concentration and the measured light intensity.

Further efforts are made to improve the accuracy of a glucose prediction model by using a hybrid machine learning (ML) approach, as described below.

1. First, a ML classification model is trained and optimized in order to classify the intensity data samples measured at the optimal wavelengths into 3 classes that are labeled hypoglycemic, normal and hyperglycemic. The data samples within the three classes are associated with glucose concentration values from 40 to 70 mg/dL, 70 to 180 mg/dL, and 180 to 250 mg/dL.
2. After identifying a class for a measured intensity data sample as described in Step 1, a feed-forward neural network (FFNN) regression model is used to predict the glucose concentration. There are three regression models (FFNN1, FFNN2, and FFNN3), and each model is used to predict glucose concentration values within each of the three classes (hypoglycemic, normal and hyperglycemic).

Using both classification and regression models results in a greatest capability in the identification of hypoglycemia, normal and hyperglycemia conditions as well as the lowest error in glucose prediction so that the real value of glucose concentration is in high agreement with the predicted value for glucose concentrations.

The dissertation is organized as follows:

Chapter 2 begins with a literature review of blood physiological factors and their relationship with blood glucose concentrations. This is followed by a review of non-invasive glucose measurement methods, the challenges associated with each method, and possible solutions to overcome those challenges. Chapter 3 includes design details associated with the first version of the glucose sensor along with preliminary results. This sensor was built to investigate the feasibility of glucose sensing using this type of optical sensor. Chapter 4 describes the steps taken to improve the design and implementation of the first version of the optical based glucose sensor. This sensor is the main focus of this study, and the new sensor design and experimental methods employed to collect the transmission intensity data for glucose within an aqueous sample are described in detail. Chapter 5 explains the data analysis approach to identify the optimal wavelengths for accurate glucose measurements. After the optimal wavelengths are identified, five machine learning techniques are tested and analyzed. These models train and optimize glucose predictions based on the sensor design described in Chapter 4. Chapter 6 includes the results obtained after evaluating the performance of each of the machine learning models and prediction results. 5 These results are compared to glucose prediction accuracy results from existing sensors from the literature. Chapter 7 includes conclusions associated with the observations and results for this study, their relevance to current glucose sensing research, and recommended future work.

2. Literature Review

Interdisciplinary collaboration is essential for realizing non-invasive blood glucose monitoring devices. The goal of this Section is to provide an overview of (1) the relative physiological and biochemistry knowledge needed to connect the sensor design to the components that need to be measured, and (2) the engineering knowledge required to design an optical based non-invasive sensor. There is a third component related to the data analysis and machine learning techniques, which will be discussed in Chapter 5.

2.1 Physiological Aspects

Accuracy of glucose measurements is dependent on blood components, glucose transportation in the body, thickness and composition of skin tissue layers, and the physiological characteristics related to blood glucose molecules. Glucose molecules interact with blood/tissue components and dominant components within blood/tissue, and this affects the accuracy of glucose sensors. These physiological characteristics are discussed in the Sections below in order to consider how some of these issues influence the sensor design.

2.1.1 Body Water Content

Figure 2.1 illustrates the distribution of water content in the body [30]. More than 50% of body weight is water, and about two thirds of the water is intracellular fluid and one third is the extracellular fluid. Extracellular fluid is mainly composed of interstitial fluid surrounding tissue cells, a small portion of intravascular fluid (or blood plasma), and a negligible portion of cerebrospinal fluid. Interstitial fluid is located between vessels and tissue cells and contains glucose. Intravascular fluid or plasma also contains glucose.

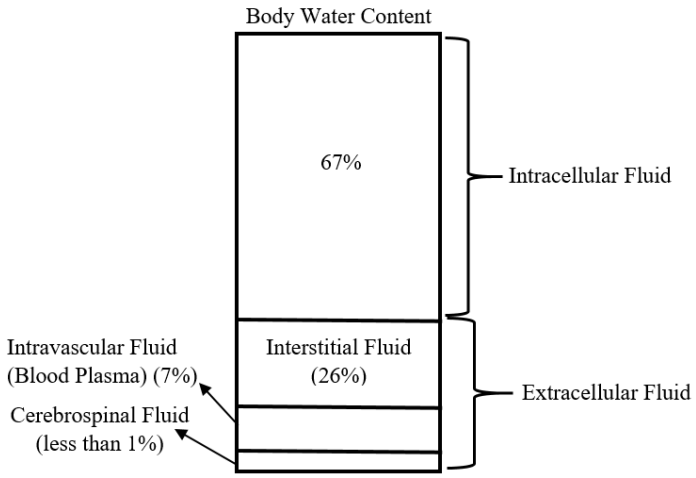


Figure 2.1: Water content distribution in the body.

2.1.2 Blood Components

The total blood volume in the human body consists of about 54.3% plasma, 45% erythrocytes (or red blood cells) and less than 0.7% buffy coat (leukocytes/white blood cells and thrombocytes/platelets). The main component, plasma, is the medium for excretory product and oxygen transportation. Plasma is primarily made up of water (approximately 90%). Blood plasma also includes dissolved proteins (7%), electrolytes or inorganic salts (0.5%), lipids (0.4-0.7%), glucose (0.07-0.1%), and amino acids (0.03-0.07%) [31]. The primary protein in plasma is albumin, and other dissolved proteins include globulins and fibrinogen. The electrolytes in plasma are mainly sodium ions (Na^{+}) and chloride ions (Cl^{-}), along with calcium ions (Ca^{2+}), Magnesium ions (Mg^{2+}) and hydrogen carbonate ions (HCO^{3-}).

There are three types of blood cells that make up approximately 45% of the total blood volume: red blood cells (erythrocytes), white blood cells (leukocytes) and platelets (thrombocytes). Only approximately 1% of blood cells are white blood cells and platelets, the rest are red blood cells. The life span of red blood cells is about 120 days [32], and red blood cells consist mainly of hemoglobin proteins, an iron-rich substance that is responsible for the red color of blood.

2.1.3 Skin Tissue Layer and Its Vasculature

Skin tissue contains arterioles, capillaries, venules, and interstitial fluid, and the later occupies significantly more volume of skin tissue than do the blood vessels. Skin tissue layers include the epidermis (upper layer), the dermis, and the subcutaneous, as illustrated in Figure 2.2. Approximately 15%-35% of the epidermis layer volume is devoted to interstitial fluid with no blood vessels. The dermis layer contains the arterioles, capillaries, venules, and approximately 40% interstitial fluid. The subcutaneous tissue layer includes fat storage, connective tissue and interstitial fluid, but less interstitial fluid than in the dermis. The subcutaneous layer also contains blood vessels [33]–[35].

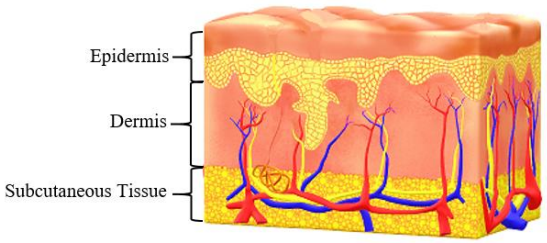


Figure 2.2: Skin tissue layers.

2.1.4 Protein and Glycated protein

Proteins are organic molecules made up of amino acids that play many vital roles in the human body. Amino acids are composed of amino group (NH_2), carboxyl group (COOH) and functional R group molecules (Figure 2.3). Polypeptide chains are formed by connecting amino acids together using peptide bonds (covalent chemical bonds) between the carboxyl group and the amino acid group molecules, Polypeptide chains have a three dimensional shape due to intramolecular interactions between the functional R groups.

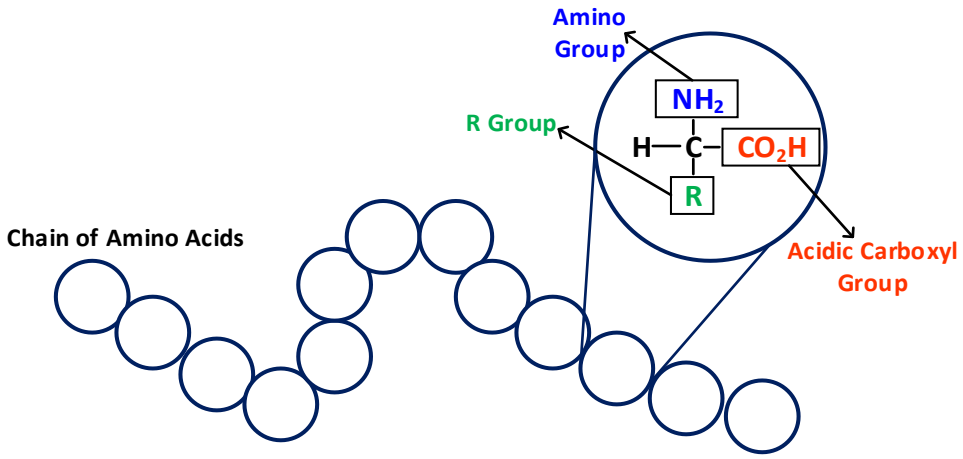


Figure 2.3: Amino Acid composition.

Recall that there are two main types of protein in the blood, albumin is the protein found in blood plasma and hemoglobin (Hb or Hgb) is the protein located in red blood cells. Both of these proteins have important physiological roles. Glycated hemoglobin is an indicator of mean blood glucose levels and diabetes mellitus. Therefore, it is important to examine the structure of both hemoglobin and glycated hemoglobin. Hemoglobin is predominantly made of HbA (about 97%), small amounts of HbA2 (about 2.5% of total) and a trace of fetal hemoglobin HbF (about 0.5% of total) [36], [37]. Figure 2.4 is an illustration of the structure of hemoglobin A (HbA) [37].

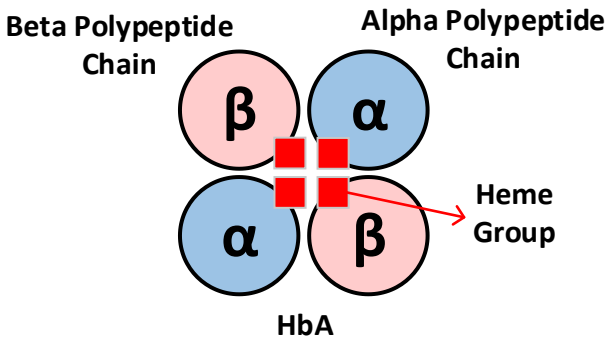


Figure 2.4: The Structure of Hemoglobin A (HbA).

HbA is composed of four polypeptide chains: two alpha polypeptide chains and two beta polypeptide chains [38]. Intermolecular interactions hold these subunits together. Each polypeptide chain has an iron containing heme group. Iron ions (Fe++) are at center of each heme group that binds the nitrogen atom from amino-acid residues in hemoglobin. Hemoglobin A (HbA) from red blood cells and albumin from plasma react with glucose in plasma to form glycated hemoglobin and glycated albumin as illustrated in Figure 2.5 [39].

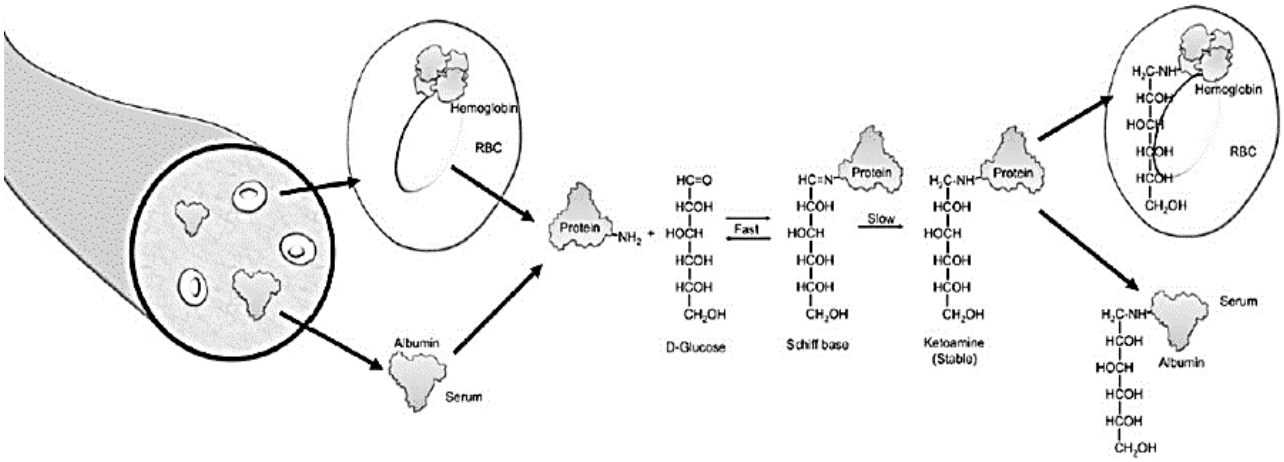


Figure 2.5: Interaction of glucose and protein to form glycated protein reprinted with permission [39].

Glycation is sometimes called non-enzymatic glycosylation. Glycation is the result of covalent bonding between glucose and protein amino groups without action of any enzymes. This reaction occurs between glucose and amino groups of a protein (NH_2) (N-end of the beta polypeptide chain in hemoglobin) to form an unstable Schiff base and H_2O. Schiff base is a compound with double bonds between the carbon atom of glucose and the nitrogen atom of lysine in hemoglobin. The Schiff base will change to a stable ketoamine by an Amadori rearrangement. The final product of glycation of proteins is known as advanced glycosylation end-products (AGEs), which are highly stable and irreversible. AGEs result from structural changes in Amadori

products through oxidation, dehydration and degradation [40]. The process of glycated protein can be explained in a concise manner by saying that carbohydrate moiety is added at the amino terminal of a protein.

Normally less than 7% of hemoglobin A (HbA) in red blood cells is glycated [37] to form glycated hemoglobin HbA1a, HbA1b, and HbA1c. Figure 2.6 illustrates the percentage of hemoglobin participating in the production of glycated hemoglobin. The collection of these products are denoted as HbA1 and are characterized as glycated hemoglobin, glycosylated hemoglobin, fast hemoglobin or glycohemoglobin [36]. The major component of HbA1 is HbA1c (70%-80% of HbA1) and HbA1c contains glucose [37]. HbA1a contains fructose-1,6-diphosphate (HbA1a1) and glusoce-6-phosphate (HbA1a2), and HbA1b contains unidentified carbohydrate [36].

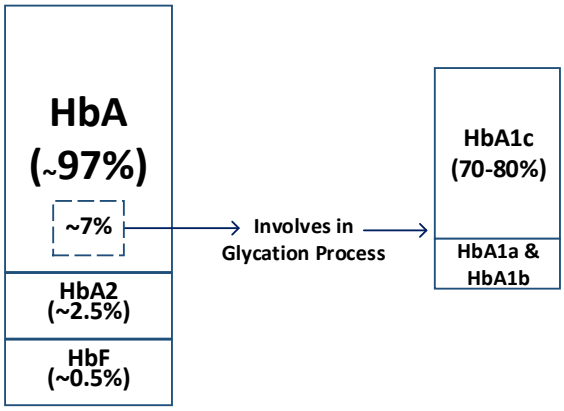


Figure 2.6: The percentage of hemoglobin involved in glycation and glycated products.

2.1.5 Correlation between Blood Glucose Concentration and Glycated Protein

The use of glycated HbA1c to detect glucose metabolism in patients with diabetes was first proposed by Ronald Koenig and coworkers in 1976 [41]. HbA1c is used widely for decades to diagnose patients with diabetes and to help manage diabetes. The American Diabetes Association (ADA) approved HbA1c as an indicator of diabetes when the value of HbA1c is $\geq$ 6.5% (48

mmol/mol), and as an indicator of high risk for future diabetes when the value of HbA1c is in the range between 5.7 and 6.4% (39-47 mmol/mol) [42].

The life span of red blood cells (which consist mainly of hemoglobin proteins) is about four months [32]. The rise in blood glucose levels is accompanied with a rise in glycated protein or HbA1c. Thus, since the breakdown of all red blood cells does not occur at the same time, the average life of blood glucose detected in HbA1c is about three months (i.e., glycated hemoglobin shows longitudinal exposure to glucose). Likewise, the measurement of HbA1c reflects the average plasma glucose concentration over a period of about three-months. In actuality, plasma glucose concentration fluctuates during the span of a day due to multiple factors such as illness, hormones, food ingestion, and exercise [39]. While daily blood glucose levels are associated with short-term changes, HbA1c measurements to quantify glucose levels reflect long-term effects (~three months). In a healthy person, 3-6 % of total hemoglobin is involved in the glycation process that produces HbA1c. This value increases to 6-12 % in patients with diabetes [41]. There are prediction models that estimate the average glucose levels in the body using HbA1c data. For example, a linear model was proposed in [43] estimates that a 1% increase in HbA1c corresponds to a 35 mg/dL increase in the average blood glucose level [44].

In addition to HbA1c, other glycated proteins that have been evaluated as effective alternatives for glycemic markers. These glycated proteins include glycated albumin, fructosamine, glycated apolipoprotein B, etc. Glycated albumin (GA) measurements reflect the average plasma glucose levels over a 2-4 week period and have the potential to act as a monthly management marker for diabetes [45]. This relative response based on changes in glucose levels reflected from glycated albumin measurements falls in between the measurements taken from an HbA1c test and the instantaneous measurement available from a blood glucose test. Fructosamine

measurements also reflect glucose changes over a period of 2-3 weeks. Fructosamine is produced from the glycation of all serum proteins (including albumin) [46]. Serum includes all of the blood plasma components (water, proteins and hormones) with the exception of fibrinogen and other clotting factors. Glycated low-density lipoprotein (LDLs) molecules (also found in plasma) is an example of a relatively short-term glycemic indicator (3-5 days). LDLs can undergo glycation with glucose by forming covalent bonds with the lysine of apolipoprotein B (the main apoprotein of LDL) [47].

Additional chemicals affect the level of HbA1c that are independent of glucose concentration. These chemicals are incorporated when translating HbA1c levels to glucose concentration values. Some of these factors are listed as below:

- Red blood cell lifespan: Changes in the survival of erythrocyte in red blood cells changes the level of HbA1c independent of blood glucose concentration. A lifespan of less than 120 days for red blood cells or an exposure of hemoglobin to blood glucose for less than 120 days results in a reduction in the level of HbA1c. Conditions that affect red blood cell lifespan include (1) blood loss (after surgery, due to frequent blood sampling or blood donation) leading to a decrease in HbA1c [48], and (2) patients with chronic renal disease. Cronic renal disease or chronic kidney disease is a health risk associated with diabetes. The worst case of kidney disease is when the kidney stops working which is called end-stage renal disease (ESRD). Chronic kidney disease causes a reduction in life span of red blood cells by 30%–70% and thus results in reduction of HbA1c [49]. Patients at this stage are recommended to undergo erythropoietin therapy to stimulate erythropoiesis. Erythropoietin therapy alters the proportion of old to young erythrocytes. Since young red blood cell have a lower rate of glycosylation compared to old red blood cells, a higher

proportion of young red blood cells compared to old blood cells reduces the amount of HbA1c [39].

- Anemia: There are several factors that can cause anemia, including a deficiency in iron, vitamin B12, and/or folate, and sickle cell disease (or other thalassemia) [50]. The risk of anemia is higher for a patient with diabetes, and most anemia cases are due to iron deficiency, insufficient iron intake, failure of iron absorption, pregnancy or loss of blood. Iron deficiency anemia elevates the HbA1c levels in diabetic individuals compared to individuals with normal blood glucose levels, especially in groups with a fasting glucose level between 100 and 126 mg/dL [51]. Anemia also leads to false diabetes readings due to elevated HbA1c. Iron deficiency associates with elevation of HbA1c even in patients without diabetes [48].
- Age and race: Research done in [52] showed that the level of HbA1c increase by 0.1% per decade in people with normal glucose tolerance and 0.07% in people with impaired glucose tolerance. HbA1c is higher in non-diabetic Blacks, Asians, and Latinos compared to white persons [53]. It was posed that non-Hispanic black people with normal glucose tolerance have 0.21% more HbA1c and non-Hispanic black people with impaired glucose tolerance have 0.35% more HbA1c compared to non-Hispanic whites [52].
- Temperature: Researcher in [54] have been studied on patients with type 2 diabetes over period of 5 years to investigate the effect of temperature on HbA1c level. They reported negative correlation between temperature and HbA1c level in Taiwanese patients especially patient under 65 years old with blood mass index less than 24. Several experiments were conducted on T1D children in different years to investigate the effect of

seasonal change on HbA1c level [55]. Results demonstrated lower level of HbA1c during summer months.

- Vitamin C, vitamin E supplementation: Vitamin E and C have been proposed as beneficial actor in glycaemic control in patient with diabetes since prevent from glycosylation of hemoglobin. It was posed that using vitamin E supplementation causes reduction in HbA1c level and also glucose level in T1D [56]. Oral supplementation of vitamin C along with metformin (diabetic medication) reduces the level of HbA1c and glucose level in T2D to maintain good glycemic control [57]. However, further evidence is needed to determine if vitamins supplementation changes analytical expression between glucose and HbA1c.

In conclusion, the monitoring of glucose levels in the body using HbA1c measurements requires: (1) awareness of the parameters (in addition to glucose) that influence the HbA1c level, and (2) measurement technique that demonstrates a clear selectivity to HbA1c. It is important to design a sensor that takes these parameters into account, and to come up with an analytical expression that improves the correlation of HbA1c to glucose levels.

2.1.6 Blood Oxygen Saturation

The normal range for the arterial oxygen level in the body is between 75 and 100 mmHg, and these levels are essential for healthy blood pressure levels and cardiopulmonary function. Red blood cells are responsible for the transportation of oxygen from the lungs to tissues throughout the body. Red blood cells contain hemoglobin and water (about 70% of red blood cell volume is water) [58]. The transportation of oxygen is mainly performed by hemoglobin molecules. Each hemoglobin molecule has four iron (Fe) atoms, and each iron atom forms partial covalent bonds with one oxygen molecule. The hemoglobin affinity for oxygen increases the solubility of oxygen in the blood [59]. Approximately 98% of the total oxygen carried in the blood is bound to

hemoglobin, and about 2% is dissolved in plasma and red blood cell water [58]. Oxygenated hemoglobin loses oxygen when it combines with hydrogen ions (resulting in deoxyhemoglobin). Oxygenated hemoglobin causes blood to have a bright red color, while deoxygenated hemoglobin is a dark red color. Oxygen rich blood flows away from the heart through arteries and arterioles, and oxygen diffuses out of capillaries and into every cell in the body. After this happens, deoxygenated blood returns to the heart through veins and venules. Venules are small blood vessels that connect the larger blood vessels or veins to the capillaries.

Oxygen saturation is the ability of hemoglobin to carry oxygen. A blood-oxygen saturation reading is be referred to as arterial blood oxygen saturation or SaO_2. SaO_2 is the ratio of the hemoglobin molecules in the arterial blood, which are saturated with oxygen (HbO_2), to the total hemoglobin in the blood [60]. The total hemoglobin is made up of functional and non-functional hemoglobin. The hemoglobin capable of carrying oxygen is characterized as functional and includes oxygenated hemoglobin (HbO2) and deoxygenated hemoglobin (HHb). The hemoglobin which is incapable of carrying oxygen is characterized as non-functional hemoglobin and includes carboxyhemoglobin (HbCO) and methemoglobin (METHb) [60]. Eq. (2.1) defines the blood oxygen saturation (SaO2) and is sometimes referred to as fractional SaO2.

$$Fractional\ SaO_2 = \frac{HbO_2}{HHb + HbO_2 + HbCO + METHb + other\ non\ functional\ hemoglobin} \qquad (2.1)$$

A pulse oximeter is a continuous non-invasive method used to measure the blood oxygen and is routinely used in clinical environments [59] [61]. When SaO2 is measured by pulse oximetry, it is expressed as SpO2. Measurements made using pulse oximeters do not include non-functional hemoglobin. When non-functioning hemoglobin is neglected, SpO2 is defined as the ratio of oxygenated hemoglobin (HbO2) to the combined oxygenated hemoglobin (HbO2) and

deoxygenated hemoglobin (HHb) [62]. Eq. (2.2) is sometimes denoted in the literature as functional SaO2 [63], [64] [62].

$$Functional\ SaO_2 = SpO2 = \frac{HbO_2}{HHb + HbO_2} \quad (2.2)$$

2.1.7 Correlation between Blood Glucose Concentration and Oxygen Saturation

Patients with diabetes develop physiological complications such as pathological changes in the coronary arteries of the heart, peripheral blood vessels, blood vessels in the brain, and the kidneys [65]. As these complications arise, diabetic patients may experience a reduction in blood oxygen saturation levels. Thus, these patients develop an inverse relationship between blood glucose levels and the blood oxygen saturation levels. Below is a brief explanation of the correlation between diabetes and low blood oxygen levels.

Diabetes is one of the main causes of chronic renal failure. Patients with severe end-stage renal disease or kidney failure are treated with dialysis or hemodialysis. Ninety percent of patients receiving dialysis treatment undergo hemodialysis, and these patients are at risk for developing cerebral disease complications such as low levels of cerebral regional oxygen saturation (rSO2). The level of rSO2 is significantly low in diabetic patients undergoing hemodialysis compared to non-diabetic patients undergoing hemodialysis [66]. As, the level of rSO2 is positively correlated with SpO2 or SaO2, it is concluded that there is low level of oxygen saturation in diabetic patients who suffer from kidney failure.

Diabetic patients who experience chronically elevated blood glucose levels can also be at risk for the development of atherosclerosis [67]. Atherosclerosis is a form of arteriosclerosis and is caused by fat build up or thickening of the inside of artery walls. This leads to restricted blood flow and a reduction in oxygen supply to tissues and poor oxygen permeability. Poor glucose control and hyperglycemia contribute to inflammation and are thus contributing factors for

atherosclerotic cardiovascular disease which is accompanied with insufficient blood oxygen due to atherosclerosis [65].

The correlation between obesity, Type 2 diabetes and blood oxygen saturation are explained in [68]–[70]. Excess calorie consumption causes an increase in the quantity of adipose tissue mass. This leads to obesity and local hypoxia [69]. Obesity is one of the main causes leading to Type 2 diabetes and hypoxic disease can be a possible pathogenic factor in Type 2 diabetes [71]. So, it can be concluded that low oxygen levels in the blood and glucose intolerance in Type 2 diabetes are interrelated.

Overproduction of endothelin-1 and fibronectin, and induction of apoptosis are common aspects in both hyperglycemia and hypoxia-induced phenomena. Hyperglycemia can induce cellular hypoxia thorough increased oxygen consumption in mitochondria [72]. Correlation of blood glucose concentration and oxygen saturation is due to a competition between glucose and oxygen solubility in the blood [61]. For instance, an increase in the blood glucose level reduces the oxygen solubility in the blood.

There are other factors that change blood oxygen saturation levels independently of glucose concentration. These include (1) an increase in carbon monoxide levels, (2) lower amounts of red blood cells and (3) hemoglobin and factors affecting hemoglobin's affinity for oxygen. The factors that increase hemoglobin's affinity for oxygen are (1) an increase in HbA1c [73], (2) an increase in blood pH, (3) a reduction in temperature, (4) a reduction in the partial pressure of carbon dioxide (PaCO2) and (5) a reduction of byproducts from glucose metabolism (glucose metabolism is the oxidation of glucose to carbon dioxide and water).

2.1.8 Overview of Intrinsic Glucose and Tissue/Blood Properties:

The intrinsic properties of glucose include (1) glucose absorption coefficient (μ_a), (2) glucose specific optical rotation (α), and (3) glucose Raman shift. These three properties are the basis for several glucose sensing techniques such as optical spectroscopy, optical polarimetry, and Raman spectroscopy. Optical spectroscopy sensors detect the interaction of light with glucose molecules and are capable of detecting changes in the absorption of light as a result of changes in glucose concentration. Likewise, optical polarimetric sensors measure glucose concentration based on the rotation angle of light and Raman spectroscopy is based on the Raman signal intensity. The values for glucose absorption coefficient, specific optical rotation and Raman signal intensity are also dependent on the wavelength of incident light. The measurement of these glucose properties as a function of wavelength are reported in the literature.[74], [75], [76] [77].

Additional non-invasive methods include scattering/occlusion spectroscopy and optical coherence tomography that measure glucose concentration based on the tissue scattering coefficient and blood refractive index of glucose. [4]. Bio-impedance spectroscopy and millimeter-wave/microwave glucose sensing are a function of the tissue permittivity and electrical conductivity. The plot of permittivity and conductivity vs wavelengths can be found in [78].

The seven non-invasive glucose sensing techniques introduced here will be explained in more detail below, especially how they relate to the physiological concepts described above.

2.2 Non-Invasive Methods for Glucose Measurements

The conventional finger-prick method with glucose strips and accompanying meter is the most reliable method for patient glucose self-monitoring, if we do not consider plasma glucose measurements used in clinical laboratories. The finger prick glucose meter has two essential components: a test strip coated with enzymes (e.g. glucose oxidase (GOx), glucose dehydrogenase

(GDH), and hexokinase (HK)) and a detector composed of proprietary electronics. When a drop of blood is applied to a test strip, the glucose within the blood sample reacts with the enzymes and the resulting electrochemical reaction produces a current signal which is linearly proportional to the glucose concentration[79]. It is well established that the finger prick method is a reliable method for accurate glucose measurements. However, consistent penetration of the skin is painful, inconvenient and carries a risk of infection. Non-invasive glucose measurement methods have the potential to (1) ease glucose detection with greater patient comfort, (2) increased real time glucose accuracy, and (3) more effective treatment options. Since the focus of this research is on optical non-invasive sensors, Sections 2.2.1 - 2.2.7 describe the operation of optical sensors discussed above.

The optical sensor is comprised of a light source illuminating the tissue and an optical transducer that converts the detected light into a measurable electrical signal. The optical system has two modes of operation (1) reflection mode, where both light source and the photo detector are placed on one side of tissue and (2) transmission mode, where the photo detector is placed on one side of the tissue and the light source is placed on the opposite side [80]. There are several locations in body close to the skin surface where blood vessel are easily accessible, among them, are fingers, ears, across the tongue, lip, forearm, and anterior chamber of the eye, etc. In order to understand glucose detection based on optical signals, it is necessary to understand how incident light interacts with human tissue. Incident light can either be absorbed, transmitted or scattered by the tissue. The type of interaction of light with tissue depends on the (1) wavelength of the incident light, (2) tissue structure, and (3) optical properties of the tissue (relative refractive index, absorption coefficient, and scattering coefficient) [81]. Figure 2.7 illustrates the mechanisms associated with interactions of light with atoms in tissue.

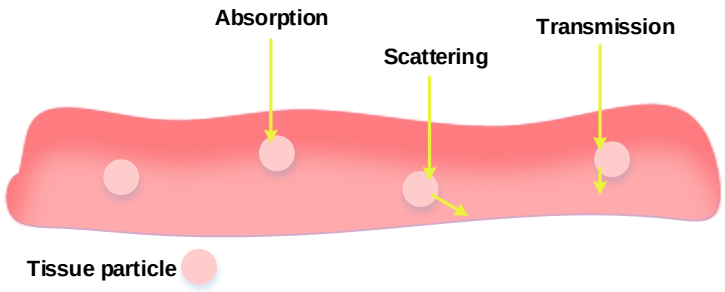


Figure 2.7: Interactions of light with tissue.

Absorption: Absorption is defined as the gain in electron energy from a low energy level to a higher energy level as a result of interaction with incoming light (or photon). Furthermore, in order for absorption to occur, the energy of the incident photon (hc/λ_1) should be equal to the difference in energy between the high (E2) and low energy (E1) levels of the electron ($E2 - E1 = hc/\lambda_1$). Depending to energy of photon (hc/λ_1), there can be different type of energy transition in a molecules including energy transition between vibrational states, rotational states, electronic states or etc.

Scattering: Absorption of light by a material can result in elastic or inelastic scattering of light. Energy of a photon can be absorbed by an electron causing excitation from low energy level (E1) to higher level of energy (E2). This electron can also lose energy by transitioning from a high energy level (E2) to low energy level (E1), and when this occurs, a photon is emitted ($hc/\lambda_2 = E2 - E1$). If this transition (high to low energy) is equal to the transition when the photon was absorbed (low to high energy), then the process is defined as elastic scattering of light (i.e. $\lambda_1 = \lambda_2$). Likewise, if the transition from high energy level to low energy level is less than or greater than (E2-E1), then this is defined as inelastic scattering (i.e. $\lambda_1 \neq \lambda_2$).

Elastic scattering includes both Rayleigh scattering and Mie scattering [81]. In Rayleigh scattering, the size of the particles involved in scattering, such as atoms or molecules, is much less than the wavelength of incident light (λ_1). In contrast, for Mie scattering, the size of the particles

(involved in scattering) is comparable to the wavelength of the incident light. Inelastic scattering including Raman scattering and fluorescence. In this case emitted light has different wavelength than incident light. The portion of inelastic scattering is negligible against of elastic scattering.

Transmission: A photon can pass through transparent or semitransparent tissue matter if there is no loss in its energy by any energy transition and scattering.

In next Subsections, the concept of glucose sensing techniques for plasma glucose quantification are discussed, related references using these methods are cited and the factors affecting glucose reading independently to glucose concentration are discussed. Non-invasive methods discussed next Subsections are mid-infrared and near infrared spectroscopy [7], polarimetry [12], Raman spectroscopy [11], occlusion/ scattering spectroscopy [82], optical coherence tomography [17], bio-impedance spectroscopy [83], and millimeter or microwave sensing [16]. These methods are involving in absorption, scattering and transmission of an electromagnetic radiation.

2.2.1 Mid-Infrared and Near Infrared Absorption Spectroscopy

Mid infrared and near infrared absorption spectroscopy are techniques used for probing a sample to find quantitative information about its constituents. The wavelength range for near infrared (NIR) is between 700 nm and 2,500 nm and between 2,500 nm and 25,000 nm for mid-infrared (MIR). The advantage of using the optical absorption method to measure glucose is the low cost associated with the light detector and light source measurement set-up, especially in the NIR range. This is due to the lower cost of the light source operating at lower wavelengths (NIR) compared to higher wavelengths (MIR). NIR light results in deeper penetration below the skin surface compared to MIR wavelengths. Sensing measurements using NIR can be operated in both reflection and transmission mode due to associated penetration depths of 0.5 mm or more. Since

MIR light cannot penetrate more than a few micrometers through tissue, MIR sensing methods are only operated in reflection mode [84].

The spectroscopy setup for MIR and NIR absorption includes a light source generating different wavelengths in the range of MIR or NIR and a photo detector in order to measure the intensity of light which is reflected or transmitted through the sample based on the light source used. Figure 2.8 includes a simple diagram illustrating the absorption of light through a sample consisting of a mixture of glucose and distilled water solution. The intensity of light that is transmitted and measured by the photo detector is a function of the concentration of absorbing molecules, the thickness of the sample, and the absorption coefficient. The absorption coefficient of absorbing molecules, (μ_a) depends on the wavelength of incident light and structure of the absorbing molecules (in this case glucose molecules). Different materials exhibit an absorption peak within a specific range of wavelengths. MIR and NIR absorption spectroscopy can measure the variation of light absorption as a function of wavelength, and can identify the wavelength of light with the highest glucose absorption. Targeting or identifying the wavelength of light with the highest glucose absorption will help to optimize the correlation between light absorption and glucose concentration.

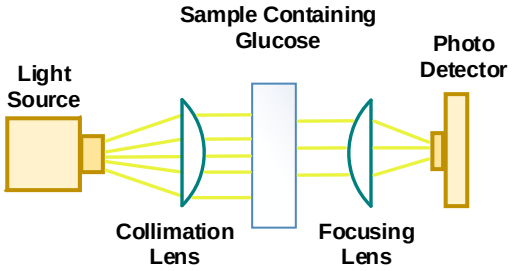


Figure 2.8: Simple schematic illustrating absorption spectroscopy (reprinted with permission) [4].

The concept of absorption spectroscopy can be understood based on the Beer-Lambert law of absorption (Eq. (2.3)) [85] [86].

$$I = I_0 . 10^{(-l.\varepsilon.c)} = I_0 . e^{(-l.\mu_a)} \quad (2.3)$$

where I_0 is the initial light intensity (W/cm^2), I is the intensity of light at any depth within the absorption medium (W/cm^2), l is the absorption depth within the medium (cm), ε is the glucose molar extinction coefficient or molar attenuation coefficient (L/(mmol.cm)), and c is the concentration of absorbing molecules in (mmol/L). The product of ε and c is proportional to the absorption coefficient (μ_a). This model assumes that the attenuation of light due to scattering is negligible in comparison to the light being absorbed. This equation can model the attenuation of light due to the absorption of light by only glucose molecules within a specified depth (l).

The concentration of blood glucose can be determined by detecting the intensity of light as it passes through the tissue layers. Blood glucose sensing is more complicated than glucose sensing through a sample of distilled water and glucose solution. Some of these issues and possible solutions are itemized below.

Absorption of light by water: Water in biological fluids absorbs a significant percentage of incident light, especially within the MIR range of wavelengths. When this occurs, the absorption of light by the water is independent and in addition to the absorption of light by the glucose molecules and therefore decreases the sensitivity to glucose molecules. A wavelength window should be identified that minimizes the absorption of light by water molecules and maximizes the absorption of light by the glucose molecules. There are two absorbance peak for water in the NIR range: one is placed between 1,350 nm and 1,520 nm, and the other is between 1,790 nm and 2,000 nm [87]. The NIR wavelength window between 1,500 nm and 1,850 nm and between 2,000 nm and 2,400 nm, where a relatively less amount of light absorption by water occurs, were investigated

for absorption-based glucose measurement [74], [88]. Multiple wavelengths of light can be used to eliminate the effect of signals coming from water in order to more easily identify the glucose signal. As an example, researcher in [89] used two discrete MIR wavelengths of 9,500 nm and 10,400 nm, where the 9,500 nm light was absorbed by both glucose and water, and the 10,400 nm was absorbed mostly by water. The differential method was then applied in order to subtract the absorption of water by both signals, resulting in a signal that mostly represents the absorption of glucose. The multiple wavelength approach for improving accuracy of glucose sensing is applied in a non-invasive glucose measurement device called TensorTip Combo Glucometer. TensorTip Combo Glucometer is designed by Cnoga Medical Ltd (Israel) and the concept approval started since 2006. The device is approved for use in numerous countries worldwide. TensorTip Combo Glucometer is capable to measure glucose in the range between 70 and 440 mg/dL. The device is comprised of four LEDs with wavelengths ranging from 600 nm to 1,000 nm and a color image sensor camera that photographs the transmitted light which passes through fingertip. Algorithms are applied in order to derive six-dimensional signal (position [x, y], time [t], color (red, green and blue)) which has correlation with blood glucose concentration. The performance of the sensor was investigated in [90], [91] by conducting a study on 14 healthy subjects, 6 T1D patients and 16 T2D patients. Based on consensus error grid, 100% of data are demonstrated to be in zone A (96.6%) and B (3.4%) [90].

Scattering of light by tissue: The scattering of light due to multiple tissue/blood components results in deviation from Beer-Lambert's Law of Absorption and results in measurement error [92]. In fact, both tissue and blood components cause light scattering which attenuates the intensity of measured light. Total attenuation of light depends on the total attenuation coefficient, μ_{total} (1/cm), which is the sum of the absorption coefficient of the absorbing species,

μ_a, and the reduced scattering coefficient of the scattering species, μ_s'. The attenuation coefficient represents how strongly light is attenuated by molecular species for a specific wavelength (Eq.(2.4)). The reduced scattering coefficient, μ_s' is a contribution of the scattering coefficient, μs and the anisotropy of light propagation in biological tissue, g, which is the average cosine of the scattering angle [93].

$$\mu_{total} = \mu_a + \mu_s' = \mu_a + \mu_s(1 - g) \qquad (2.4)$$

If g is close to 1, more light is scattered in the forward direction compared to backward scattering. The value of g for biological tissue is found to be between 0.65 and 0.95 [93].

It is important to reduce light scattering, and therefore, it is possible to select the wavelength of the light source so that the condition $\mu a > \mu_s'$ is satisfied (i.e., more absorption vs scattering of light due to tissue). NIR light results in more scattering by tissue compared to MIR light. However, NIR results in higher penetration depth through tissue, compared to MIR. This improves sensitivity to glucose molecules since light can penetrate in depth and face with more glucose molecules (within tissue and blood) compared to MIR light which only targets superficial glucose (within tissue).

Other measurement error sources: The measurement is sensitive to several factors including the amount of pressure applied at the probe/skin tissue interface, movement artifact, and interfering light noise and temperature. The experimental set-up can be designed to eliminate error by designing and constructing a stable probe and by using a modulated or pulsed light source that results in a high signal to noise ratio compared to using a constant light source [80] [94].

The effect of perturbing factors should be included in the algorithms which will help to model the relationship between light absorption and glucose concentration. Glucose quantitative data can be derived by applying the following analytical and calibration methods to absorbance vs

wavelength data: multiple wavelength linear regression, partial least squares, principal component regression, Deming regression [95], genetic algorithm and so on. These algorithms help to generate a model that accurately predicts the glucose concentration by considering the effect of multiple variables on the output. The algorithm will weigh the effect of multiple variables by minimizing and or eliminating the influence of competing signals on the real value of the glucose concentration. Finally, the effectiveness of the algorithm can be determined by using evaluation criteria by comparing the estimated glucose concentration with the known value of glucose concentration.

2.2.2 Optical Polarimetry

Polarimetry measures the angle of rotation of linearly polarized light as a result of its interaction as it passes through an optically active solution. Figure 2.9 is a simple schematic of a polarimeter which is comprised of a (1) light source, (2) linear polarizer, (3) polarization analyzer, and (3) photo detector.

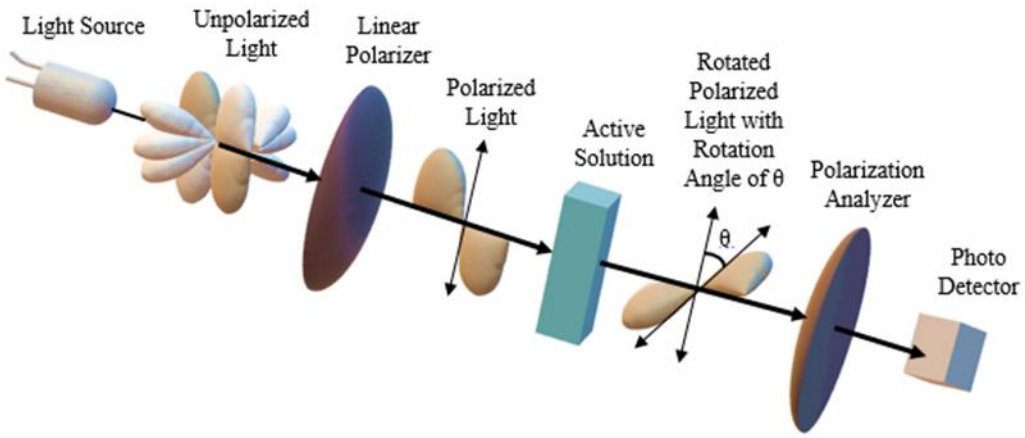


Figure 2.9: Simple schematic illustrating polarimetry

The angle of rotation of the electric field depends on the concentration of glucose in the optically active aqueous solution which is modeled by Eq.(2.5) [96].

$$\theta = \alpha_{\lambda}^{T} l\, c \qquad (2.5)$$

where θ is the measured angle of rotation (degrees) of the electric field, α_{λ}^{T} is the specific rotation for the active substance (degrees mL)/dm g), l is the optical path length (dm), and c is the concentration of the active substance (g/mL). The value of the specific rotation, α_{λ}^{T}, depends on the wavelength of the light source and the temperature of the sample. The polarizer is also used as a polarization analyzer to determine the plane of the polarized light after it passes through the sample. When the axis of polarization in the analyzer matches the angle of rotation θ of the electric field, then maximum intensity of light will be detected by the photo detector. On the other hand, the photo detector will not detect light when the polarization axis of the analyzer is perpendicular to the angle of rotation of the electric field.

The measurement of the angle of rotation can be accomplished using two methods. The first is by measuring the amplitude of detected light intensity which is function of rotation angle (θ) as already mentioned in the previous paragraph. The second method is by measuring the phase shift between two detected light beams that are generated by a beam splitter. In the latter, a beam splitter is placed just before the sample and divides the light source into two beams of light. One of beams passes through the sample while the other is reflected in different direction. The two beams are collected by detectors, and the phase shift between the two light beams as measured by the detectors is dependent on the angle of rotation (θ) [94]. Reference [97] designed a polarimetric system based on the phase shift method. This system is composed of a (1) helium neon laser light, (2) rotating linear polarizer, (3) two stationary linear polarizers, (4) two analyzers and (5) two detectors. They reported a direct linear correlation between the glucose concentration and the angle of rotation (in the range of millidegrees) for a vitro glucose test cell with a path length of 1 cm. The experiment was repeated with similar results for an excised human eye.

The anterior chamber of the eye is a preferable anatomic location for applying the polarimetric technique compared to using skin tissue. The eye contains a clear optical medium called the humour aqueous of the eye that is ideal for measuring the glucose concentration using visible light, which is harmless to the human eye. In contrast, skin tissue exhibits higher scattering which results in complete depolarization of light. Thus, glucose concentration measurements within skin tissue cannot be measured accurately as a result of the low signal to noise ratio. Issues regarding polarimetry of eye are variation in angle of rotation due to (1) variation in birefringence of the cornea between individuals [97] [98], and (2) variation in optically active components in the aqueous humour such as albumin and ascorbic acid [99] [100].

A disadvantage of using the polarimetric method for glucose monitoring of humour aqueous is the physiological lag time (below ten minutes) between the peak of glucose in humour aqueous and the associated glucose peak measured in blood plasma, that prevent from real time measurement [101]. The lag time is in the range of 2.9 to 5.4 min based on measurements of glucose concentration within the anterior chamber of the eye for New Zealand white rabbits [102]. Faraday-based polarimeters with red light sources were used to measure the angle of rotation with a sensitivity below 0.4 millidegrees, which corresponds to less than 10 mg/dL of glucose concentration. Furthermore, polarimetric system should be designed in a handy miniaturized form for user friendly self-monitoring.

2.2.3 Raman Spectroscopy

Raman spectroscopy setup comprises of a high intensity light source and very sensitive Raman spectrum photo detector. Figure 2.10 is a simple schematic representation of the functional components of Raman spectroscopy.

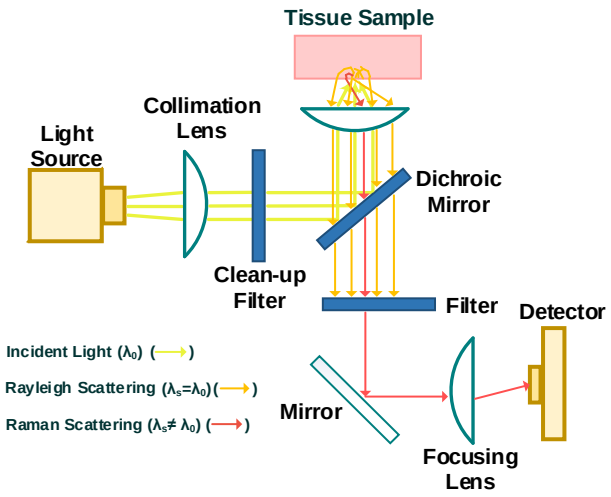


Figure 2.10: A simplified schematic illustrating Raman spectroscopy (reprinted with permission) [4].

When light with a monochromatic frequency interacts with matter, a portion of the light is scattered. The scattering is mostly elastic, with a very small percentage of inelastic scattering that results in multiple wavelengths (i.e. frequencies) compared to the incident wavelength of light [103]. Inelastic scattering of light is also called Raman scattering or is referred to as the Raman effect based on observations by Chandrasekhara Venkata Raman, who observed inelastic scattering of light experimentally for first time in 1928. The peak height or intensity of the Raman spectrum relates to a particular substance concentration. The shift in the frequency of the scattered light depends on the type of molecules or the chemical structure of the samples, and is independent of the wavelength of the light source. It was measured that Raman fingerprints of glucose are placed at 911 cm^{-1}, 1060 cm^{-1} and 1125 cm^{-1}, with the highest intensity Raman signal at 1125 cm^{-1} [104]. Raman shift is calculated as Eq.(2.6) where l_0 is wavelength of incident light and l_s is wavelength of Raman scattered light.

$$Raman\ shift\ (cm^{-1}) = \frac{10^7(nm)}{(cm)} \times \left(\frac{1}{l_0(nm)} - \frac{1}{l_s(nm)}\right) \qquad (2.6)$$

The vibrational modes for glucose molecules ($C_6H_{12}O_6$) are associated with C-O, C-C and C-H stretching bonds and are observed between 800 cm^{-1} and 1200 cm^{-1} for C-O and C-C and around 2900 cm^{-1} for C-H [105]. Thus, typical glucose Raman fingerprints are observed at 911 cm^{-1}, 1060 cm^{-1} and 1125 cm^{-1}, with the highest intensity Raman signal at 1125 cm^{-1} [11].

Non-invasive glucose Raman-based detection is possible with a calibration stability of at least 10 days [11]. The measurement set up for this system includes an 830 nm light source irradiating at 250 μm below the skin surface (targeting the interstitial fluid region) of the thumb in 35 patients. It was demonstrated that 93% of measured data points were placed in the region of A+B using consensus error grid analysis.

Some issues associated with the measurement of glucose concentration using Raman spectroscopy are background fluorescence signal due to presence of protein, body water content and variation in tissue between individuals. Skin tissue characteristics vary between each individual and tissue characteristics may affect the measured intensity of the Raman fingerprint of glucose.

Unlike NIR and MIR spectroscopy, the presence of water has minimal effect on Raman spectroscopy measurements of glucose aqueous solutions because the Raman spectrum of water has a weak cross Section. However, the amount of water content in the glucose solution can be estimated using the Raman fingerprint of water. The large Raman shift due to the OH stretching mode of water at 3400 cm^{-1} can be used to estimate the water content [106]. Nearly all proteins respond to incident light by emitting a fluorescence signal that is the result of a series of electron transitions between two singlet states [107], where the wavelength of the fluorescence signal is longer than the wavelength of the incident photon. The presence of blood proteins in a glucose sample produces a background fluorescence signal which contaminates the glucose Raman

spectrum since the intensity of fluorescence signal is equal to or larger than the Raman signal [85]. A possible solution is to measure the glucose concentration within the anterior chamber of the eye (where less protein exists) versus within the skin tissue.

2.2.4 Occlusion/Scattering spectroscopy

Scattering spectroscopy can be used to measure the glucose concentration based on the scattering property of light in the tissue. Figure 2.11 includes a schematic of a scattering spectroscopy set-up with a red or near infrared light source and a photo detector array illustrating the scattering behavior of a tissue sample at (a) low glucose concentration and (b) high glucose concentration [4]. In the case of the low glucose concentration, Figure 8 demonstrates that the scattering angle exceeds that of the high glucose concentration sample, and thus the intensity of the scattered light at the detector for the low concentration sample is less than that of the high glucose concentration sample.

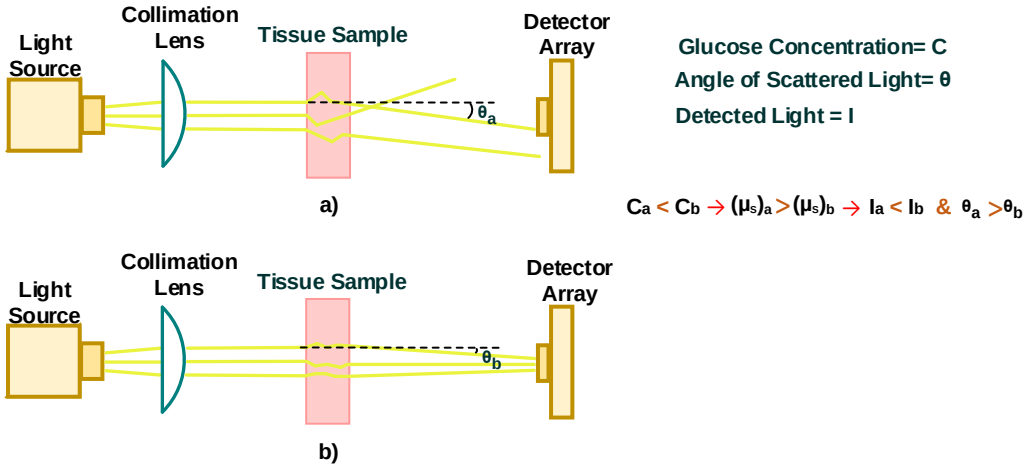


Figure 2.11: A simplified schematic illustrating scattering spectroscopy of a tissue sample with a (a) low glucose concentration (b) high glucose concentration (reprinted with permission) [4].

The glucose concentration affects the scattering behavior of tissue, and the reduced scattering coefficient of tissue (μ_s') depends on glucose concentration [108], [109]. As the blood

glucose concentration increases, the reduced scattering coefficient of tissue μ_s' decreases. The reduction of the reduced scattering coefficient of tissue means that when blood glucose increases, there is (1) a resulting increase in the intensity of detected light passing through the tissue [82], and (2) the angle of scattered light decreases. Beer's law states that light attenuation for light scattered by a sample is proportional to the negative exponent of the total attenuation coefficient (μ_{tot}=μ_a+μ_s'). So, a decrease in the reduced scattering coefficient results in a higher light intensity at a particular depth in the sample relative to the initial incident light intensity. The reduced scattering coefficient of tissue and glucose molecules is greater than the absorption coefficient of tissue and glucose molecules as the wavelength decreases from MIR to NIR range. Regarding the second observation, the angle of scattered light is ranged between 0° to 10° and decreases in values when the glucose concentration increases due to a shorter mean free path between glucose molecules [110]. Also, applying over systolic pressure at tissue site to occlude the blood flow causes increase in magnitude of intensity of detected signal due to erythrocyte aggregation. Scattered pattern at tissue location can be monitored with more sensitivity to glucose molecules [82].

The OrSense's NBM-200G device is an example of a commercial non-invasive glucose sensor based on occlusion spectroscopy. The testing of the sensor on the fingertips of 12 T1D and 11 T2D patients resulted in 95.5% of the measurement data within the clinically acceptable A (69.7%) and B (25.7%) regions of the Clarke error grid analysis chart [82]. Although the OrSense has a CE safety designation, it is not commercially available at this time.

In this method as with other methods, physiological differences between individuals affect the accurate interpretation of the results. Variations in tissue component compositions, such as free fatty acid concentrations and intrinsic erythrocyte aggregation, affect the scattering of light in the

tissue measurement site. Variations in oxygen saturation between individuals and the level of light absorption due to hemoglobin concentrations also make it difficult to accurately interpret results. The effect of these interfering measurements can be minimized by using light sources with multiple wavelengths and by applying a suitable and sophisticated algorithm to extract glucose information from multispectral data [4].

2.2.5 Optical Coherence Tomography

The optical coherence tomography (OCT) system uses low coherence light within red or NIR range (with coherence length between 10 μm and 15 μm) to measure an interferometric signal. Figure 2.12 is a schematic of the OCT system illustrating the reference arm, sample arm, interferometer optics and photo detector/camera set-up. A light beam is split into two beams, one of which is backscattered from the tissue sample in a sample arm and another beam is reflected back from a moving mirror in a reference arm [93], [111]. The combination of the reflected light and backscattered light results in the interferometric signal which is detected by the photo detector. The peak intensity is dependent on the glucose concentration at different depths (up to 1 mm).

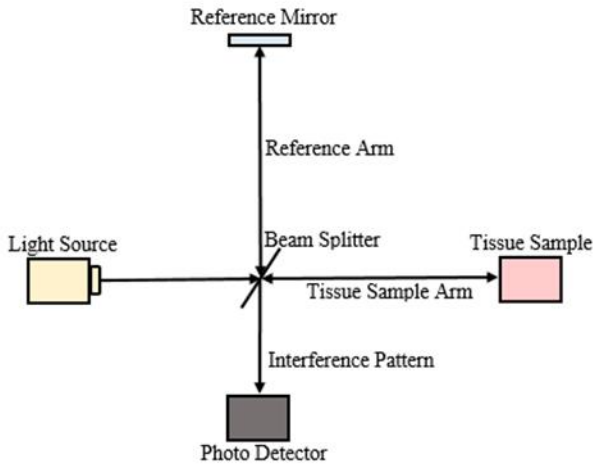


Figure 2.12: Simple diagram of set up for optical coherence tomography of tissue.

OCT signal can be measured at specific depth of tissue layer without contamination by interfering signal from other tissue layer. Using second moving mirror into the tissue sample arm allow scanning of probing beam laterally over the tissue surface, so there will be two dimensional images in both lateral and in-depth [112], [113].

The relationship between the intensity of the OCT signal and the glucose concentration can be modeled by Eq.(2.7), where the square of the intensity is related to the tissue scattering coefficient [93], [114].

$$I^2(l) = rI_0^2 \frac{\mu_b(l)}{4\pi} Le^{-2ln\mu_s} \qquad (2.7)$$

where r is the reference reflection coefficient, I_0 is the intensity of the incident light, l is the temporal coherence length of the incident light, n is the mean refractive index of the tissue (≈1.38), and l is the penetration depth. The parameter μ_s is the wavelength dependent scattering coefficient, and the parameter μ_b is the backscattering coefficient after light has penetrated the tissue a depth equal to l.

The OCT signal intensity is mostly a function of distance from the surface of the skin. The slope of a straight line fitted to the OCT signal depth profile depends on glucose concentration. The correlation coefficient between glucose concentration and the OCT signal slope varies periodically (with a period of 100 – 150 μm) between -0.9 to 0.9 depending on the depth where the tissue layer is scanned [115]. The period of change in the correlation coefficient is in agree with distance between neighbor collagen bundles within the skin [115]. The OCT measurement on 15 healthy subjects demonstrated that the OCT signal slope decreases up to 2.8% per 10 mg/dL increase in plasma glucose concentration when the slope of the OCT signal is measured at a depth between 200 μm and 600 μm from the skin surface [111]. In vivo experiments on farm pigs show

the maximum correlation between glucose concentration and the OCT signal slope at the papillary-reticular and dermis-hypodermis junctions [116].

There are several factors that affect the accuracy of the glucose OCT measurement, such as patient-related motion artifacts, temperature fluctuation (more than ±1℃), tissue heterogeneity, serum albumin and lactic acid. In addition, changes in the concentration of osmolytes in the body (such as KCl, Urea and NaCl, the major blood osmolyte components) can also have an effect on the scattering coefficient of the tissue. Although the intensity of OCT signal is sensitive to concentrations of these components, which can vary between individuals, the effect of glucose on the signal is still dominant [112], [117]. A disadvantage of the OCT method is that there is a lag time associated with the glucose measurement coming from the interstitial fluid (using OCT method) compared to the actual glucose level in the blood [112].

2.2.6 Bio-impedance Spectroscopy

Bio-impedance spectroscopy employs the measurement of impedance levels within tissue using a small AC current signal with a frequency lower below 1 MHz. Biological tissue can be modeled as an electric circuit of resistors and capacitors. Capacitance and resistance originate from the cellular membrane structure and the body water fluid (intra and extra cellular fluid), respectively [118]. Cell membranes are semipermeable and separate intracellular spaces from extracellular spaces. A simplest electrical model for tissue is a parallel combination of a conductor and capacitor, however, more realistic tissue electrical models have been proposed [119]–[121]. Conductivity of tissue is related to movements of ions within the biological fluid and permittivity is related to the tissue's ability to store charge or rotate molecular dipoles in the presence of an electric field [119], [122]. Eq. (2.8) describes the impedance of tissue based on the simplest electrical model as a function of tissue conductivity σ and permittivity ε.

$$I^2(l) = rI_0^2 \frac{\mu_b(l)}{4\pi} Le^{-2ln\mu_s} \tag{2.8}$$

In Eq. (2.8), G and C represent conductance and capacitance of the equivalent model. Parameters d and A are the thickness and cross-sectional area of the tissue sample, and ω is the angular frequency of the applied signal to the tissue. The impedance of tissue depends on the frequency of the applied signal. At low frequency, the current signal flows only through the extracellular fluid and contributes to the conductance portion of the tissue impedance. However, at high frequency, the current flows through both the intra and extracellular fluid and penetrates the cellular membrane, thus contributing to both the conductance and capacitance part of the tissue impedance [118].

Figure 2.13 illustrates the physiological components that play a role in tissue impedance behavior. Tissue impedance depends on the electric characteristics of the cell membranes. Since RBCs suspended in the blood occupy around 45% of the blood volume, the RBC membrane has an important role in the total tissue impedance [123], and has an effect on the capacitance value in the equivalent circuit of the tissue.

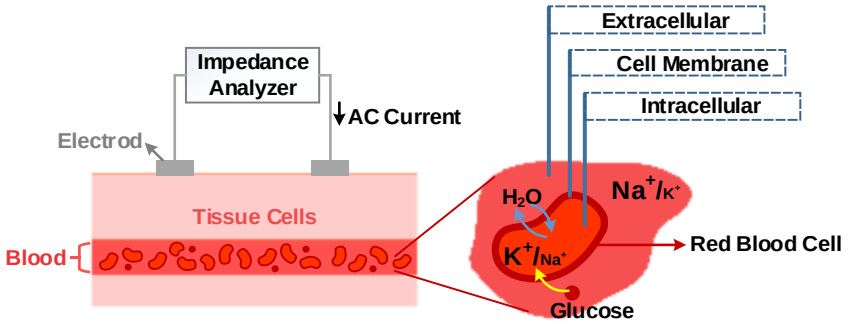


Figure 2.13: A simplified schematic illustrating tissue impedance spectroscopy (reprinted with permission) [4].

When the blood glucose concentration increases, the serum osmolality increases which results in the movement of water out of the cell and into the extracellular spaces through the cell

membranes. This causes the sodium ion [Na+] levels in extracellular spaces to decreases due to the dilution (see Figure 2.13). Cellular dehydration promotes redistribution of potassium ions [K+] from the intracellular to the extracellular spaces, resulting in an increase in the serum [K+] ion level in extracellular spaces [4], [123]–[125].

As the [K+] and [Na+] ions are balanced, there are associated changes to the permittivity and conductivity of surrounded medium and cell membranes, including the RBC membrane. These activities result in changes to the tissue impedance. According to a study conducted by Li et al., the permittivity and conductivity of an aqueous solution decreases when the glucose concentration increases for a frequency range between 1 kHz and 1 MHz [119]. Based on Eq. (2.8), the magnitude of the impedance increases with an increase in the glucose concentration.

There are differences in tissue components and red blood cell morphology between individuals. This results in different tissue impedance among people regardless of blood glucose level and so, affect glucose measurements while using the bio-impedance spectroscopy method. A possible solution to improve accuracy of glucose reading is combination of different glucose sensing methods. The combination of two methods can compensate the error caused by each method separately and thus minimize the effect of confounding factors. A combination of (1) scattering spectroscopy at three wavelengths (850 nm, 950 nm and 1,300 nm) and (2) impedance spectroscopy in the frequency range between 10-76 kHz was proposed in [126]. Blood glucose levels of ten volunteers was estimated based on two methods separately and then the combined results were obtained using an artificial neural network algorithm in order to improve the accuracy of the glucose estimation. In this research the accuracy of 100% was achieved based on Clarke error grid analysis. The approach of using multiple technologies was employed in a non-invasive glucose sensor device named GlucoTrack which was developed by Integrity Applications

Company (Israel). GlucoTrack device is based on combination of three technology of ultrasound, thermal and impedance spectroscopy. The performance of GlucoTrack device was evaluated by targeting the ear lobe of 91 diabetes in [127]. It was shown that 96% of the glucose reading fell in the clinically accepted A and B zone using Clarke error grid analysis.

2.2.7 Millimeter Wave/Microwave/Ultra-High Frequency wave sensing

Millimeter waves, microwaves and ultra-high frequency waves are electromagnetic (EM) radiation with a frequency range between 30 and 300 GHz; 3 and 30 GHz; and 300 MHz and 3 GHz, respectively [128]. When EM radiation is applied to tissue, some portion of the EM radiation energy is reflected back from the tissue surface. The rest of EM radiation energy is transmitted through the surface toward underlying layers in which some portion of transmitted signal is absorbed. Absorption, reflection and transmission of EM waves through the tissue depends on the dielectric characteristics, which depends on the glucose concentration. Reflected or transmitted EM radiation is sensed using a single-port or two-port EM wave sensor, respectively. The measured signal is analyzed using a vector network analyzer (VNA) and computer software that extracts glucose concentration information. A simplified schematic for a reflection mode EM measurement device is shown in Figure 2.14 [4].

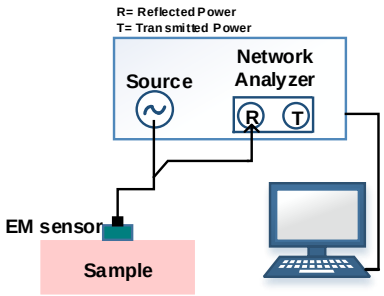


Figure 2.14: A simplified schematic illustrating a reflection mode EM measurement (reprinted with permission) [4].

The permittivity and conductivity of the tissue depends on the frequency of the applied EM signal and the glucose concentration [129]–[131]. The relationship between the dielectric properties of tissue mimicking phantom and the frequency of the applied EM signal was investigated at a frequency range between 0.3-20 GHz by Yilmaz et al. [130]. The relative permittivity measurement shows a decreasing trend (from approximately 68 unit to 22 unit), and the conductivity measurement demonstrates an increasing trend (from approximately 2 S/m to 28 S/m) due to an increase in the frequency. An overall increasing trend in the conductivity and a decreasing trend in the permittivity was also observed by Gabriel et al. over an increase in frequency between 10 Hz and 20 GHz for diverse types of human and animal tissue (liver, muscle, skin, tongue, adipose tissue) measured at body temperature [132].

Dependency of glucose concentration to relative permittivity was investigated by Jang et al. [129]. In this research a negative relation between relative permittivity and glucose concentration in a solution was demonstrated over frequency range between 1 GHz to 15 GHz. At a single frequency of 2.9 GHz, increase in glucose concentration from 0 to 400 mg/dL results in a total reduction of 0.2 unit in relative permittivity [129]. Negative relation between glucose concentration and relative permittivity have also been observed in [130], [131].

Measurement of the dielectric properties of a material are a function of the resonant or non-resonant response. The resonant response methods are classified primarily as resonator or resonant perturbation types [128], [133]. A change in the glucose concentration will cause a change in the dielectric properties of the tissue and thus impedance, which is the input into the sensor circuit. Sensors include micro-strip patch antennas, split-ring resonators [129], [134], [135], patch resonators [130] and dielectric resonator antennas. Electromagnetic sensors measure the narrowband dielectric properties of tissue based on the resonant frequency. The equivalent circuit

model for the sensor is a resonant circuit oscillating at resonance frequency. The resonant frequency of the sensor can be observed by measuring scattering parameters vs frequency using a vector network analyzer (VNA). The scattering parameters are the reflection coefficient (S11) and the transmission coefficient (S21). The output from the sensor circuit will result in (1) a shift in the resonant frequency of the scattering parameters, (2) a change in the amplitude of the scattering parameters at the resonant frequency, and (3) a change in the 3 dB bandwidths at the resonance frequency. These output measurements reflect changes in the glucose concentration that correlate to the dielectric properties of tissue.

The non-resonant response techniques include the free space method, the transmission line method, and the open-ended coaxial probe method [133], [136], [137]. These non-resonant methods use sensors to detect broadband dielectric properties which depend to glucose concentration in tissue [137]. The coaxial line sensor [138], open ended coaxial line sensor [139], waveguide-based sensor [140] and ultra-wideband antennas [16] are used to measure dielectric properties of tissue, and as with the resonant techniques, a VNA is used to measure the tissue scattering parameters as a function of frequency. The scattering parameters can then be converted to dielectric properties of tissue using the software embedded in the VNA or by using conversion methods such as the Nicolson-Ross-Weir method or the NIST iterative conversion method [136], [137], [141]. Once the dielectric property data is collected, it is calibrated to the glucose concentration.

2.3 Discussion: Physiological Factors and Existing Non-Invasive Glucose Sensors

This Chapter represents a discussion of the various physiological factors and existing non-invasive glucose sensors and techniques. Interdisciplinary knowledge about engineering experimental disciplines and glucose-related physiological theory are instrumental for the design

of an accurate non-invasive glucose sensor. Experimental and physiological factors affect the accuracy of glucose readings for each of the non-invasive techniques. Identification of the causes of these errors require appropriate solutions. One of the common contributions to errors are the differences in morphology and/or concentration of blood and/or tissue components, such as water, protein (albumin and hemoglobin), and fatty tissue. For example, a variation in the concentration of red blood cells can influence the accuracy of glucose readings for bio-impedance spectroscopy and millimeter-wave/microwave sensing. In some instances, the integration of multiple technologies (optical or/and electrical based) into a single platform has the potential to improve the accuracy of glucose measurements. This can affect the complexity of the design, processing algorithms, device size, cost, etc. In this dissertation, the goal is to increase the sensitivity to glucose, and thus the accuracy, by designing an optical sensor that supports multiple wavelength measurements compared to using single wavelength based glucose sensor.

An optical sensor has two modes of operation, transmission and reflection mode. In this dissertation the transmission mode is used, where the detector is placed on the opposite side of the light source. In transmission mode, the light passes through the sample before being entering the detector, and the interaction between the transmitted light and the glucose molecules is greater compared to reflection mode. The enhanced interaction results in an increase in the sensitivity of the sensor to track changes in glucose concentration. The other advantage is that, in transmission mode, the photons penetrate both tissue and blood vessels. There is a lag time between the glucose levels in the tissue compared to the glucose in the blood. Sensors working in reflection mode concentrate on surface interactions with tissue, and this means that this type of sensor cannot measure real time glucose levels in the blood.

There are several attempts to measure glucose within physiological fluids, such as interstitial fluid, salvia and aqueous humour of the eye. As mentioned before, there is an unavoidable lag time associated with glucose in the blood vs. the level of glucose in the physiological fluids. The lag time differs between individuals and further work is required to collect and report the lag times and how they vary for each subject [4]. In this dissertation, a wearable sensor on the fingertip is envisioned that relies on the measurement of glucose within highly dense vessels for real time glucose measurements.

Chapter 3 describes the preliminary study of the design and development of a transmission mode optical sensor that includes a single light source and detector. The feasibility of glucose measurements within aqueous solutions was explored based on the measured intensity data of a light source with a single wavelength. This preliminary study provided insight regarding how to improve the sensor parameters, structure and operation in order to increase its stability and accuracy of glucose readings. The preferred solution includes a sensor with a light source with multiple wavelength measurements and this design is described in Chapter 4.

3. Preliminary Design of a Non-Invasive Glucose Optical Sensor with a Single Wavelength Light Source

This Chapter explains the design considerations for the preliminary design of an optical based glucose sensor with a single wavelength light source (the design later evolves into a multi-wavelength light source design). The suitable wavelength range for glucose measurements is among three ranges (visible, near infrared and mid infrared) based on the literature. The experimental design, data collection and data calibration are described. The results associated with this preliminary design demonstrate a dependency of light intensity to the glucose concentration in solution. Finally, the performance of the preliminary design is analyzed and recommendations/conclusions are presented.

3.1 Glucose Spectra

A preliminary investigation to understand the basic glucose spectrum is first performed utilizing commercial instruments. Several instruments are used to examine the absorption and scattering behavior of light interacting with glucose molecules within the range of ultraviolet (UV), near infrared (NIR) and mid infrared (MIR) light. Glucose spectra is presented for the different wavelength ranges and finally, the optimal wavelength range most suitable for glucose measurements is identified based on the measurement results and based on some information from the literature.

To illustrate the glucose absorbance in the UV range, an Ocean Optics with Flame-Sony spectrometer is used to measure the absorbance of glucose within the range between 200 nm to 575 nm at room temperature of (20 $°C$). In the Ocean View software, the absorbance is calculated based on Eq. (3.1).

$$A = -\log\left(\frac{I - I_s}{I_0 - I_s}\right) \tag{3.1}$$

where I_0 is the intensity of detected light as it passes through the reference sample, which is only distilled water in a cuvette. Parameter I is the intensity of the detected light as it passes through the solution of distilled water and glucose, and parameter I_s is the intensity of stray light. Stray light is background light noises that are detected even when the light sources are off. Stray light limits the dynamic range of absorbance in the absorption band. The destructive effect of I_s is dominant when the intensity of the light (I) from the sample is very small (in the other words when absorbance from the sample is high). In this case, the stray light comprises a larger fraction of the total light that is detected by the detector. Therefore, since the value of the signal measured by the detector is very small, it is difficult to measure the true value of absorbance by the sample. To compensate for the effect of stray light, the intensity I_s is subtracted from both I and I_0 in Eq. (3.1). Background noise entering the detector can also be reduced by using a light source that emits light in a limited wavelength band, one that has a sharp response of intensity as a function of wavelength. An optical band pass filter is used at the entrance of the detector to allow only a signal within the range of interest to strike the detector, shielding, etc.

Figure 3.1 shows light absorption spectra from three solutions, a 450 mg/dL glucose solution (red curve), a 1000 mg/dL glucose solution (green curve), and a 450 mg/dL albumin solution (black curve). The albumin solution is made up of distilled water and egg white albumin powder with intermediate purity (albumin, laboratory grade, 100 g from CAROLINA). Albumin is a class of water soluble proteins, is also found in eggs whites, and has a similar amino acid content as human serum albumin. Albumin absorption is measured just to give an example of a component within the blood that is UV active. The resulting spectra shows no difference in absorbance between the 450 and 1000 mg/dL glucose solutions within the UV range. Thus, a light

source in the UV range is not optimal, and the possible reasons for this observation is explained in the next paragraph.

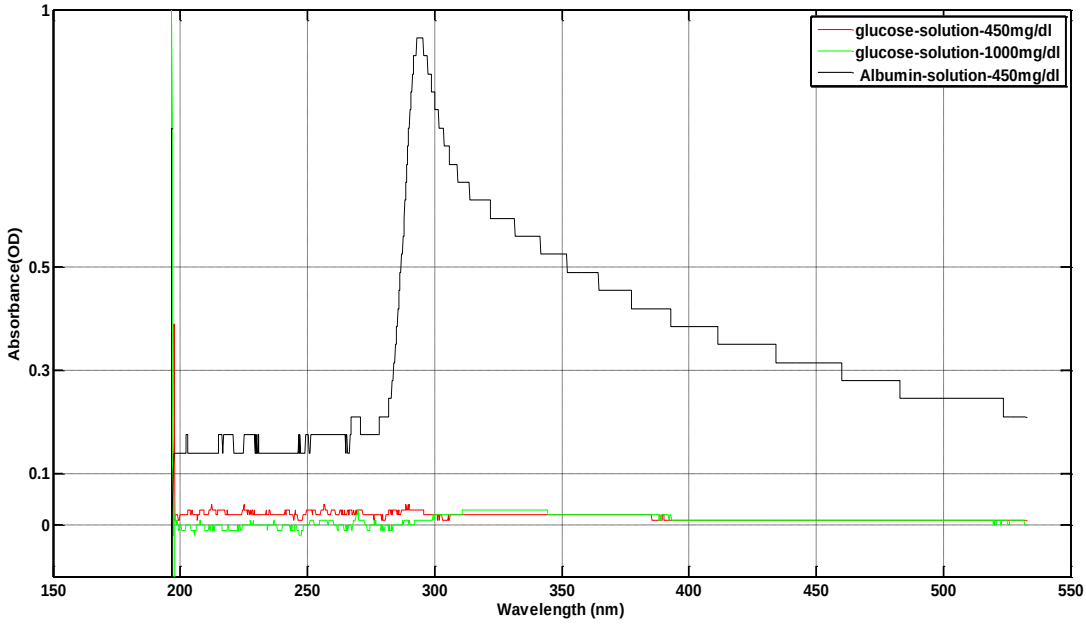


Figure 3.1: UV absorption spectra for glucose and albumin solutions.

The Ocean Optics spectrometer is equipped with a detector which is a Sony ILX511B linear silicon CCD array. The number of active pixels in this detector array is 2048, each pixel size is 14 µm × 200 µm, and the sensitivity is 75 photons/count at 400 nm and 41 photons/count at 600 nm. This detector array is used to measure absorbance in the unit of optical density (OD) and the ideal range for absorbance is between 0.5 OD to 1 OD. Working in this range results in a maximum sensitivity to changes in absorbance and a better signal to noise ratio. The spectrometer cannot respond well to small changes in absorbance when the value of the measured absorbance is already low. On the other hand, in the case of high absorbance, the signal to noise ratio is relatively poor as a lower number of photons are entering the detector. The value of absorbance can be adjusted in the desired range of 0.5 OD to 1 OD by changing the cuvette path length or the concentration of the samples. The measured glucose absorbance shown in Figure 3.1 has the value less than 0.5

OD, and the low value of OD may be a reason for the similarity between the absorbance values of the two samples, even with very different glucose concentrations.

The glucose absorbance in the MIR range was analyzed using a Fourier transform infrared spectroscopy (FTIR) device (PerkinElmer Spectrum 100). The absorbance of two samples, a distilled water sample and a 1,000 mg/dL solution of glucose and distilled water sample over a MIR wavelength range extending from 950 cm^{-1} to 4000 cm^{-1} (corresponding wavelengths of 2,500 nm to 10,526 nm). To measure the absorbance, the intensity, I_0 (also known as reference sample or background), is measured using no sample. There is an optical window on a metal surface where a droplet of an aqueous solution is placed and absorbance is measured using the interferometry method. In Step 1, a droplet of the distilled water sample is placed on a metal surface to measure the intensity of the light caused by the distilled water (I) and measures the absorbance of the distilled water while the background light is measured as the reference measurement (I_0). In Step 2, a droplet of the glucose in distilled water solution is placed on the metal surface to measure the absorbance of this solution. Figure 3.2 shows the absorbance of the distilled water sample (blue curve), and the absorbance of glucose within distilled water (red curve). In Step 3, the black curve is the result of the glucose/distilled water solution minus the distilled water absorbance that is then multiplied by 10. Thus, there are peaks of absorption observed for glucose molecules within the MIR range as illustrated in Figure 3.2. For glucose sensing based on the absorption method, the wavelength of the light source should be selected at these optical windows, where the absorption due to the solution of glucose and water is more than the absorption due to just water. Otherwise, the water will obscure the glucose signal.

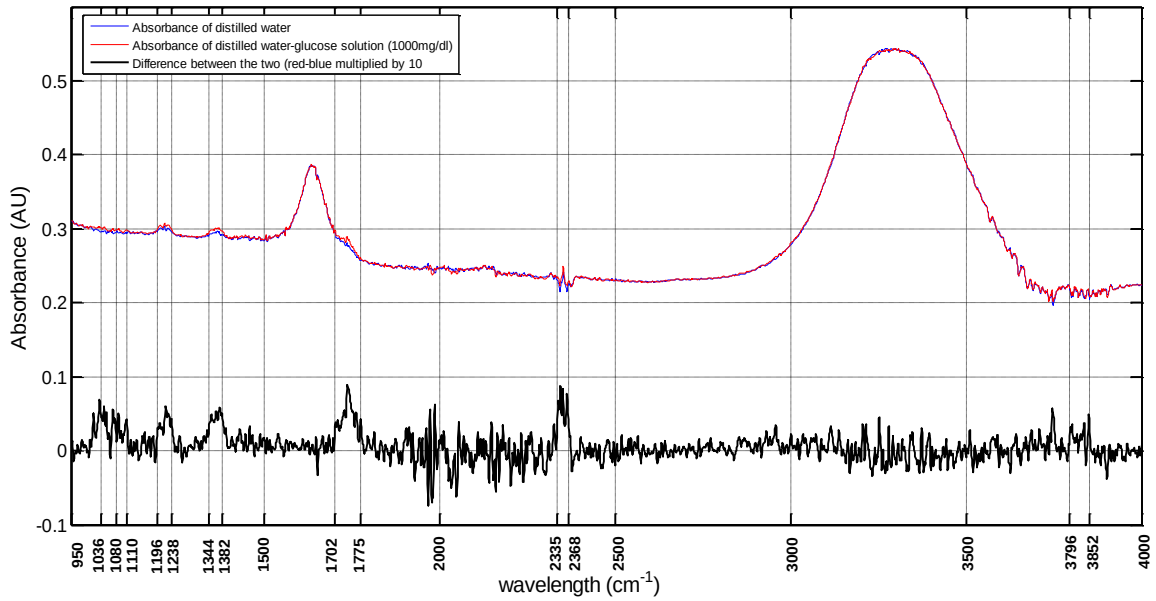


Figure 3.2: Absorption spectra for distilled water (blue), 1,000 mg/dL solution of glucose and distilled water (red), and the difference between the two (red signal – blue signal = black signal multiplied by 10 for ease of visibility).

One of the most significant challenges in non-invasive optical glucose sensing is the significant absorption of light by water. Blood plasma consists of about 92% water. This results in a significant light intensity signal coming from water compared to glucose (percent of light absorption by water > percent light absorption by glucose), and this reduces the sensitivity of the non-invasive optical sensor to changes in glucose concentration. Therefore, it is essential to select a wavelength range that results in low light absorption by water (and high absorption by glucose) and this coincides with a wavelength that results in a low absorption coefficient for water. The water absorption coefficient is measured in a wide range of wavelengths from 10 nm to 1 mm in [142], and the water absorption coefficient decreases when the wavelength of the incident light decreases from higher wavelength in MIR range to the lower wavelength in the NIR and visible ranges [142]. To minimize the obstructing effect of water and maximize the sensitivity to glucose, a wavelength within the NIR range is suitable for glucose sensing. This is consistent with literature

sources [143]–[145]. Although there are several absorption peaks associated with glucose in the MIR range (identifiable in Figure 3.2), a wavelength within the NIR range is preferred over MIR since it is absorbed less by water and therefore can penetrate deeper into tissue. Furthermore, NIR has a higher energy and can be transmitted through tissue, where it is absorbed by glucose molecules within the skin tissue fluid and by blood vessels. This is expected to increase the sensitivity of the sensor since light sources at NIR wavelengths will encounter many more glucose molecules.

In the next Step, we explain the correlation between glucose concentration and light intensity in the NIR range. An HR-1024i spectrometer (Spectra Vista Corporation, NY, USA) measures the attenuation of light as a function of glucose concentration for the range between 800 nm and 1,800 nm at room temperature. The light attenuation can be due to scattering and absorption of light by glucose molecules (and by other species) in solution. This spectrometer measures the ratio of a target scan to a reference scan (I/I_0). In our experiment, the reference scan is from a sample of distilled water and the target scan is from a sample of glucose solution with distilled water. I_0 is the light intensity measured by a photo detector when the sample used is distilled water and (I) is the light intensity measured by a photo detector when the sample is 400 mg/dL glucose in distilled water solution.

The graph in Figure 3.3 includes the log ($100/(I/I_0)$) as a function of wavelength. The light attenuation due to glucose molecules can be calculated using the formula log ($100/(I/I_0)$) based on Beer's law. There are two water absorption peaks between 1,350 – 1,520 nm and 1,790 – 1,960 nm that are shaded in blue [4], [146], [147]. Thus, it is important not to select wavelengths in these two intervals for glucose sensing light sources since there is a strong water absorption interference, especially in the case of low glucose concentration. Based on the light intensity measurements in

Figure 3.3, the four potential optimum wavelengths for glucose detection are 950 nm, 1,025 nm, 1,200 nm and 1,600 nm.

In Section 3.2, a description of the optical based non-invasive glucose sensor design is described. The optimal light source from the four potentially optimal wavelengths is identified in the first step of the preliminary study (950 nm, 1,025 nm, 1,200 nm and 1,600 nm). The wavelength resulting in the highest sensitivity to changes in glucose concentration is incorporated in the preliminary sensor design.

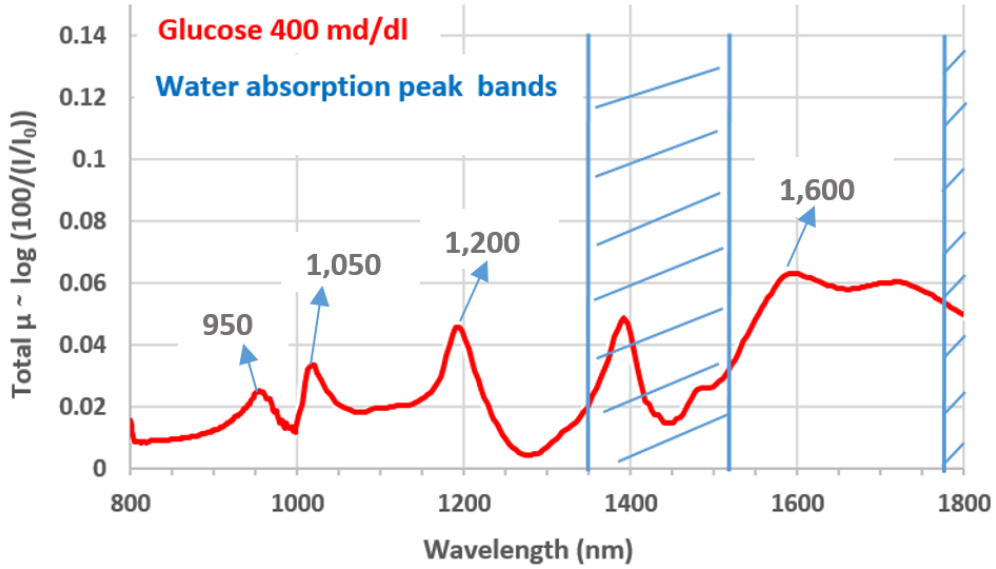


Figure 3.3: NIR spectra for 400 mg/dL solution of glucose in distilled water.

3.2 Design and Development of Optical Sensor with Single Wavelength Light Source

Based on the initial study described above, a single wavelength sensor is designed in this Section. Figure 3.4 includes a schematic of the optical based glucose sensor design using a single wavelength light source. The basic components are identified and include the following: photodiode (FGA015, Thorlabs Inc., Newton, NJ), light sources (LED940E, LED1050E, LED1200E and LED1600E, Thorlabs Inc., Newton, NJ), lenses (LA1540-C-ML, Thorlabs Inc.,

Newton, NJ), and cuvette (G-10-COVER, Science Outlet). An LED with a specific wavelength emits light that is collimated by a collimator lens. The parallel light passes through the sample of glucose-distilled water solution at which time a portion of light is absorbed by the water and glucose molecules (i.e. the light is scattered and/or is transmitted through the sample). The transmitted light is focused by a focusing lens and before it is absorbed by a photodiode. The photodiode converts the absorbed light into the analog current signal which is proportional to the glucose concentration.

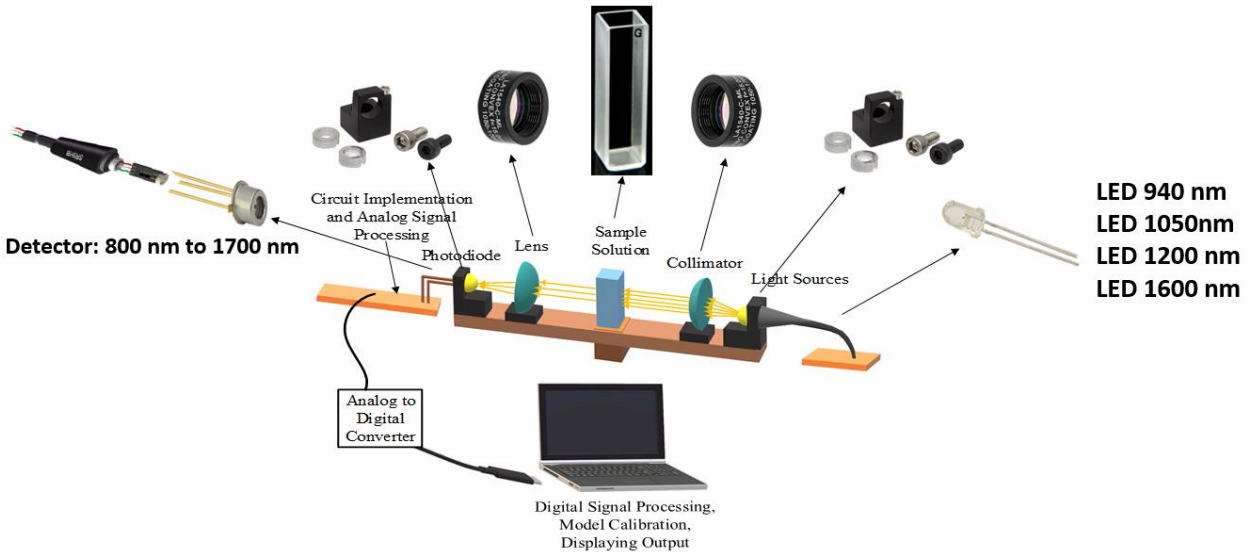


Figure 3.4: Schematic for optical based glucose measurement set-up with single wavelength light source.

The optical set-up configuration is implemented as shown in Figure 3.5. The detector, LED light source and sample cuvette are secured on an additive manufactured platform for stability and to minimize measurement error due to component movement. The cuvette contains the glucose sample and the glucose concentration is adjusted for multiple glucose concentrations.

The photodiode contains a transducer that converts the light intensity into a current signal. The current signal is filtered using analog circuits constructed on a breadboard, and the filtered

signal is converted to a digital signal using the USB oscilloscope in the Analog Discovery 2 instrument (Digilent Inc., Pullman, WA). The Analog Discovery 2 includes a data conversion IC (AD9648) that serves as a 14-bit resolution analog to digital converter (ADC) with sample rate of 100 MS/s. It is also equipped with an analog device (AD9717) that serves as a 14-bit resolution digital to analog converter (DAC) with sample rate of 100 MS/s. The hardware in the Analog Discovery 2 collects the analog signal and displays it on a laptop screen using Waveforms Software as a voltage signal with a certain amplitude. The signal is recorded for 10 ms and the average voltage amplitude value is collected from the Analog Discovery 2 and is assigned to each glucose concentration.

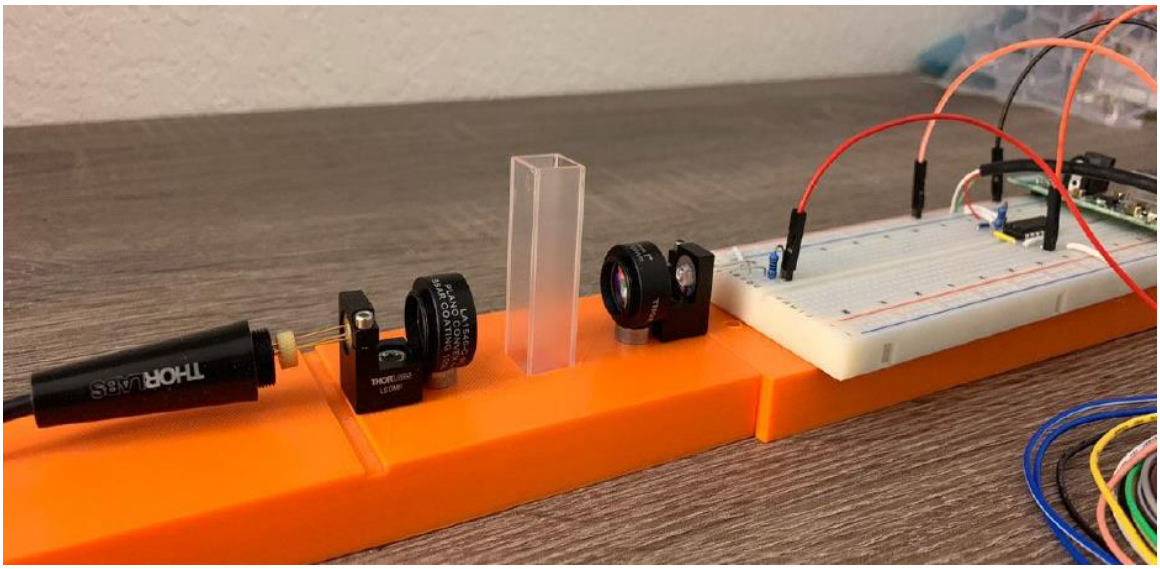


Figure 3.5: Platform configuration for optical based glucose sensor with single wavelength light source.

3.3 Experimental Method

A 950 mg/dL aqueous glucose solution is made by dissolving 950 mg of dextrose (or D-Glucose) (Carolina Biological Supply, NC, USA) in 100 mL of distilled water (Carolina Biological Supply, NC, USA). In order to reduce the error caused by motion artifacts, the cuvette containing

the 950 mg/dL glucose solution is fixed on the designed platform as demonstrated in Figure 3.4. A pipette is used to reduce the volume of the solution by 25% and this volume is replaced with distilled water in order to dilute the solution. As a result, the concentration of the glucose solution varies as follows: 950, 712, 523, 400, 300, 225, 168, 126, 95, 71 and 53 mg/dL. As the solution is diluted, the intensity of light transmitted through the solution is collected from the output of the sensor at each of these concentrations as a voltage signal. The voltage associated with each concentration is measured from the output of the analog circuit using the Analog Discovery instrument. The voltage is displayed on a laptop via the Waveforms Software interface and the averaged voltage for each sample over a 10 ms sweep is read. The glucose measurements described here are conducted for four optimal wavelengths. The values of these wavelengths are based on the potentially optimal wavelength values identified in Section 3.1 (950 nm, 1,025 nm, 1,200 nm and 1,600 nm). The LED light sources commercially available with the closest corresponding wavelengths include 940 nm, 1,050 nm, 1,200 nm and 1,600 nm. Therefore, these values are used in place of the potentially optimal wavelengths. The voltage measurements for each of these four wavelengths includes the following:

- voltage readings associated with each of the 11 glucose concentrations
- two measurements are made for each glucose concentration and these two are averaged.
- each voltage measurement is repeated five times over five different weeks.

Therefore, a total of 220 glucose concentration data points are collected for all four wavelengths.

3.4 CALIBRATION OF DATA

Light attenuation within the aqueous glucose solution sample is due to the absorption and scattering of the light by water and glucose molecules. The attenuation coefficient (μ_{tot}) depends on the glucose concentration (c) and is proportional to the logarithmic ratio of I_0 / I , where I_0 is

the light intensity as it passes through the sample of distilled water, and I is the intensity of light as it passes through the sample of glucose and distilled water. The relationship between the glucose concentration and the intensity of the measured light intensity is defined by Eq.(3.2).

$$\begin{matrix} \mu_{tot}\ depends\ on\ c \\ \mu_{tot} \propto\ log\ (\frac{I_0}{I}) \end{matrix} \quad \Rightarrow \quad log\ \left(\frac{I_0}{I}\right) depends\ on\ c \tag{3.2}$$

The intensity of light passing through the cuvette is measured by the photo detector and is converted to a current signal. The current signal is filtered as it passes through the analog circuit and the output of the circuit is a voltage signal. For a given wavelength, the glucose concentration is dependent on the logarithmic ratio of $I_0/\ I$ which is proportional to the logarithmic ratio of $Volt_{\text{water}}/Volt_{\text{glucose-water}}$ as defined by Eq. (3.5), where $Volt_{\text{water}}$ is the voltage corresponding to a light intensity measurement for a sample of distilled water only, and $Volt_{\text{glucose-water}}$ is the voltage corresponding to a light intensity measurement for a sample of glucose in distilled water solution.

$$log\ \left(\frac{I_0}{I}\right) depends\ on\ c \quad \& \quad log\ \left(\frac{I_0}{I}\right) \propto log\left(\frac{Volt_{water}}{Volt_{glucose-water}}\right) \tag{3.3}$$

$$\Rightarrow\ log\left(\frac{Volt_{water}}{Volt_{glucose-water}}\right)\ depends\ on\ c$$

For each wavelength, the logarithmic ratio of $Volt_{\text{water}}/Volt_{\text{glucose-water}}$ is calculated for all the sample measurements. Then, linear regression analysis is used to model the data in order to predict glucose concentration values for a given input value(s) (i.e. light intensity).

Linear regression is a supervised machine learning algorithm which finds a linear relationship between input variable(s) x_i (in this case it is the logarithmic ratio of voltage data points as expressed in Eq. 3.2) and an output variable y_i (in this case it is the known glucose concentration) by using the optimized value for each of the parameter(s). The most common way

to measure the performance of a regression model is by using the Root Mean Square Error (RMSE) as a cost function (Eq.(3.4)).

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - bx_i)^2} \tag{3.4}$$

In Eq.(3.4) the value of i changes from 1 to n, where n is the number of measurements, bx_i is the predicted value for the glucose concentration which can also be denoted by $\hat{y}_i$. and y_i is the actual value for the glucose concentration. A good fit between a model and measured data is when the value of RMSE is a minimum.

Another performance metric to evaluate the prediction model is the coefficient of determination which is also called R-squared and is defined by Eq. (3.5),

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{3.5}$$

where y_i is the actual value of the glucose concentration for the i^{th} data point, $\hat{y}_i$ is the predicted glucose value for the i^{th} data point, and $\bar{y}$ is the average of the actual values for the glucose concentration.

In this work, MATLAB™ is used as the data analysis program using a linear regression algorithm to fit a linear model (line) on the plot of the measured data values ($Log(\frac{Volt_{\text{water}}}{Volt_{\text{glucose-water}}})$) vs the actual glucose concentration. The linear model allows the prediction of the glucose concentration based on the measured intensity data.

3.5 Results and Discussion

The measured voltage data is plotted as a function of actual glucose concentration for each of the wavelength experiments described above. A linear line is fitted on the data using linear regression and the value of RMSE and R-squared is calculated for each of the 5 experiments

associated with each wavelength. Figure 3.6 shows the plot of measured voltage data $log(\frac{Volt_{\text{water}}}{Volt_{\text{glucose-water}}})$ vs actual glucose concentration for experiment 1 out of 5 for each of the 4 wavelengths.

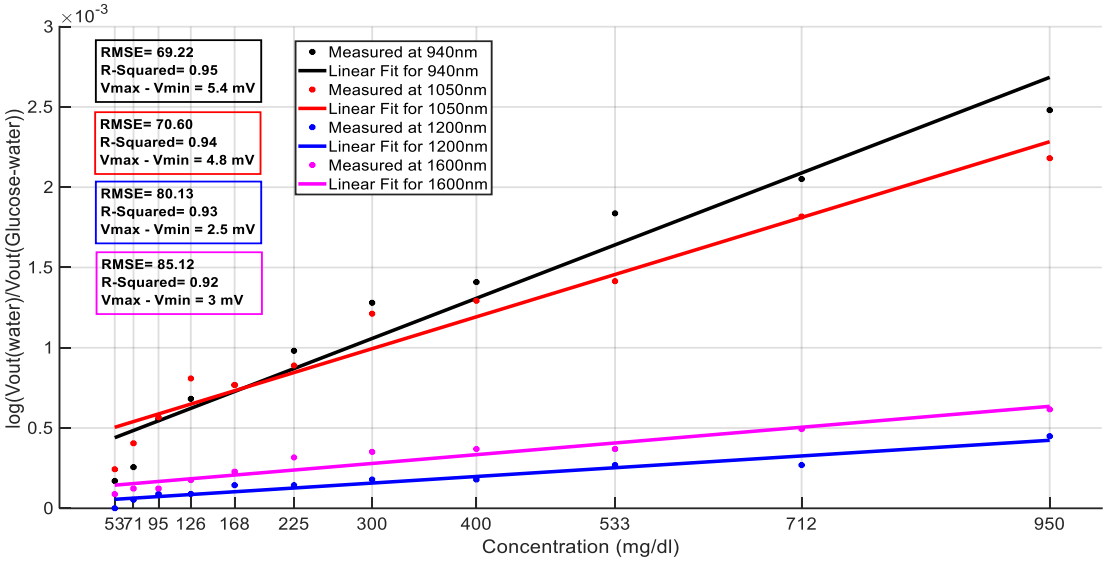


Figure 3.6: Linear regression model applied to measured voltage data as a function of actual glucose concentration values for 4 wavelengths (experiment 1 out of 5).

In order to identify the optimal wavelength that results in the most accurate glucose predictions, the average values of RMSE and R-squared are calculated for each wavelength for experiments 1-5 (see Table 3-1: The average values of RMSE, R-squared and range of voltage change (Vmax-Vmin) for each wavelength during 5 experiments).

Table 3-1: The average values of RMSE, R-squared and range of voltage change (V_{max}-V_{min}) for each wavelength during 5 experiments

	Averaged RMSE	Averaged R-Squared	Averaged Voltage Range (V_{max}-V_{min})
940nm	58.60	0.962	6.14 mV
1,050nm	60.48	0.958	4.06 mV
1,200nm	75.93	0.938	1.47 mV
1,600nm	72.86	0.944	1.76 mV

Based on Table 3-1: The average values of RMSE, R-squared and range of voltage change (Vmax-Vmin) for each wavelength during 5 experiments, the lowest and highest average RMSE and R-Squared values, respectively, are observed when the light source has a wavelength of 940 nm. This means that the glucose concentration is predicted with higher accuracy using a light source with a wavelength of 940 nm. On average, the change in the measured voltage range (V_{max}-V_{min}) is maximized when the incident light has a wavelength of 940 nm. This means that a change in the glucose concentration can be measured with higher sensitivity when using a 940 nm light source.

The preliminary results in this Chapter proves the feasibility of predicting glucose readings by analyzing the light intensity signal measured using a simple optical sensor with a single light source. There is definitely a dependence on glucose concentration for all four wavelengths, but the optimal performance is observed when the incident light source has a wavelength of 940 nm. The rest of this dissertation focuses on improving the accuracy of the glucose measurements, and the following indicates how this improvement is going to be achieved.

- A glucose sensor that works in transmission mode and measures the intensity signal within the visible and near infrared range - more in depth measurements.
- A multiple wavelength light source - higher signal due to glucose molecules and improvement of the sensitivity to and selectivity of glucose with more quantitative information about the glucose content in the solution.
- Adjustable distance between the cuvette and light source and between the cuvette and detector - optimize the accuracy of the glucose sensor.
- Use a printed circuit board (PCB) vs a breadboard (from preliminary design) - improve repeatability of the sensor

- Use digital filtering vs analog filtering (from preliminary design - improve the signal to noise ratio.
- Secure all of the components on the 3D printed sensor platform - reduce error caused by sensor component displacement.
- Use a peristaltic pump to gradually and continuously reduce glucose concentrations - improve error caused by manually changing glucose concentration by using a pipette.
- Limit the glucose range to match the physiological range between 40 and 250 mg/dL - more accurate glucose predictions.
- Use machine learning techniques that are more sophisticated than simple linear regression - create accurate glucose prediction models.

4. Design and Implementation of VIS-NIR based Optical Glucose Sensor Suitable for Multiple Wavelength Measurements

This Section presents details about the design and implementation of an optical based glucose sensor that measures light intensity passing through a sample at multiple wavelengths. This Section describes basic components required for the design of the glucose sensor, the strategy used to assemble different components, the collection of intensity data at multiple wavelengths, and the experimental methods to measure glucose concentration within an aqueous solution.

4.1 Sensor Schematic and Critical Parameters

Figure 4.1 illustrates the schematic of the glucose sensor design that operates in transmission mode and supports multiple wavelength measurements. The design is comprised of the following: (1) an LED printed circuit board (PCB) that serves as light sources; (2) a photo detector printed circuit board that detects light for a wide range of wavelengths; (3) a cuvette suitable for spectroscopy and the region of interest; and (4) a peristaltic pump with an adjustable flow rate to vary the glucose concentration in a controlled and measurable manner. Samples are tested in an optical glass cuvette. When light impinges on the sample, some portion of light is absorbed by the physiological factors within the sample and the rest passes through the sample with a different angle from the incident light (scattering) or with the same angle (transmission). Absorption, scattering and transmission of light through the sample depends on the optical properties of the sample, and these depend on the concentration of components within the sample. The light passing through the sample passes through the photo detector and is then digitized. The digitized data is displayed on a laptop and is further processed to extract quantitative information related to the glucose concentration.

The sensor design includes the following: (1) fixed components with minimum movement to reduce the measurement error caused by components displacements; (2) electrical circuits built on a printed circuit board for stability and noise reduction; (3) a peristaltic pump with a stepper motor and driver to automatically and precisely change the glucose solution concentration and to remove the measurement error caused by manually changing glucose concentrations; and (4) the ability to adjust the distance between the light source and the cuvette, and between the detector and cuvette, from 1 cm to 5 cm. Optimal distances are selected to maximize the sensitivity to glucose changes.

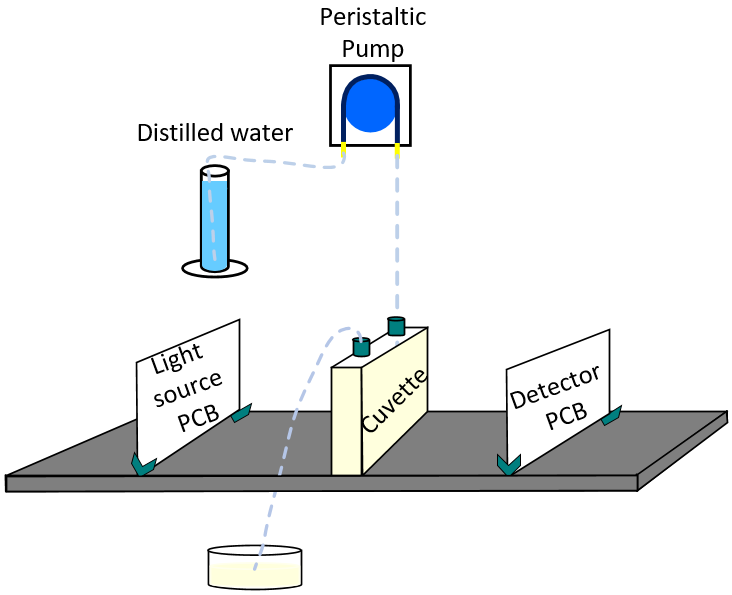


Figure 4.1: Schematic for the optical glucose measurement set-up.

4.2 Glucose Sensor Components

The items described in this Section are the electrical and optical components that are used to build the optical glucose sensor. Detailed descriptions and specifications are presented for each of the components.

4.2.1 SparkFun Triad Spectroscopy Sensor - AS7265x

The SparkFun Triad Spectroscopy Sensor - AS7265x (SparkFun Electronics, Colorado, USA) serves as the photo detector in the UV, VIS and NIR range [148]. The Triad consists of three devices; AS72651, AS72652, and AS72653 (ICs with black square shape in the center of Figure 4.2) that measure light intensity in the range between 372 nm and 966 nm in 18 individual bands. The part numbers for the three sensor devices are AS72651-BLGT, AS72652-BLGT, and AS72653-BLGT and they are available from AMS (Austria Mikro Systeme) AG manufacturer [149].

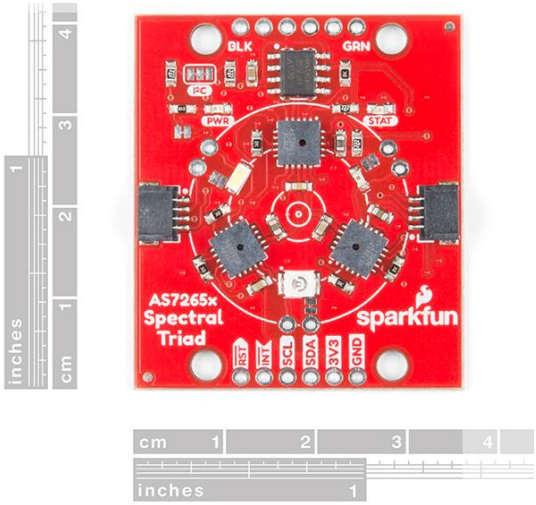


Figure 4.2: SparkFun Triad Spectroscopy Sensor - AS7265x [148] (reprinted with permission).

Figure 4.3 illustrates the 18 wavelength bands with 20 nm full width at half maximum (FWHM). The energy of the photons measured by the AS7265x sensor is in units of $\mu W/cm^2$. Each channel of the AS7265x sensor reads 35 counts for an incident 1 $\mu W/cm^2$ of light in that channel band. The accuracy of the channel is ±12% counts/$\mu W/cm^2$.

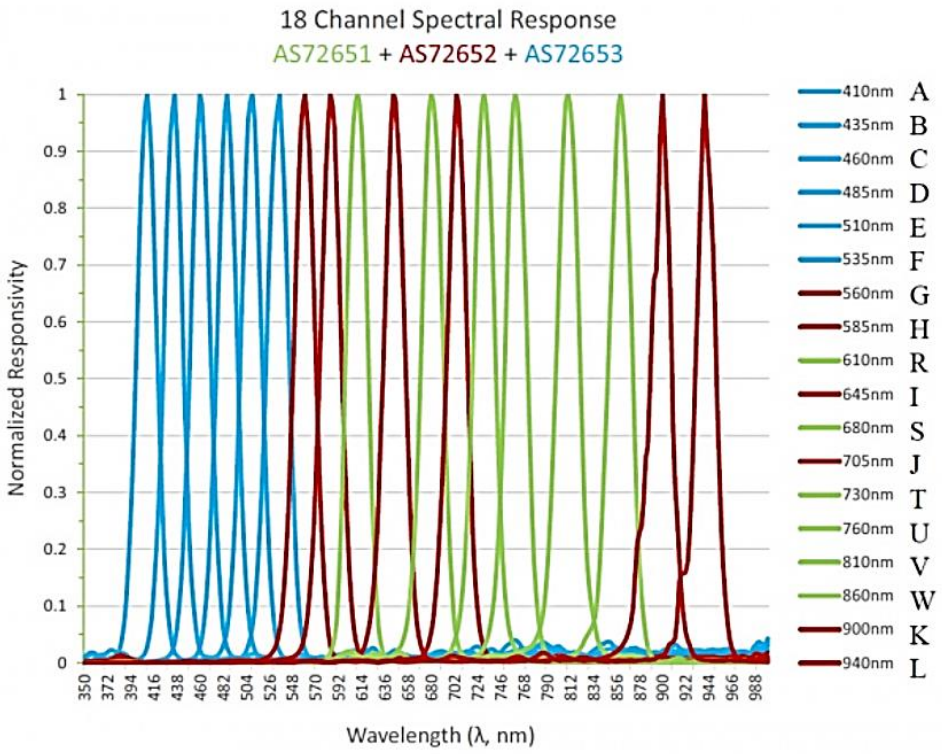


Figure 4.3: 18 channel spectral response of AS7265x [150] (reprinted with permission).

The AS72651 communicates with the two other devices on the AS7265x, the AS72652 and the AS72653, over the Inter-Integrated Circuit (I^2C) bus. These three devices on the AS7265x include a photodiode array and each photodiode array includes 6 channels as illustrated in Figure 4.4.

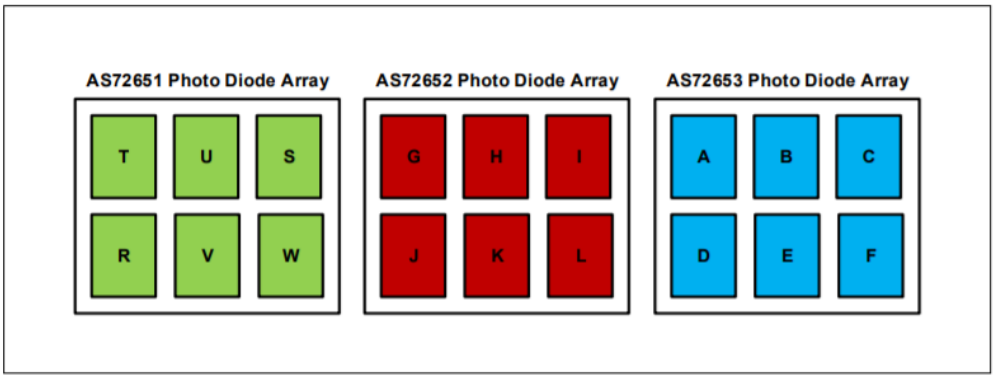


Figure 4.4: AS7265x photodiode arrays [151] (reprinted with permission).

Each channel in Figure 4.4 has a Gaussian filter characteristic with 20 nm FWHM. The Gaussian filter helps to establish stability of the sensor in terms of drift over time. Furthermore, each photodiode is supplied with a temperature sensor to compensate for the temperature drift [151].

The three devices (AS72651 to 3) also have an aperture through which the light passes and is received by the detector. The aperture-limited average field of view is ±20.5° in order to deliver specified accuracy.

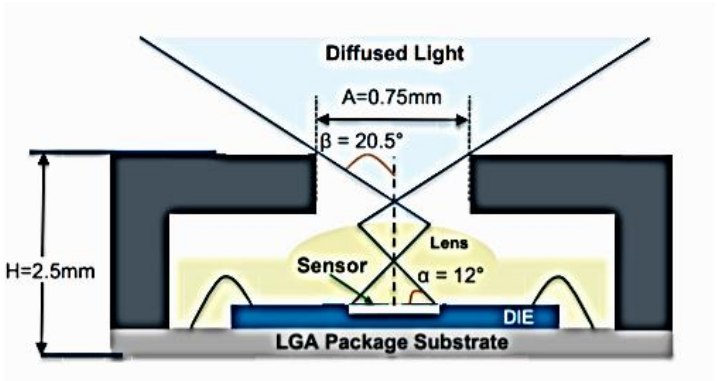


Figure 4.5: The AS7265x aperture average field of view is ±20.5°[151] (reprinted with permission).

The Triad AS7265x also includes a 5700 K (K stands for Kelvin) white LED, a 405 nm UV LED, and an 875 nm IR LED with the capability to illuminate the sample. However, these LEDs are not the light sources for the proposed glucose sensor because the sensor is operated in transmission mode, and this means that light sources need to be placed opposite the photo detectors.

In the case that someone decides to use the Triad as a light source, there are three extra footprints on the Triad PCB (outside the center of the AS7265x in Figure 4.2) that permit the connection of three external LEDs, which are soldered at these locations. In this case, the integrated LEDs should be disconnected from the Triad PCB by cutting the jumper next to each LED. The

footprint belonging to the three external LEDs are routed to the LED driver pins of the AS72651, AS72652 and AS72653, respectively. The LED drivers can deliver a programmable current with values of 12.5 mA, 25 mA, 50 mA or 100 mA.

4.2.2 RedBoard Qwiic Interface

Two digital interfaces can communicate with the Triad AS7265x control/access spectral, the serial universal asynchronous receiver-transmitter (UART) interface and the I^2C/RedBoard Qwiic interface. The RedBoard Qwiic interface (SparkFun Electronics, Colorado, USA) shown in Figure 4.6 is used to communicate with the Triad AS7265x via I^2C/Qwiic [152]. The RedBoard has an Arduino-compatible microcontroller (with 16 MHz clock speed) that can be programmed over a USB Micro-B cable using Arduino Integrated Development Environment (IDE) Software. The data that is measured by the Triad AS7265x within the 18 individual bands is read using IDE written in C++ code, and the IDE software displays the data on a laptop screen.

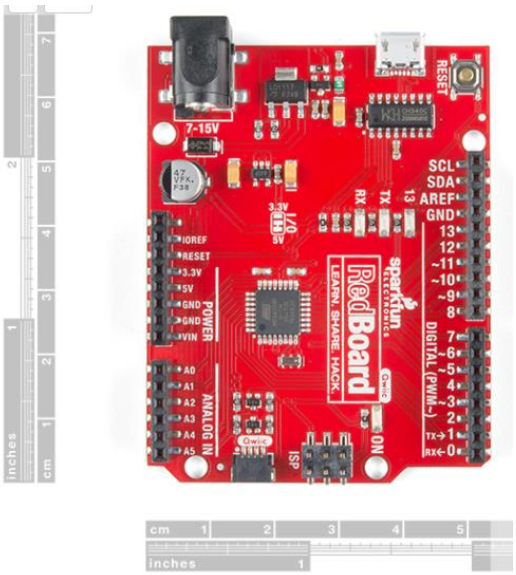


Figure 4.6: RedBoard Qwiic [152] (reprinted with permission).

4.2.3 Light Sources

This Section explains the design and fabrication process of the printed circuit board (PCB) for the UV, white and NIR LEDs light sources.

4.2.3.1 UV, Visible and NIR LEDs:

The spectral response of the UV LED (VLMU3100-GS08, Digi-Key Electronics, Minnesota, USA) is illustrated in Figure 4.7, with a wavelength centered at 405 nm. The UV LED specifications are summarized in Table 4-1.

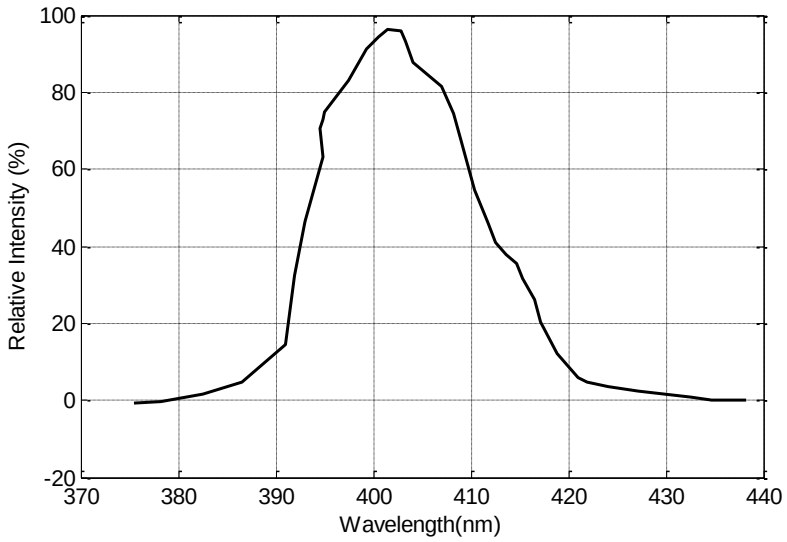


Figure 4.7: Relative intensity vs wavelength for UV LED [153].

Table 4-1: UV LED's specifications

Type	Ultraviolet (UV)
Current DC forward (max)	30 mA
Radiant intensity- min @ if	2.5 mW/sr @ 20mA
Wavelength	405 nm
Voltage forward (Typ)	3.2 V
Viewing angle	120°

The spectral intensity distribution of the white LED (L130-3080001400001, Digi-Key Electronics, Minnesota, USA) is illustrated in Figure 4.8 and the specifications are listed in Table 4-2.

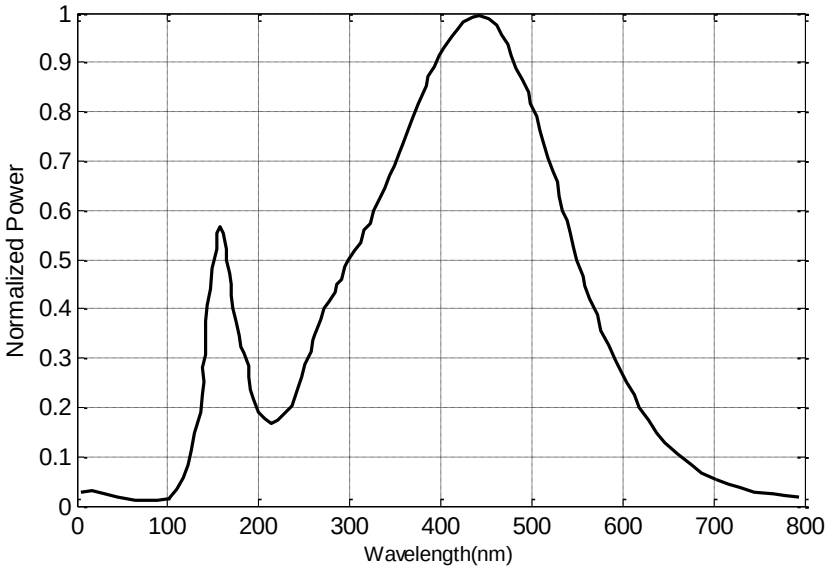


Figure 4.8: Normalized power vs wavelength for white LED [154].

Table 4-2: White LED's specifications

Color	White, Warm
CCT (K)	3000K
Flux @ 25°C, current-test	21m (Typ)
Current - test	60 mA
Voltage forward (Typ)	3.1 V
Viewing angle	-
Lumens/watt @ current - test	113 lm/W
Current - max	100 mA

Figure 4.9 illustrates the spectral distribution of the NIR LED (SIR19-21C/TR8, Digi-Key Electronics, Minnesota, USA) with a wavelength centered at 875 nm. The LED's specifications are summarized in Table 4-3.

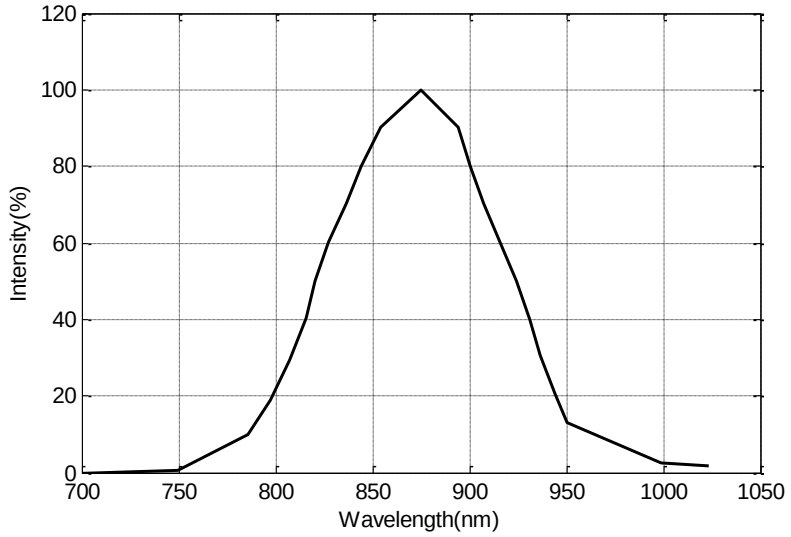


Figure 4.9: Intensity vs wavelength for NIR LED [155].

Table 4-3: NIR LED's specifications

Type	Infrared (IR)
Current DC forward (max)	65 mA
Radiant intensity- min @ if	0.2 mW/sr @ 20mA
Wavelength	875 nm
Voltage forward (Typ)	1.3 V
Viewing angle	145°

4.2.3.2 Resistors and Capacitors:

Surface Mount Technology (SMD) resistors and capacitors required for the design of the light source circuits are purchased from Digi-Key Electronics, Minnesota, USA. The resistors are 143, 150 and 294 Ω with part numbers RC0603FR-07143RL, CR0603-FX-1500ELF and RC0603FR-07294RL, respectively. The capacitors with the values of 10 µF and 0.1 µF have part numbers C0603C106M7PAC7867 and C0603C104K9RAC7867, respectively.

4.2.3.3 PCB Design and Fabrication of Light Sources

The PCB for the light source is designed using an online PCB design tool (EasyEDA) [156]. The schematic design, illustrated in Figure 4.10, includes three LEDs, the white LED, the 405 nm UV LED, and the 875 nm IR LED. These LEDs combined have the capability to illuminate the sample in a spectral range between 410 nm and 940 nm.

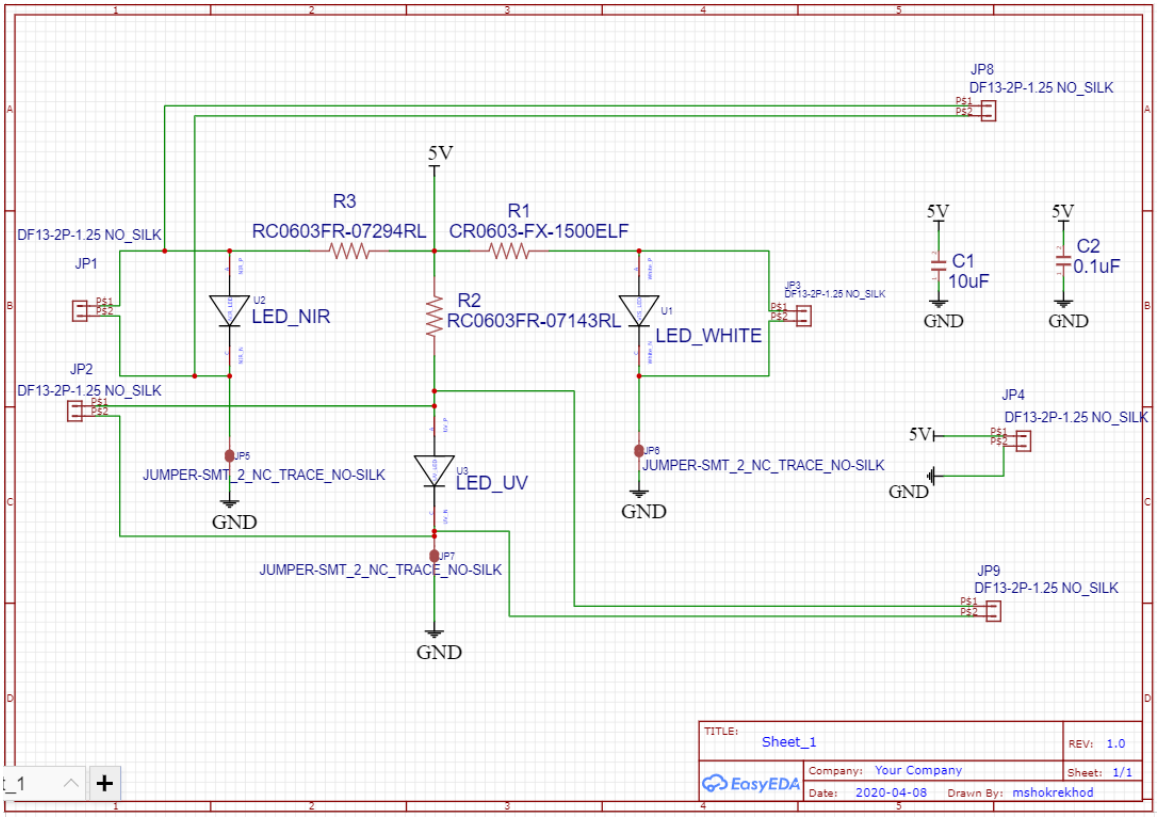


Figure 4.10: Schematic of the light source PCB design.

The LEDs are supplied with a 5 V voltage source connected in series with a resistor. The value of the resistors for the UV, White and NIR LEDs are 144, 152 and 296 Ω, respectively. These resistance values result in a current signal of 12.5 mA through each LED, and the voltage drop across each LED is equal to its typical forward voltage value. There are two bypass capacitors with values of 0.1 and 10 μF as shown in Figure 4.10 that are used to eliminate AC noise that may be present within a DC signal.

The cathode of each LED is connected to a jumper which is connected to the ground node. In the case of closed jumpers, the LEDs are supplied by a constant 5 V voltage source. However, the jumper connection can be cut or removed if an external driver needs to be applied to the LEDs. To be able to apply an external driver to an LED, the cathode and anode terminals of each LED are routed to header pins with a 2.54 mm pitch spacing. The LEDs can also be connected to external drivers via header sockets routed to the terminals of an LED. As mentioned in Section 4.2.1 SparkFun Triad Spectroscopy Sensor - AS7265x the Sparkfun Triad AS7265x has LED drivers that can be programmed to control the current that passes through each LED within the range between 12.5 mA and 120 mA. The other advantage of using the Triad AS7265x LED driver is the possibility of creating a modulated light beam (vs a constant light beam).

It is essential to ensure that all component footprints are accurately created in the PCB design. The LED footprints are created by adding a new PCB Library in EasyEDA based on the footprint information provided in the datasheet of each LED. The solder pad layout for the UV LED, white LED, and NIR LED can be found in [153], [154], and [155], respectively. The footprints (RESC1608X55N) for the 0603 (1608 Metric) resistors used in the PCB design are downloaded from the SnapEDA library.

A schematic of the PCB design is included in Figure 4.11. The connections between the component pads are created at the 'TOP' and 'BOTTOM' layers. The PCB area is covered with copper at both the 'TOP' and 'BOTTOM' sides where both planes are connected to ground net. The size and area of the light source PCB matches that of the Sparkfun Triad PCB in order to maintain symmetry, and is equal to 35.56 mm × 40.64 mm.

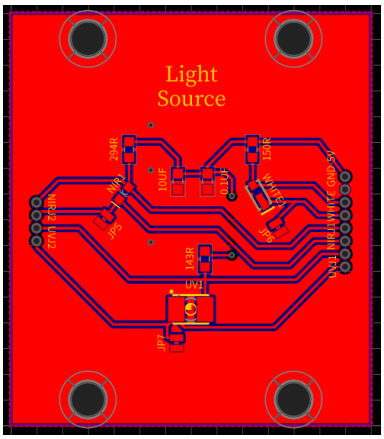


Figure 4.11: Image of the light source PCB design fabrication file.

The fabrication file for the PCB design can be downloaded as a Gerber file. The Gerber generated dialog shows the price of the PCB fabrication according to PCB quantity, thickness, surface finish and copper weight. The PCB Gerber file is downloaded and sent to a factory for production. In this case, the PCB design is ordered from JLCPCB Company, the largest PCB prototype enterprise in Hong Kong.

4.2.4 Standard Rectangular Cell

A rectangular optical cell suitable for spectroscopy in the range between 334 and 2,500 nm is purchased from Starna Cells Inc., CA, USA and is illustrated in Figure 4.12. This cell has four clear glass windows and two holes at the top. The holes located on the top of the cell are sealed with polytetrafluoroethylene (PTFE) stoppers, which provide a suitable seal for volatile liquids. Each stopper has circular holes at their center that allow water to flow into the cell via one of the

holes using a fluidic pump, and the solution to be pushed out of the cell via the hole in the second stopper. The cell has a volume of 14 mL, and an interior width, length and height of 10 mm, 40 mm, and 48 mm, respectively.

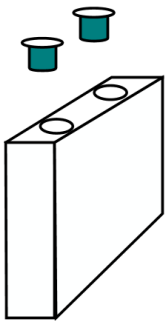


Figure 4.12: Illustration of optical cell suitable for spectroscopy in UV, VIS and NIR range [157].

4.2.5 Peristaltic Pump

A precision peristaltic pump (PE-WX14-S21-1, REET Corp., Connecticut, USA) is used to establish a flow rate between 2.5 μL/min and 24 mL/min as shown in Figure 4.13.

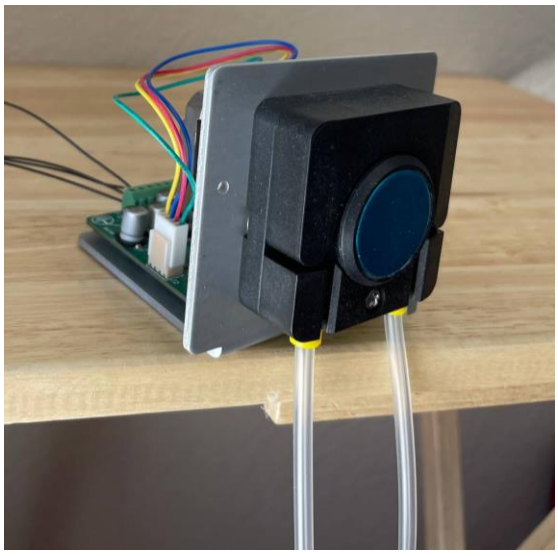


Figure 4.13: A precision peristaltic pump [158].

The pump has 4 rollers and requires a power supply in the range of 11.4 to 25.2 V DC with a power rating ≤ 10 W. The pump is equipped with a driver/controller board that can change the motor speed to set different flow rates. The stepper motor has a maximum angular speed of 60 rpm with a resolution of ±0.1 rpm.

A tubing of customizable size is connected to the pump to control the flow rate based on the motor speed. A tube with an inside diameter of 3 mm and an outside diameter of 5 mm is connected to one of the stoppers, and for this size of tube, a flow rate up to 24 mL/min is achievable as demonstrated in Figure 4.14 (see line marked 3×1 mm).

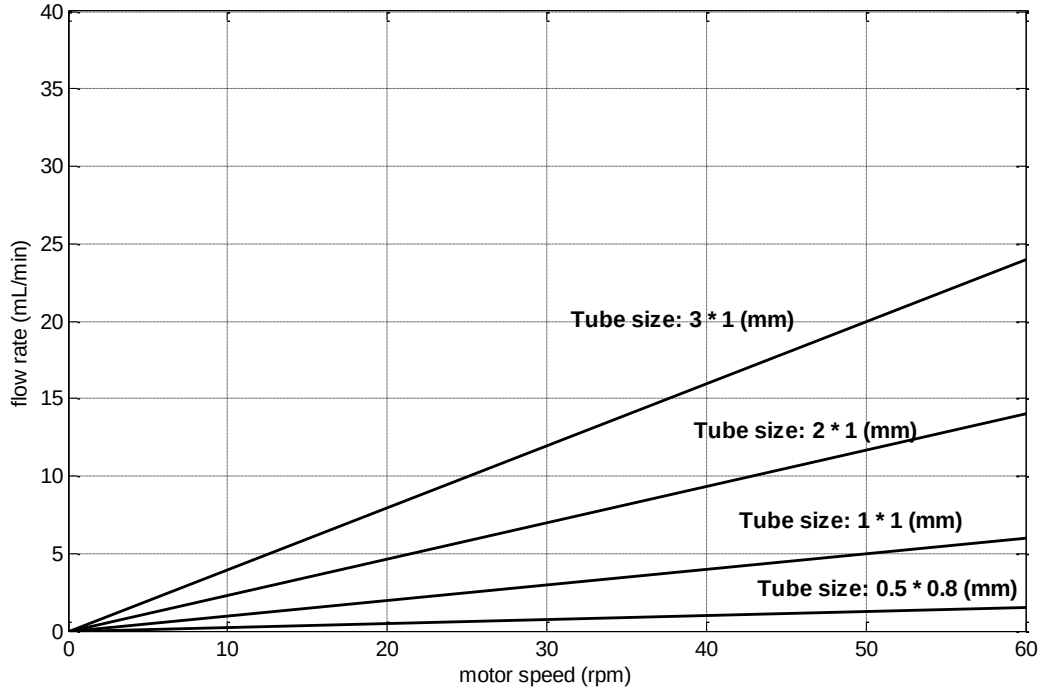


Figure 4.14: Flow rate (mL/min) vs motor speed (rpm) for four different tube sizes.

Figure 4.15 includes the driver board of the pump identifying the BCD dial switch used to control the motor speed. The BCD dial switch offers 15 different speeds with a range from 1 to 60 rpm. Once the BCD dial switch is fixed, the value is set for the motor speed as outlined in Table 4-4.

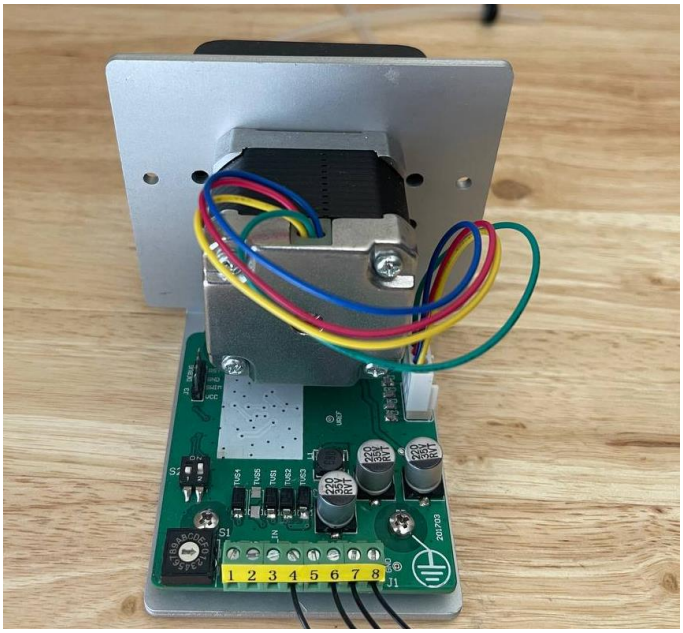


Figure 4.15: A BCD dial switch (bottom left) is located on the pump driver board.

Table 4-4: The pump motor speed as a function of the BSC dial switch position

Control mode	BCD dial switch position	Speed (rpm)
Fixed speed control	1	1
	2	3
	3	5
	4	8
	5	10
	6	15
	7	20
	8	25
	9	30
	A	35
	B	40
	C	45
	D	50
	E	55
	F	60
Setting for variable speed use	0	-

4.2.6 Design and Fabrication of Glucose Sensor Platform

The glucose sensor design consists of several components assembled on a mounting hardware platform. The mounting hardware platform consists of three additive manufactured parts:

- the main sensor platform (Figure 4.16)
- an L shaped bracket to hold the photo detector PCB (Figure 4.17)
- a similar L shaped bracket to hold the light source PCB (Figure 4.17)

These platform components ensure a stable and secure platform for the glucose sensor design with enough flexibility for adjustments. Each component is designed using an online computer aided design (CAD) software [159] and manufactured using a 3D printer.

A schematic of the main platform layout using CAD software is illustrated in Figure 4.16. The design includes a cuvette holder in the middle of the main platform, and two holes at each end. The L shaped bracket design is shown in Figure 4.17. Screws secure each L shape bracket (or PCB holder) to each side of the platform via the two holes. The PCB holders can move freely back and forth across the platform, and this adds flexibility to change the distance between the PCB holder and the cuvette.

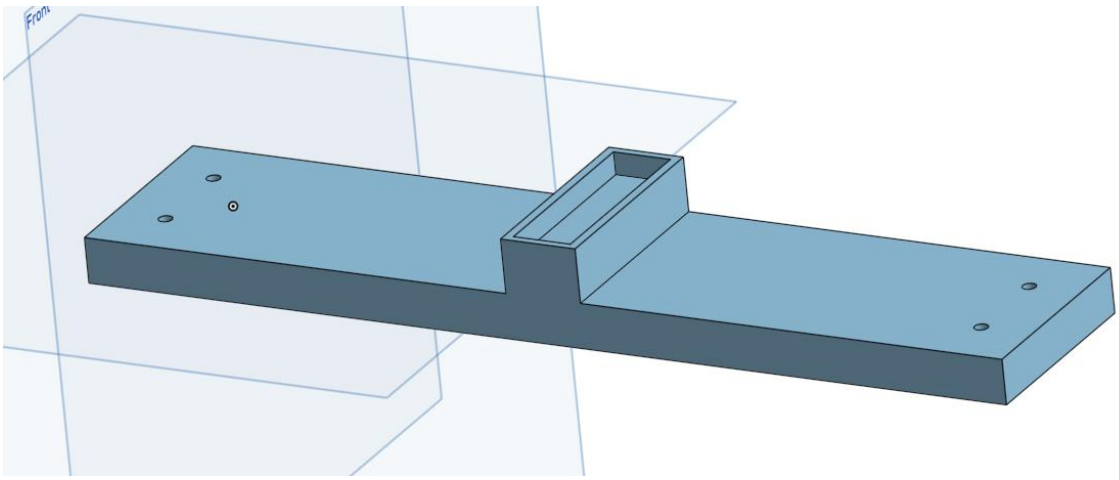


Figure 4.16: Layout of the glucose sensor platform design.

Figure 4.17 includes a schematic of the L shaped bracket that is used to secure the detector and the light source PCBs. The detector and light source PCBs are drilled into the platform using each L shaped bracket with four M3×16 mm screws and four female thread nuts (M3×10 mm×5 mm). Five rows of holes, with a distance of 10 mm between each pair of holes, are designed on the base of the L shaped bracket. The L shaped bracket can be adjusted along the main platform in the horizontal direction and can be fixed at a certain distance from the cuvette by using two male screws (M3×12 mm) and two female thread nut (M3×10 mm×5 mm).

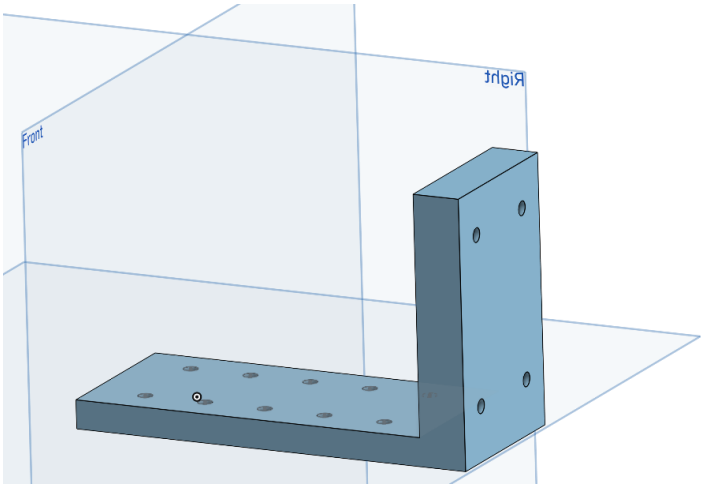


Figure 4.17: Layout of the L shaped bracket designed to hold the light source or photo detector PCBs

4.3 Build the VIS-NIR based Optical Glucose Sensor

An image of the custom-built VIS-NIR based optical sensor set-up is shown in Figure 4.18. The stabilization of the detectors, light sources, and cuvette configuration on the platform minimize measurement error that could result from any component displacement or movement. The minimum distance from the light sources to the cuvette or from the detectors to the cuvette is 10 mm. It is possible to increase these distances up to 50 mm with incremental steps of 10 mm. In this work, the optimal distance between the light sources and the cuvette, and between the detectors

and the cuvette are determined to be 20 mm and 10 mm, respectively. By testing the glucose sensor when the distances are set to the optimal values, the measured light intensity shows a higher correlation with changes in glucose concentration.

A peristaltic pump is used to adjust the glucose concentration of the solution contained in the cuvette container in a controlled manner. Distilled water is pumped into the cuvette through a tube that is secured by a stopper at one of the holes on top of the cuvette. Pumping a volume of water through one hole of the cuvette causes the same volume of the solution to be forced out of the cuvette through the second hole on top of the cuvette. This results in a change in glucose concentration within the cuvette that will be described in detail in Section 4.5.1 Sample of Glucose and Distilled Water. The use of a pump to dilute the solution automatically eliminates any measurement error caused by manually changing the concentration using a pipette.

When light from the light sources impinge on the cuvette sample, a portion of this light is absorbed by chemical species within the glucose sample and the rest passes through the sample. The light is either scattered (at an angle different from the incident light) or transmitted (with no change in incident angle) through the sample. Absorption, scattering and transmission of the light through the sample depends on the optical properties of the sample, which depend on the concentration of species components. The light passing through the sample is detected by the Triad Sensors (by photo detectors) at 18 different wavelengths in a range between 410 nm and 940 nm. The intensity data measured at these wavelengths is displayed on a laptop using IDE interface software. The measured intensity data will be processed and analyzed to identify wavelengths where the light intensity has significant correlation with glucose concentration, and then the intensity data at the identified wavelengths will be used to extract the glucose concentration.

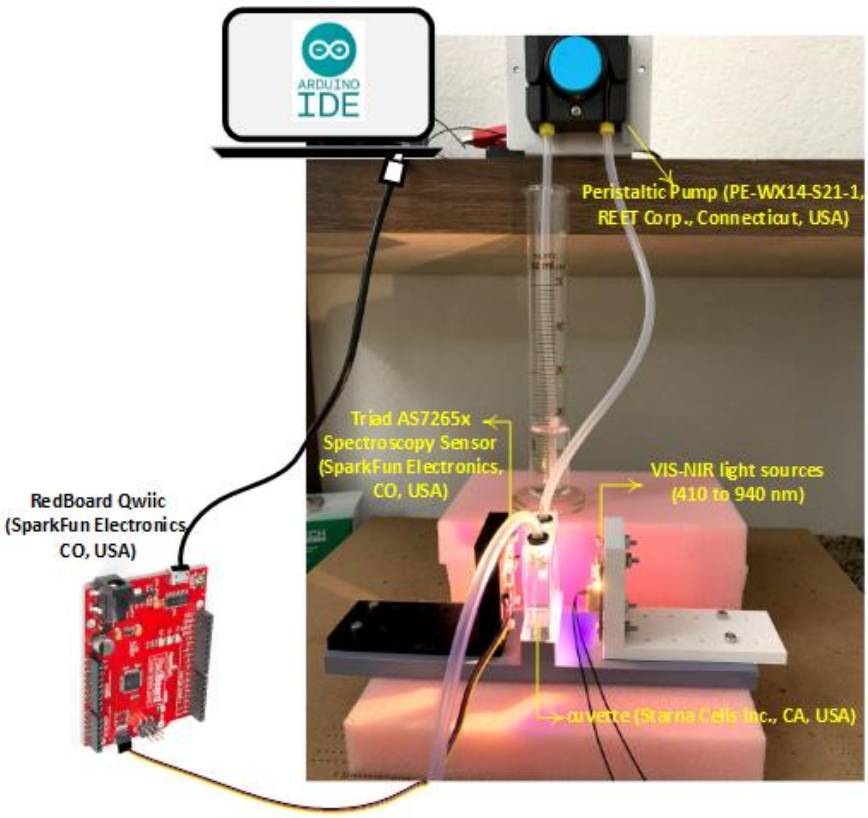


Figure 4.18: VIS-NIR based Optical Glucose Sensor Set-up.

Figure 4.19 shows the light intensity measured at 18 wavelengths when all three LEDs light sources are on for three types (Type 1, 2 and 3) of measurements. Type 1 is when there is no cuvette between the light source and the detector (black line in Figure 4.19), Type 2 is when the cuvette includes only distilled water (blue line in Figure 4.19), and Type 3 is when the cuvette includes a 120 mg/dL aqueous glucose solution (red line in Figure 4.19). The operation of each of the three light sources was tested manually by covering each LED one by one while observing the change in amplitude of the collected spectra. Figure 4.19 shows that the amplitude of the measured signal coming from the visible light source is much higher compared to the UV and NIR light sources. One possible reason for this observation is the higher power associated with the visible light source, which results in a higher brightness. Glucose absorbs light within wavelengths where

the measured signal for the glucose solution (red line) is lower than the amplitude of the measured signal for only water (blue line). The blue (water only) and red (glucose and water) lines are very close to each other. To understand how the measured signal is affected by the glucose molecules, multiple experiments are performed for different glucose concentrations within aqueous solutions as described in Subsection 4.5 Experimental Methodsanalysis used to extract the quantitative information related to the glucose concentration is described in Chapter 5.

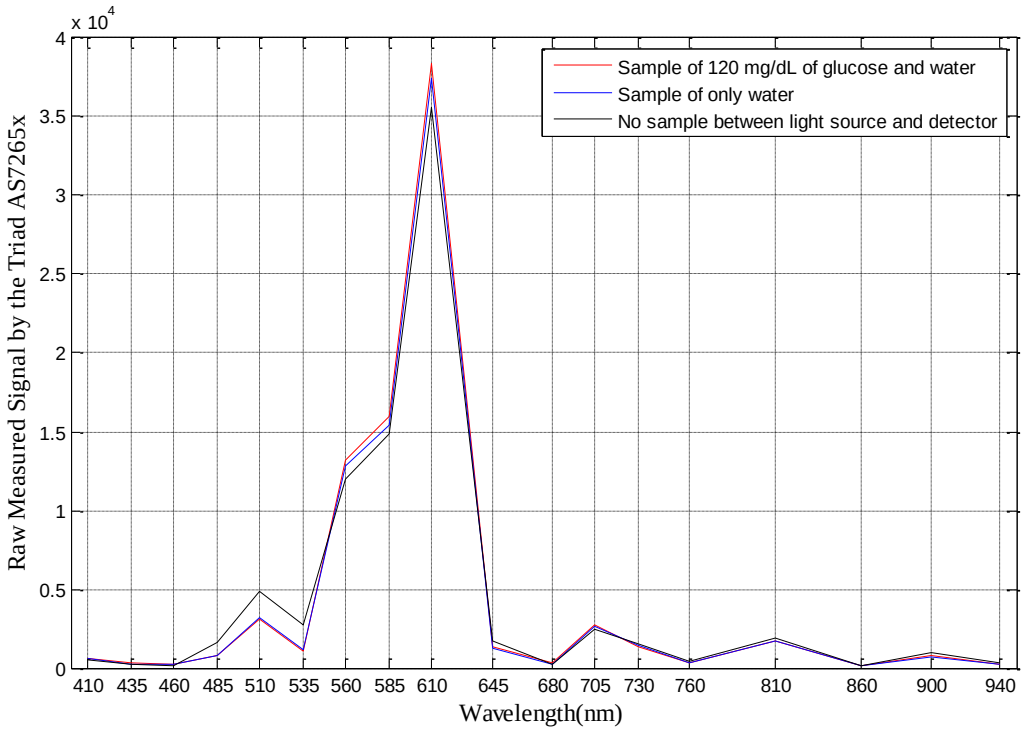


Figure 4.19: Light intensity measured at 18 wavelengths by the Triad AS7265x when all 3 LED light sources are on.

4.4 COLLECTION OF DATA

The Triad sensor includes 3 devices (AS72651, AS72652, AS72653). Each device has a 6-channel photodiode array that converts light intensity into a proportional current signal. Each channel is involved in measuring the intensity at a specific wavelength. Since the Triad has 3 devices and each device has 6 channels, the light intensity is measured at 18 different wavelengths.

In addition, each device includes an analog to digital converter (16-bit resolution ADC) that integrates the current signal from each channel's photodiode. The ADC continues to integrate the current signal at the rate of the integration time (between 2.87 ms and 711 ms), and then converts the current signal into a digital signal. The digital signals coming from the ADC are saved into AS72651 I²C virtual register sets. Then, the RedBoard Qwiic is used to read the measured signals at 18 wavelengths.

The data can be stored in registers as a 2 byte integer or a 4 byte floating point. Figure 4.20 is an example of 4 byte (32 bit binary) data composed of 8-bit unsigned integer and 23-bit fraction.

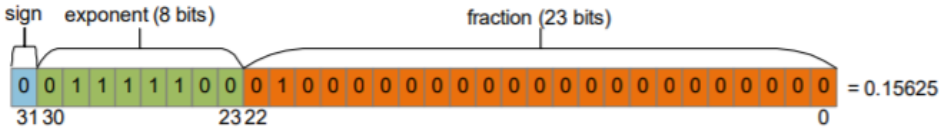


Figure 4.20: An example of 4 byte floating point data [151].

The floating point value (FPvalue) for the above example is calculated based on Eq.(4.1).

$$FP_{value} = (-1)^{sign}\left(1 + \sum_{i=1}^{23} b_{23-i}.2^{-i}\right).2^{(e-127)} = (-1)^{0}\left(1 + \sum_{i=1}^{23} b_{23-i}.2^{-i}\right).2^{(124-127)} \quad (4.1)$$

$$= 0.15625$$

In order to read the data stored in registers and access to device control functions, the RedBoard is connected to the Triad sensor over an I²C interface. The RedBoard is an Arduino-compatible microcontroller and can be programmed using Arduino IDE software. There are various functions that can be called from the SparkFun library and used for different purposes, such as:

- enable or disable LEDs that are integrated into the Triad
- tell the sensors to take measurements while the LEDs are on or off
- take readings from all 18 channels

- adjust the integration time
- adjust the LED current signals
- read the measured temperature, etc

The default setting for the sensor integration time is 50 ×2.8 ms or 140 ms. The integration time can be changed using the code: “sensor.setIntegrationCycles(any value between 0 to 255)”. For example, the code “sensor.setIntegrationCycles(120)” results in an integration time with a value of 121 × 2.8 ms or 338.8 ms.

The default gain for each photodiode channel is 1. The gain can be set to different values such as 1, 3.7, 16 and 64 using the line code: “sensor.setGain(AS7265X_GAIN_1X)”.

By programming the RedBoard Qwiic interface, the data from 18 channels can be taken through I^2C, and displayed using a Serial Monitor tool in Arduino IDE software.

4.5 EXPERIMENTAL METHODS

This Subsection explains the experimental methods employed to test each glucose solution sample. First, the sensor is evaluated using simple solutions that are composed of glucose and distilled water. Second, the sensor is evaluated using a solution that more closely resembles that of a body fluid, by adding protein to the glucose and phosphate buffered saline solution. The methodologies used to measure the intensity data for different combinations of concentration of species (glucose and protein) within distilled water are explained in the Subsections 4.5.1 Sample of Glucose and Distilled Water and 04.5.2 Sample of Glucose, Protein, and Phosphate Buffered Saline Solution, respectively.

4.5.1 Sample of Glucose and Distilled Water

A 250 mg/dL solution is made by dissolving 250 mg of dextrose (or D-Glucose), anhydrous, granular, reagent Grade (Carolina Biological Supply, NC, USA) in 100 mL of distilled

water (Carolina Biological Supply, NC, USA). The peristaltic pump transfers distilled water into the 250 mg/dL solution via one of the holes on top of the cuvette with a flow rate of 7 mL per minute. Flow of water into the cuvette causes the same volume of the solution to be forced out through the second entrance hole on top of the cuvette. It takes about 3.5 minutes for the pump to add 24 mL of distilled water into the cuvette. Since the cuvette has a volume of 14 mL, this results in a change in glucose concentration from 250 to 40 mg/dL as the sample is diluted. Every 17 seconds, 2 mL of distilled water is pumped into the solution. The resulting changes in glucose concentration every 17 seconds are listed in Table 4-5.

Table 4-5: Glucose concentration changes by dilution of the solution every 17 seconds

Time index (n = 17 sec)	Glucose level (mg/dL)
0	250
1n	250 × 12/14 = 214.29
2n	214.29 × 12/14 = 183.67
3n	183.67× 12/14 = 157.43
4n	157.43× 12/14 = 134.94
5n	134.94× 12/14 = 115.67
6n	115.67× 12/14 = 99.15
7n	99.15× 12/14 = 84.98
8n	84.98× 12/14 = 72.84
9n	72.84× 12/14 = 62.44
10n	62.44× 12/14 = 53.52
11n	53.52× 12/14 = 45.87
12n	45.87× 12/14 = 39.32

The Triad sensor measures the intensity of light that passes through the sample with an integration time of 140 ms. The light passing through the sample is detected using 18 different wavelengths ranging from 410 nm to 940 nm. Arduino IDE software is used to take the measured data over I^2C, read the data via a serial port every 1 second, and finally display the data via a serial

display tool within the Arduino IDE software. The glucose concentration is gradually lowered from 250 mg/dL to 39.32 mg/dL which takes 3.5 minutes. The total number of measurements during 3.5 minutes is equal to 3.5 minutes × 1 measurement/1 second * 60 seconds/minute ≈ 200 measurements. For each of the 200 concentration data points, the intensity data is detected using 18 wavelengths. The measured data is saved as a $k\ x\ l$ matrix, where k is equal to the total number of measurements (200) and l is equal to the total number of wavelengths (18).

After measuring the light intensity transmitted through the glucose-distilled water solution, the same experiments are repeated for distilled water without glucose. The entire process described here is repeated 10 times during different days, resulting in 10 experiments for the glucose-distilled water series and 10 experiments for the distilled water series. Note that for each experiment, the intensity values for glucose concentration between 40 mg/dL and 250 mg/dL are collected using 18 different wavelengths. Since approximately 200 intensity measurements are collected as the concentration is decreased from 250 mg/dL to 40 mg/dL using each of the 18 wavelengths, and each test is repeated 10 times for each wavelength, this results in approximately 2,000 total tests per each of the 18 wavelengths. Thus, for each of the 18 wavelengths there are 2,000 data points for the glucose-distilled water experiments and 2,000 tests for the distilled water experiments. Table 4-6 illustrates how the data is classified into 21 classes within the glucose concentration range between 40 mg/dL and 250 mg/dL. Each class includes data that corresponds to a glucose concentration within a discrete 10 mg/dL range. For example, class 1 includes data associated with glucose concentration values between 40 and 50 mg/dL. There are 10 tables like Table 4-6, one for each of the 10 experiments for the glucose-distilled water solution. The data associated with the glucose/distilled water solution is paired with corresponding data for the distilled water solution with no glucose. The measured intensity data for the glucose-distilled water solution is

normalized to the measured intensity data for the distilled water alone, thus removing the effects of any possible factors that have an effect on the intensity measurement results for glucose. The data processed in this manner will represent a more accurate measurement of the glucose concentration.

Table 4-6: Abbreviated table illustrating the data collected for 1 of the 10 glucose-distilled water solution experiments

Index	Glucose level (mg/dL)	Wave1	Wave2	…	Wave18	Class label
1	40	701	305	…	193	1
2	40.9	701	305	…	193	1
. .	. .	. .	. .	…	. .	. .
11	49.8	699	305	…	192.7	1
12	50.6	695	305	…	192.7	2
. .	. .	. .	. .	…	. .	. .
21	61.2	698	304	…	192.1	3
. . .	. . .	. . .	. . .	…	. . .	. . .
200	250	690	297	…	178	21

4.5.2 Sample of Glucose, Protein, and Phosphate Buffered Saline Solution

One of the most common solutions used for in vitro experiments is phosphate buffered saline (PBS) solution. PBS tablets (Grainger, Inc. USA) are used to make PBS solutions with a pH of 7.4. A pH value of 7.4 matches the average pH of the blood in a human adult. By adding one PBS tablet to 100 mL of distilled water, a 100 mL PBS base solution containing 137 mM of NaCl and 2.7 mM of KCl is made.

Protein and glucose are added to the PBS base solution to prepare a solution that assimilates the composition of human blood plasma. As mentioned in Subsection 2.1.2 Blood Componentsthe

most abundant plasma constituent is water (~91%), followed by proteins (~7%). The most dominant protein in plasma is serum albumin (~57%) with a concentration range between 3.5 and 5.5 g/dL. In this study, bovine serum albumin fraction V (Grainger, Inc, USA) is added to the PBS solution to form a solution with albumin protein. Three protein solutions are prepared by dissolving:

- 4 g albumin in 100 mL of PBS solution (solution 1)
- 4.5 g albumin in 100 mL of PBS solution (solution 2)
- 5 g albumin in 100 mL of PBS solution (solution 3)

Then, 250 mg of D-Glucose (Carolina Biological Supply, NC, USA) is dissolved in each of the three solutions, thus creating 250 mg/dL glucose solutions with 4 g/dL, 4.5 g/dL and 5 g/dL albumin. Three experiments (1-3) are set up to dilute the glucose in solutions 1-3 in the cuvette from 250 mg/dL to 40 mg/dL glucose using the same method described in Subsection 4.5.1 Sample of Glucose and Distilled Water and illustrated in Figure 4.21. More specific details are provided below.

In experiment 1, the PBS solution has a glucose concentration of 250 mg/dL and albumin with concentration of 4 g/dL. This solution is used to fill the cuvette with a volume of 14 mL. Then, 24 mL of PBS that only contains albumin with the same concentration (i.e. 4 g/dL) is pumped into the cuvette. Pumping this amount of solution takes about 6 minutes to lower the glucose concentration gradually until the concentration is decreased from 250 to 40 mg/dL. During the time it takes to dilute the glucose in solution from 250 to 40 mg/dL (~ 6 minutes), the Triad sensor measures the intensity data at 18 wavelengths every second (approximately), and this results in about 360 readings for each wavelength. Similar tests are performed with solutions containing 4.5 g/dL and 5g/dL albumin in experiments 2 and 3, respectively. The platform remains stationary

during the collection of the spectrum for the three experiments. These three experiments are repeated 5 times, and results in approximately 5,400 (3×5×360 = 5,400 total data points for the 3 experiments per each of the 18 wavelengths. After removing outliers, 5,233 data points are left and are used to create prediction models that estimate the glucose concentration at each wavelength.

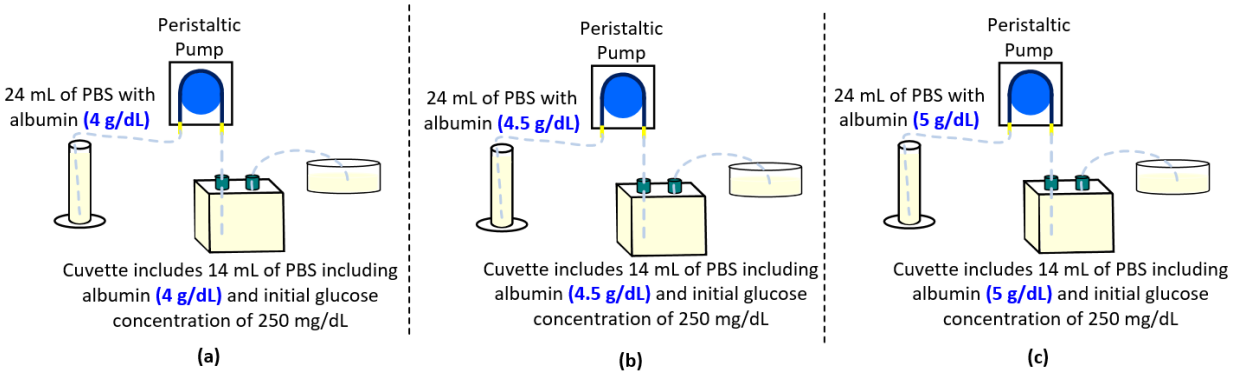


Figure 4.21: Experimental set-up used to decrease the glucose concentration from 250 mg/dL to 40 mg/dL by dilution of PBS solution while albumin concentration is fixed to (a) 4 g/dL, (b) 4.5 g/dL, and (c) 5 g/dL.

4.6 Results and Discussion

In this Chapter, the design and development of an optical based glucose sensor that measures light intensity is described. The light intensity data corresponding to 18 wavelengths is collected for different concentrations of glucose within two type of solutions, a glucose and distilled water solution, and a glucose within albumin PBS solution. Figure 4.22 demonstrates the raw signal measured for 360 different glucose concentrations (between 250 and 40 mg/dL) within the albumin PBS solution for the first experiment (1 out of 15 experiments). The amplitude of the signal measured for wavelengths between 535 nm and 645 nm is high, and this may be due to the higher power of the white LED (compared to the UV and NIR LEDs). The absorption coefficients of the glucose, water, and albumin is dependent on the wavelength of incident light, and the value

of the absorption coefficient affects the amplitude of the measured signal. The measured signal will be low for wavelengths where the species has higher absorption. The raw data shown in Figure 4.22 requires extra processing to turn it into useful information.

There is need to perform signal post-processing to extract quantity information about glucose concentration based on the measured spectra. In next Chapter, for each experiment, the measured spectra is first filtered using low pass filter in MATLAB^{TM} to remove outliers, then the spectra of glucose is normalized by spectra of water, and this allows for better visualization of how measured signal are affected by the changes in glucose concentration. Finally, the statistical method is applied to the normalized spectra to identify wavelengths where intensity of light is much affected by glucose concentrations. Once the optimal wavelengths are identified, the next step is to apply machine learning methods on the processed data at optimal wavelengths in order to create prediction model that predicts glucose concentration.

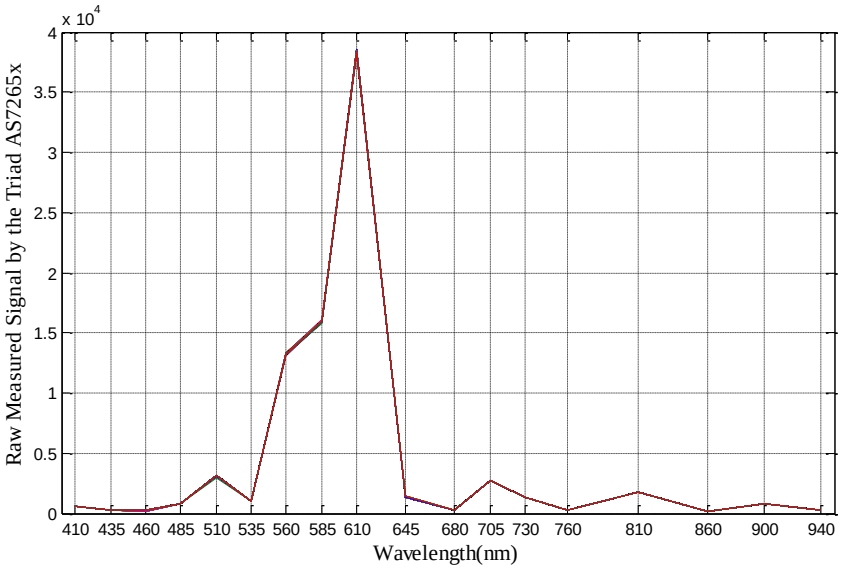


Figure 4.22: Raw intensity signal measured at 18 wavelengths for different concentration of glucose within albumin PBS solution.

5. Data Analysis and Glucose Prediction Models Using Machine Learning Techniques

This Chapter covers data analysis approaches and the machine learning algorithms that are used to predict glucose concentration based on measuring data at different wavelengths. In this Section, first the optimal wavelengths for accurate glucose measurements are identified. Then, different machine learning techniques are employed to create different glucose prediction models using the intensity data measured at the identified wavelengths (i.e. optimal wavelengths). In this Chapter, model hyperparameters are also optimized to results in maximum accuracy in predicting glucose concentrations, and reliably identify hypoglycemia and hyperglycemia.

5.1 Data Analysis and Modeling for Predicting Glucose Concentration within Distilled Water

This Section focuses on the statistical and machine learning techniques used to decipher the most accurate glucose concentration from intensity data associated with the glucose in distilled water solution.

5.1.1 Identifying Optimal Wavelengths for Glucose Prediction

The selection of the optimal wavelengths for glucose prediction is based on the intensity data measured by the Triad Sensor using 18 distinct wavelengths. The intensity data is converted from analog to digital form using an AS7265x integrated analog to digital converter. The intensity data for each of the 10 experiments is fed into Arduino IDE software and input into MATLAB™ for signal processing. The data set is filtered using a finite impulse response low pass filter with an order of 34 and a bandwidth of 100 mHz. Figure 5.1 illustrates how the optimal wavelengths for glucose prediction are identified based on the measured data.

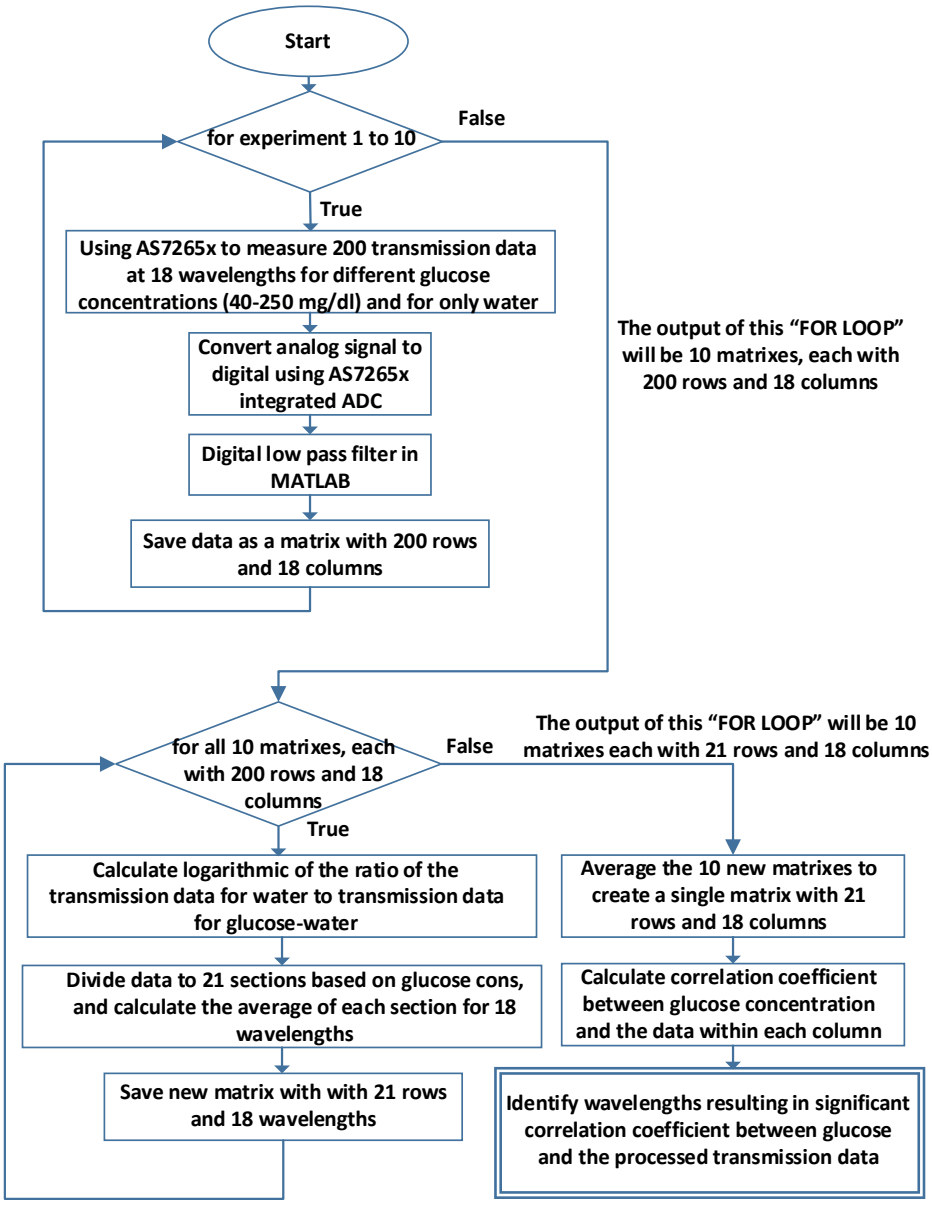


Figure 5.1: Block diagram showing procedures to identify optimal wavelengths for glucose measurements.

For each experiment and at each wavelength, each measured data point for distilled water is divided by each data point for the glucose-distilled water solution, resulting in normalized data, in the form of a matrix similar to Table 4-6. The logarithmic of the normalized data is then calculated for each of the 10 experiments. The intensity values are then averaged for each class for

each wavelength. For example, all the class 1 intensity data is averaged for each wavelength, all the class 2 intensity data is averaged for each, until all 21 classes are averaged for each wavelength. This results in a reduced matrix with 21 rows (corresponding to the average values for each class) and 18 columns (corresponding to each wavelength), and this is done for each of the 10 experiments. Finally, the average of all 10 matrices resulting from 10 experiments is calculated, resulting in a single matrix with 21 rows and 18 columns. The data associated with the final matrix is plotted as average of the normalized intensity vs wavelengths as illustrated in Figure 5.2. The data in Figure 5.2 includes 21 lines, one for each class, and illustrates the glucose detection measurements as a function of wavelength.

Pearson correlation coefficient is used to measure the linear association between glucose concentrations and the average values of the normalized intensity at each wavelength. The intensity of light measured at four wavelengths of 485, 645, 860 and 940 nm results in the highest correlation coefficient of 0.98. These wavelengths are selected as the optimal wavelengths to measure glucose concentrations [160].

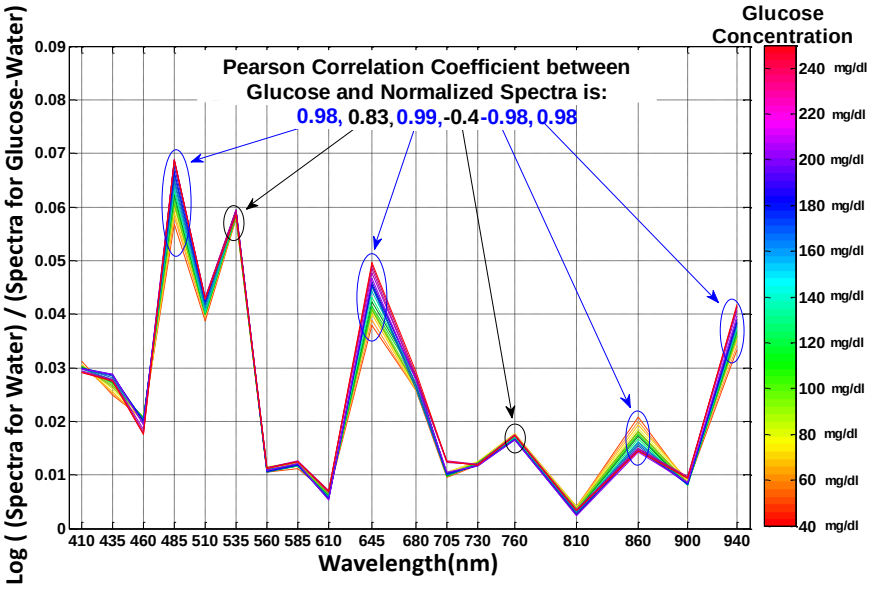


Figure 5.2: Average of Normalized Spectra for Different Glucose Concentration [160].

5.1.2 Glucose Prediction Using Machine Learning Techniques

The glucose concentration values used in this study range between 40 and 250 mg/dL and are defined by a vector y, with one row and 2,000 columns. The four wavelengths identified as optimal wavelengths (485, 645 and 860 and 940 nm) are based on Pearson correlation coefficients above 0.98 between the glucose concentration values and the average values of the normalized intensity at each wavelength. The light intensity data associated with the 4 optimal wavelengths is represented as a matrix X with 4 rows and approximately 2,000 columns. The 2,000 columns represent 200 data points per experiment, and the repeat of each experiment 10 times. Each row includes the intensity data measured using one of the four optimal wavelengths, and each column corresponds to a specific glucose concentration. A standardization method is used to transform all the values in each row of matrix X into a distribution with a mean of 0 and a standard deviation of 1. Data standardization results in similar scale for all variables, and this can help to increase the convergence speed of the learning algorithms and improve the model performance.

Five different machine learning models are trained using classification and regression methods as outlined in Figure 5.3, and are compared in terms of their ability to generate accurate glucose predictions. The regression methods consist of multiple linear regression (MLR) [22] and feed-forward neural network (FFNN) [25] and the classification methods include k-nearest neighbor (KNN) [161], decision tree (DT) [162] and support vector machine (SVM) [163].

In machine learning, hyperparameters need to be optimized to improve the model performance. In this study, a finite set of values are specified to a given hyperparameter, and a 10-fold cross validation (10-fold CV) method is used to find the best value for the hyperparameter. Based on the 10-fold CV, the data set is randomly split up into 10 “folds,” where each is roughly the same size. A machine learning model is trained using data from the first 9 folds (the training

set), and the performance of the model is evaluated on the 10th fold (the validation set). This process is repeated until each of the ten folds has served as a validation set, while the rest have served as a training set. The average prediction accuracy from the 10 different validation sets is calculated for each value of the hyperparameter. The averaged results are then interpreted to select the best value of the hyperparameter that results in the best fit to the data set. This process is repeated for each of the 5 models in Figure 5.3.

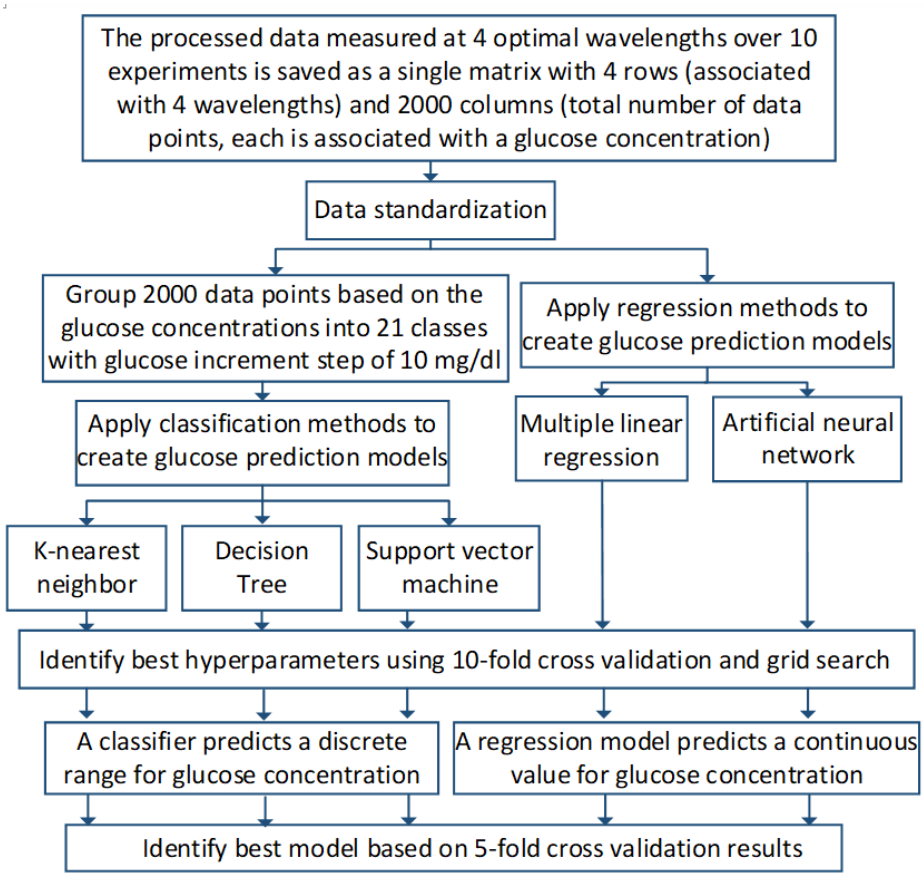


Figure 5.3: Block Diagram Showing Glucose Prediction Approach.

After the performance of each model is improved by finding the optimum hyperparameter, the 5 models are compared to identify the best model. A 5-fold cross validation method is used to evaluate the performance of each model by measuring the average of the prediction accuracy achieved on 5 different validation sets. The size of each validation set is 20% of the total data

points and the remaining 80% of the data is used as a training set. The model is evaluated using a validation set (or unseen data) in order to estimate the ability of the trained model to generalize from a training set to previously unseen data.

In this work, Python is used to train and optimize the multiple linear regression and classification models using the Scikit Learn Library, which contains efficient tools for machine learning. The neural network regression model is also trained and optimized in Python using the NumPy Library, which assists with calculations in N-dimensional arrays.

5.1.2.1 Multiple Linear Regression

The multiple linear regression (MLR) method is used to predict the value of an output target based on a linear combination of scaled input variables. The input variables for this work are the measured intensity at the four optimal wavelengths (denoted by x1, x2, x3 and x4), and the output is the glucose concentration predicted value (denoted by $\hat{y}$). The relationship between the input and output variables is represented using a regression equation $\hat{y}^{(i)} = \theta_0 + \theta_1 x1^{(i)} + \theta_2 x2^{(i)} + q_3 x3^{(i)} + \theta_4 x4^{(i)}$, where $\hat{y}^{(i)}$ refers to the glucose predicted value for the i^{th} data sample. The objective of the model is to find the best parameters (θ) that will minimize the mean squared error (MSE) value. The error term, or residual, is the difference between the predicted ($\hat{y}$) and the actual (y) glucose concentration values. The most common methods used to find the optimal values of the parameters that will minimize the error are (1) ordinary least squares which uses linear algebra operations to compute the model parameters, and (2) iterative optimization, such as gradient descent, which gradually updates the model parameters until the prediction error is at a minimum value. The accuracy of a trained MLR model is evaluated using the performance metrics such as root mean squared error (RMSE) and coefficient of determination (R^2).

The MLR model can be an appropriate model for accurate predictions if the following requirements are met: (1) a linear relationship between the input and output variables exists, (2) there is little or no multicollinearity between the input variables, (3) there is a normal distribution of the residuals, and (4) homoscedasticity of the residuals exists [164], [165]. Otherwise, non-linear regression is an alternative. The first requirements can be checked by (1) creating a scatter plot of the output vs each input variable and assessing the linearity (1st requirement), (2) computing the correlation coefficients between any two input variables to assess mutlicolinearity (2nd requirement), and (3) creating a residual histogram to assess the normality of the distribution (3rd requirement). The last requirement, homoscedasticity, can be checked by creating a scatter plot of the residuals vs predicted outputs (4th requirement). The residuals should be symmetrically and randomly distributed around the zero line on the plot of residuals vs predicted outputs, and a deviating trend to the data should not exist.

A further clarification of the two last requirements is illustrated in Figure 5.4, which includes the residual plots for the MLR model that is trained using the training set (80% of the total data) and is evaluated using the test set (20% of the total data). Figure 5.4 shows that the residuals have an apparent random dispersion around the x axis. Although there is not a clear trend, it appears that the residuals associated with the higher glucose predictions have a slight tendency to spread farther away from the x axis. This indicates that the relationship between the input variables and the output target is not perfectly linear. From the right side of the plot, the histogram of the residuals has the shape of a normal distribution with the mean close to zero, but the R^2 value is 0.91, which indicates that there is room to improve the accuracy of the glucose model by using non-linear regression algorithms such as neural networks [160]. Neural networks is generally

expected to give a better fit to the data set compared to MLR, and is evaluated in the next Sub-section.

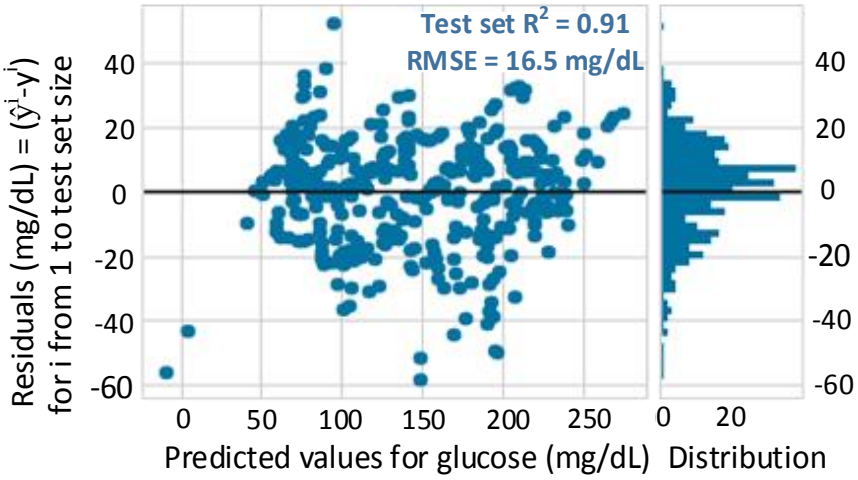


Figure 5.4: MLR residuals vs glucose predictions (left side plot) to check for homoscedasticity, and the residual histogram (right side plot) to check for the normality of the distribution.

5.1.2.2 Feed-forward Neural Network for Regression

Artificial neural networks is a subset of machine learning and uses a group of algorithms to solve both linear and non-linear problems. There are different types of neural networks, and each has its own structure and application, such as feed-forward neural network (FFNN), recurrent neural network (RNN), convolutional neural network (CNN), hybrid neural network, etc. [166], [167]. In this study, we select FFNN for the purpose of predicting glucose concentrations.

FFNN consists of multiple processing units called neurons which are arranged in different layers. The network generally consists of an input layer, one or more hidden layers and an output layer. Each layer contains neurons which are linked together to model the data and make a prediction at the output layer. In this study a neural network with one hidden layer is trained as illustrated in Figure 5.5. The input layer is denoted as matrix X (see Figure 5.5 (a)) and includes a

total of ‘m’ data samples and four variables (x1, x2, x3 and x4). The four variables coincide with the measured intensity at each of the four optimal wavelengths. The number of neurons in the hidden layer is tuned using a grid search method as explained at the end of this Subsection, and is determined to be 13. The number of neurons at the output layer is equal to 1, and is the predicted glucose concentration, denoted by the vector $\hat{y}$.

The weight matrix for the hidden layer and the output layer are denoted by $W^{[1]}$ and $W^{[2]}$, respectively (see Figure 5.5 (b)). The training of the neural network model starts by a random initialization of the weights ($W^{[1]}$ and $W^{[2]}$) and biases (b1 and b2). The values for $W^{[1]}$ and $W^{[2]}$ are set between -0.5 and 0.5, and the values for b1 and b2 are both set to zero. The input data in the form of the matrix X is multiplied by the weights ($W^{[1]}$) that are assigned to the hidden layer. The resulting matrix from this step is added to the b1 bias and this creates a matrix in the hidden layer that is denoted by $Z^{[1]}$ (see Figure 5.5 (c)). Then, each neuron in the hidden layer takes part in the activation function calculations. In this network, the hyperbolic tangent (Tanh) function is used as the activation function for the hidden layer. The Tanh function performs a mathematical operation on $Z^{[1]}$ and transforms it into the matrix $A^{[1]}$ which is the output of the neurons in the hidden layer ($A^{[1]} = \text{Tanh}(Z^{[1]}) = (1-e^{-Z[1]}) / (1+e^{-Z[1]})$) as illustrated in Figure 5.5 (d).

Similar calculations are required at the output layer. The output of the hidden layer $A^{[1]}$ is multiplied by $W^{[2]}$ and added to the bias b2 to create a matrix denoted by $Z^{[2]}$ as illustrated in Figure 5.5 (e). The matrix $Z^{[2]}$ is then plugged into the rectifier linear unit ReLU, an activation function for the output layer, which results in $\hat{y}$ (see Figure 5.5 (f)). The ReLU function returns 0 if $Z^{[2]}$ is negative and for any positive value of $Z^{[2]}$ it returns that value back ($\hat{y} = \text{ReLU}(Z^{[2]}) = \max(0, Z^{[2]})$) [160].

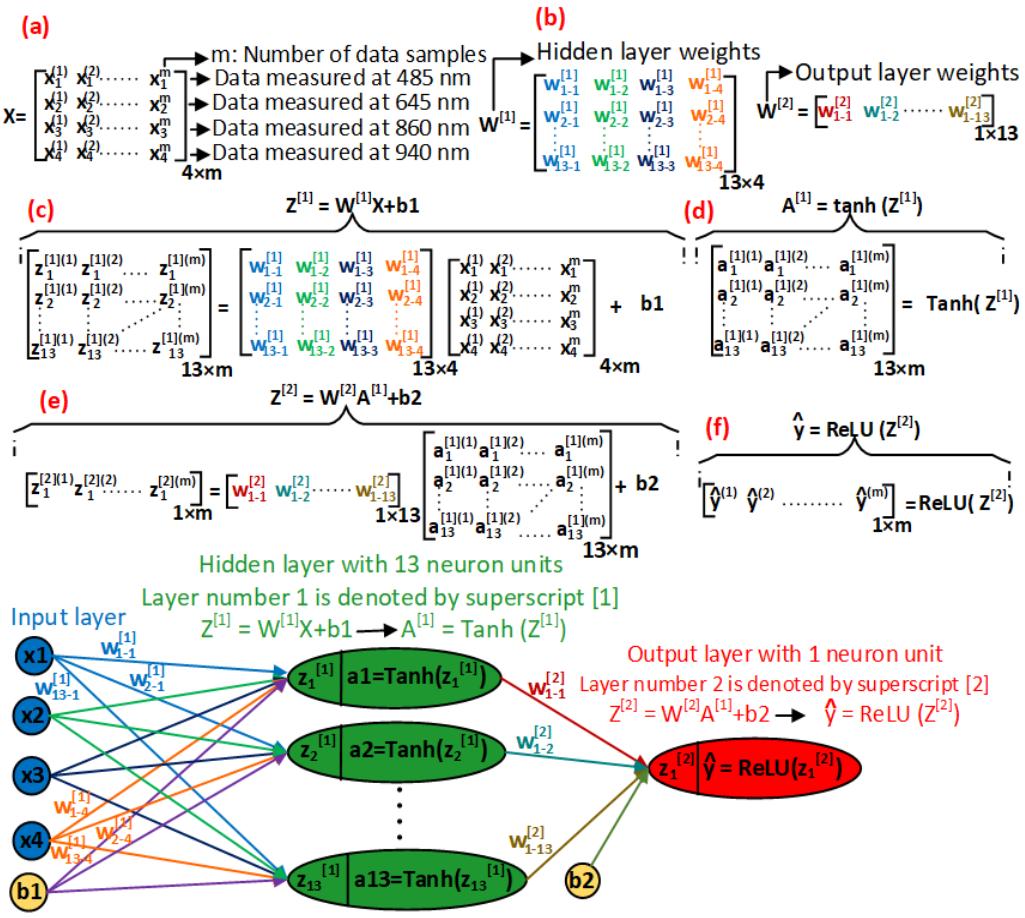


Figure 5.5: Structure of the Neural Network with one hidden layer [160].

To train the network, a back-propagation training principle is used in conjunction with an optimization method called “mini-batch gradient descent (MBGD) with momentum”. The back-propagation method is used to calculate the error associated with each neuron, and to adjust the weights ($W^{[1]}$ and $W^{[2]}$) and the biases (b1 and b2) using MBGD optimizer during an iterative process. There is another parameter named “learning rate” that is used to control the amount of changes in the weights and biases. The weights and biases repeatedly update until the network converges to an acceptable error. The error is measured using an error function that is also known as a “cost function” and can be computed as described in Eq. (5.1) .

$$Cost = \frac{1}{2m} \sum_{i=1}^{m} (\hat{y}^{(i)} - y^{(i)})^2 \qquad (5.1)$$

where $\hat{y}^{(i)}$ and $y^{(i)}$ are the predicted and real values of the glucose concentration for the i^{th} sample, respectively, and m is the number of samples.

The mini batch gradient descent method splits up the training set into smaller sets that are called mini batches. We use mini batches with the size of 64, and the total number of mini batches (denoted by n) is calculated by dividing the number of training samples by 64. When the algorithms run through all the mini batches, this is called one epoch, and for every epoch, the weights and biases are updated per each mini batch. Therefore, there will be a total of n updates for weights and biases at each epoch. The process of updating weights and biases are repeated during multiple epochs. In this work, the number of epochs is set to 5,000 since after 5,000 epochs there is no significant change in the error and thus the training can be stopped.

The MBGD optimizer exhibits benefits over other optimization algorithms such as stochastic gradient descent (SGD) and batch gradient descent (GD). SGD uses one data point to compute gradients, and updates the weights for every single data point, whereas (GD) uses all data points to compute gradients and there is a single update per all data points. MBGD falls somewhere between SGD and GD since it uses mini-batches (a small portion of the data points) to update weights. This makes MBGD computationally more efficient compared to GD, and more robust in convergence (or faster in learning) compared to SGD.

To speed up the training process for MBGD, the momentum technique is also used. In MBGD with momentum, the weights and biases are updated using the exponentially weighted average of the gradients (denoted by V_{dw} and V_{db}). The momentum parameter (denoted by beta) is usually given a value of 0.9. The process of training the neural network using MBGD with momentum, is shown as a flowchart in Figure 5.6 [160].

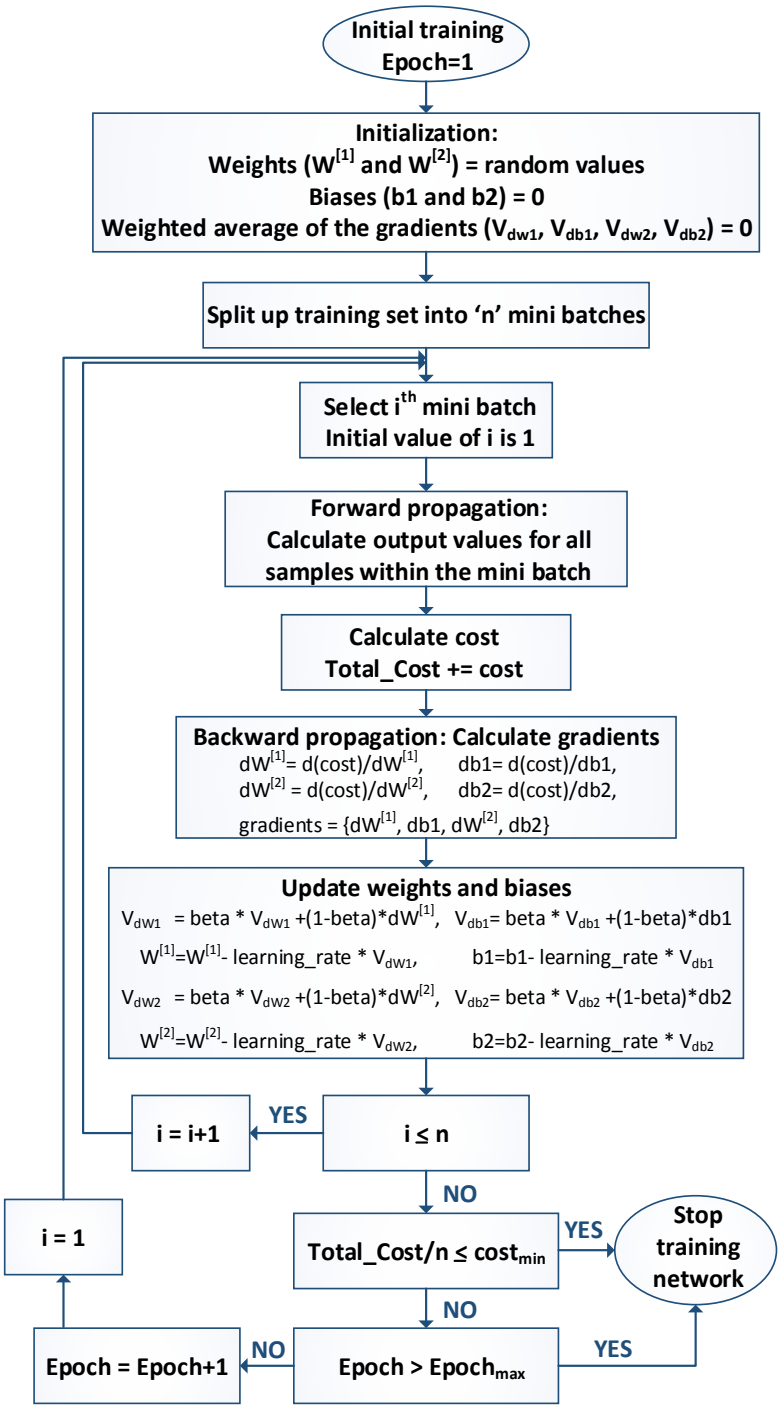


Figure 5.6: Training flowchart for feed-forward neural network based on MBGD with momentum [160].

Different methodologies have been used to optimize neural networks, and to identify the optimal set of hyperparameters such as "learning rate" and "hidden unit size," as reviewed in [168],

[169]. We use a 10-fold cross validation method and a grid search method to tune the learning rate and the hidden unit size. The grid search is a process in which a set of values are given to each parameter and then different models are trained for different combinations of the two parameters. Each trained model is then evaluated using 10 different validation sets to identify the best combination of the two parameters that results in a more accurate prediction. Figure 5.7 illustrates a 4×7 optimization grid in which each cell is the average of the root mean square error (RMSE) achieved on 10 different validation sets. For a given learning rate, the average value of RMSE decreases with increase in the size of the hidden layer units. We select 13 and 7×10^{-4} as the optimal values for the hidden unit size and the learning rate, respectively [160]. This combination of the two parameters results in a neural network architecture with the smallest possible size of hidden units and a low enough RMSE.

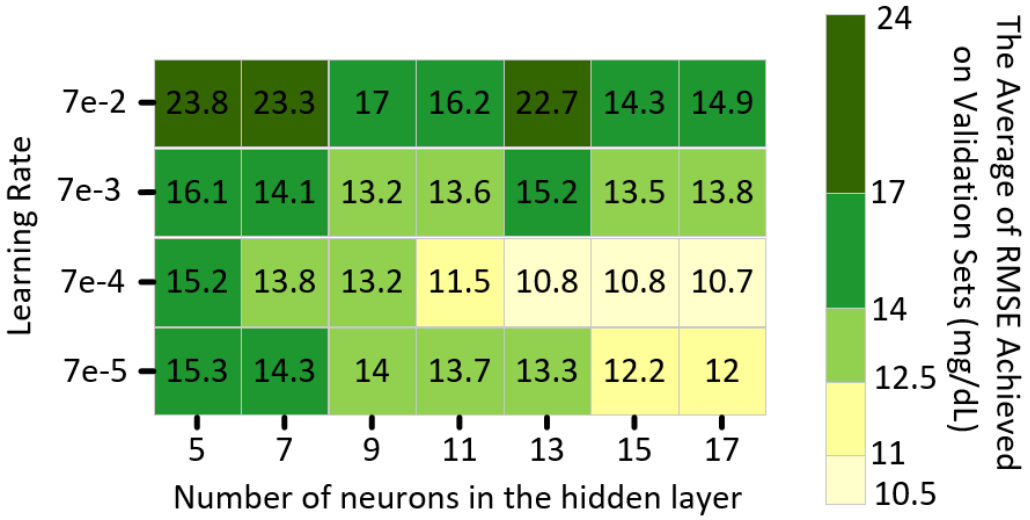


Figure 5.7: Tuning hyperparameters of the neural network model using 10-fold cross validation and grid search method [160].

The FFNN model is created based on the optimal values of hyperparameters (hidden unit size = 13 and learning rate = 7×10^{-4}), and is evaluated using the test set. Figure 5.8 shows the test set RMSE vs training epochs. As the number of epochs increases, the more number of times

weights and biases are updated. During training epochs, optimal set of weights and biases are identified that result in lowest RMSE. The final RMSE after 5,000 epochs is equal to 10.8 mg/dL.

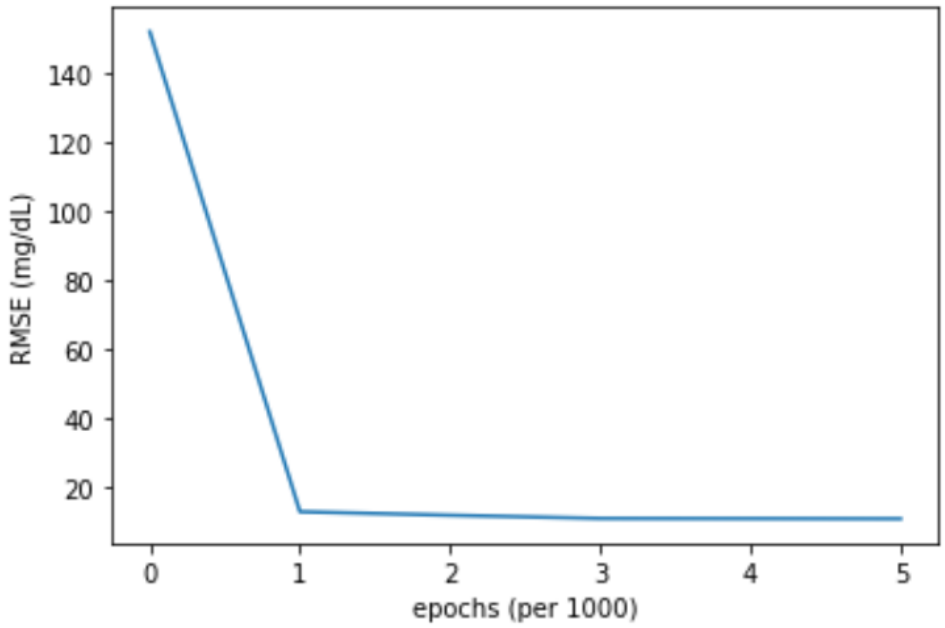


Figure 5.8: Test set RMSE vs training epochs for FFNN with learning rate = 0.0007 and hidden unit size = 13.

5.1.2.3 K-Nearest Neighbor Classifier

K-nearest neighbor (KNN) is a method that classifies unlabeled data points based on their similarity to other data points in close proximity. The similarity between two data points can be measured using different distance metrics such as Euclidean, Minkowsky, Manhattan, city-block, and Chebyshev [161], [170]. Figure 5.9 illustrates the idea of classifying an unlabeled data point based on the KNN method for an arbitrary data set with two features (denoted by x1 and x2) and only 2 classes (depicted by star and triangle). To classify the unlabeled data point (depicted by a circle), the distances between the unlabeled data point and the labeled data points are computed. According to the distances, the number of k observations in the data set that are nearest to the unlabeled data point are identified. Then, a voting rule is used to classify the unlabeled data.

Two common and simple voting rules to classify the unlabeled data points are (1) majority voting rule and (2) inverse distance weighting rule. Both voting methods are illustrated in Figure 5.9.

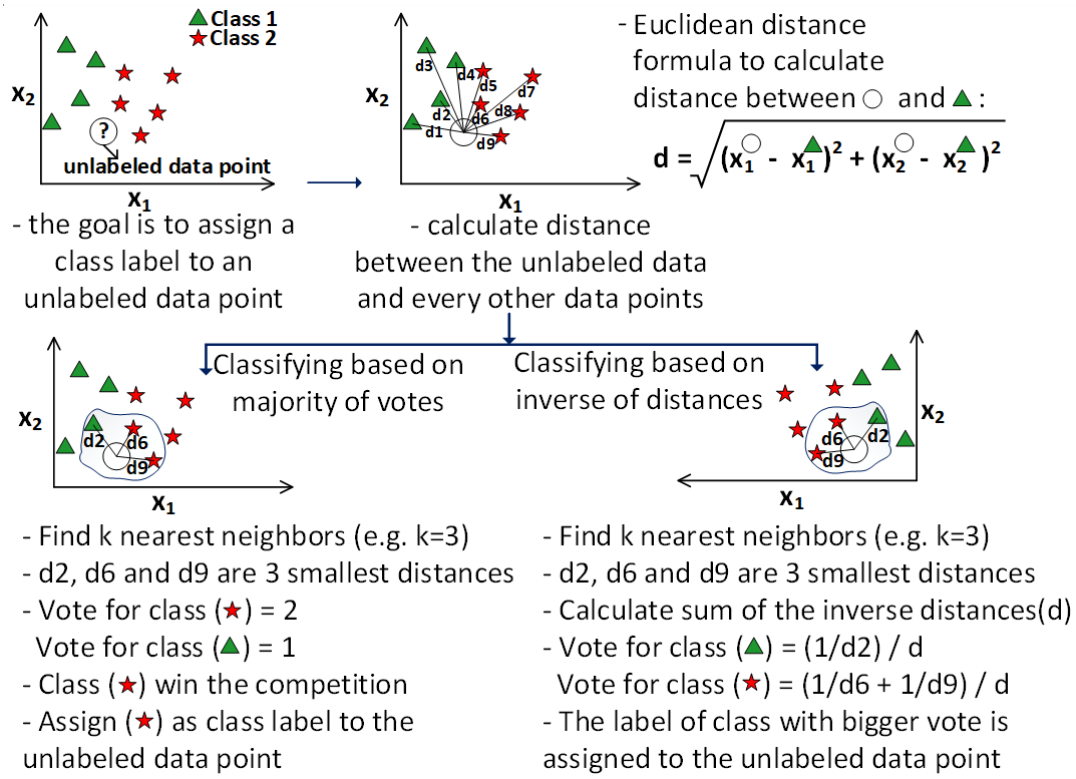


Figure 5.9: Classification of unlabeled data points using k-nearest neighbor classifier based on two different rules: majority voting rule and inverse distance weighting rule [160].

In the majority voting rule, the most popular class labels from the k-nearest neighbors (e.g. k=3) are identified, where the distance is calculated based on the Euclidean formula. The votes of the k-nearest neighbors are based on class label and each vote has equal weight. This means there exists an equality of the k-nearest neighbors in the process of voting regardless of their distance from the unlabeled data point [171] (e.g. 2 votes for star and 1 vote for triangle). When using the majority voting rule, it is difficult to accurately classify the unlabeled data point if the value of k is large and if the neighbors are from multiple classes. To address this, the inverse distance

weighting rule can be used to improve the prediction accuracy of the classifier. In this case, the votes of the k-nearest neighbors are weighed according to their distance from the unlabeled data point. This is done by calculating the inverse of the Euclidean distance between the unlabeled data and each of the k-nearest neighbors, and then normalizing the value with the summation of all the inverse distances. The class with the maximum normalized value wins the vote and the associated label is assigned to the unlabeled data point. Thus, the closest neighbors have much more decision power in determining the classification for the unlabeled data point.

The optimal value of k is identified using the ten-fold cross validation method. Figure 5.10 illustrates the Jaccard index as a function of k for 10 different validation sets. The average Jaccard index for all ten validation sets is plotted as a solid line for each value of k-nearest neighbors.

A value of 6 is selected as an optimal value for the number of k-nearest neighbors. The range of changes in the Jaccard index is less varied for k=6 compared to other values of k. It means that the KNN classifier with k=6 can perform well on different data sets as well as result in an acceptable prediction accuracy with the average value of 0.85. By selecting k=6, the risk of overfitting associated with a lower value of k (1, 2, 3) is minimized, and the robustness of the classifier against outliers within a data set is improved.

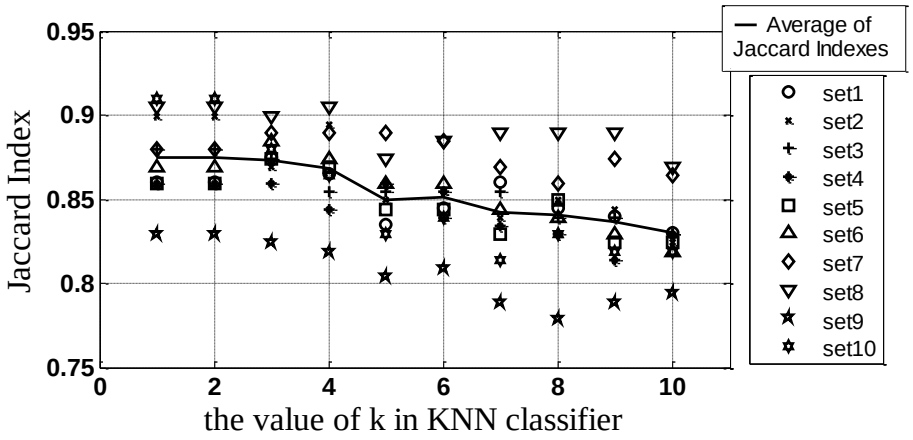


Figure 5.10: Jaccard index vs k parameter in KNN for 10 different validation sets [160].

5.1.2.4 Decision Tree Classifier

The decision tree (DT) model is a tree shaped diagram that represents all possible decision paths to classify data points, with a structure that includes decision nodes, branches and leaf nodes. A decision node represents a condition that is applied to an attribute, a branch corresponds to the result of a condition on the attribute, and a leaf node (or terminal node) represents a class label as the final prediction. Based on our data, there are four attributes (x1, x2, x3, and x4) that correspond to the intensity data measured at the four optimal wavelengths [160].

The decision tree in Figure 5.11 is used to illustrate the data classification process. The top node in the decision tree represents the entire data set S1 which is partitioned to create more nodes and leaves. The data set S1 is partitioned into S2 and S3 based on a condition that is applied on an attribute, such as x1>0.4, where 0.4 is an intensity value for the attribute x1. Attribute conditions need to be evaluated to identify the best condition. An optimal attribute condition is one that divides a single node into two nodes that are as pure as possible. The purity of a node can be measured by computing the entropy of the data set S within that node as defined in Eq. (5.2) [162].

$$Entropy(S) = -\sum_{i=1}^{C} p_i \log_2(p_i) \qquad (5.2)$$

where p_i is the probability of data points that belong to the i^{th} class and are within the subset S, and c is the total number of classes. The purest node has an entropy equal to 0 and this will occur if all of the data points within the node all belong to the same class. The maximum entropy value is dependent on the value of c, and a node with a diverse class label distribution has the highest impurity or maximum entropy. Entropy measures the purity of a single node, and to select the best attribute, the purity of the entire split needs to be computed using metrics such as Information Gain. Information Gain is the entropy before the split minus the weighted entropy after the split. The weighted entropy represents the entropy for each node multiplied by a proportional factor

based on the number of data points associated with each node. Figure 5.11 includes the information gain calculations for the split based on the attribute condition x1>0.4 [160].

Among all possible attribute conditions, the best attribute is the one with the highest information gain. This attribute is associated with a split in the dataset that results in an overall decrease in entropy.

The process of splitting a node continues until it reaches a stopping criterion. Examples of these criteria are when the entropy of a node is less than a pre-specified threshold, or when the depth of a tree reaches a pre-specified level. When this occurs, the nodes become leaves, or terminal nodes of a tree. It is possible that leaf nodes include a subset of data points with different class labels. When this occurs, the predominant class label in the subset is identified and is represented as the class label predicted by the leaf node.

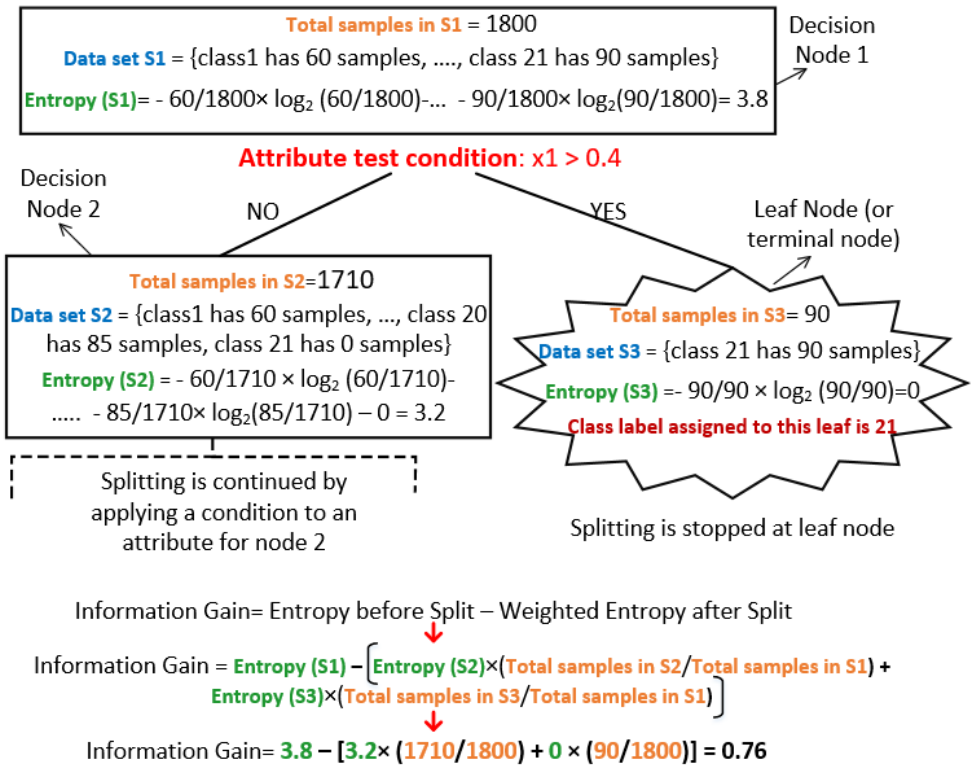


Figure 5.11: Data classification with a decision tree [160].

The predictive performance of a DT model also depends on the DT hyperparameter values. These include (1) the maximum depth of a tree, denoted as max_depth, (2) the minimum number of samples required to split a decision node, denoted as min_samples_split, and (3) the maximum number of leaves in a tree, denoted as max_leaf_nodes [172]. These hyperparameters restrict the depth and width of a tree, and thus prevent the tree from extending beyond a certain level of complexity. More importantly, they improve the generalizability of the DT classifier to make an accurate prediction based on the test data.

The three hyperparameters for the decision tree classifier are included in Table 5-1, along with a list of test values. The decision tree classifier is built for each possible combination of the three hyperparameter values, and is evaluated using the 10-fold CV method. The assessment of the prediction accuracy for each classifier is computed by averaging the values of the Jaccard index for the 10 validation sets. The Pearson correlation between each possible hyperparameter and the average Jaccard index is also calculated. A *max_leaf_node* resulting in the highest correlation coefficient of 0.7 has a significant impact on the prediction accuracy of the DT classifier. Based on the results in Table 5-1, a prediction accuracy of more than 75% occurs when the *min_sample_split* is not more than 10, the *max_depth* is at least 10, and the *max-leaf-node* is at least 200. By zooming in on these hyperparameter critical limits, the optimal values for the *min_sample_split*, *max_depth*, and *max_leaf_node* parameters are determined to be 8, 12, and 210, respectively. These values help to ensure a DT classifier with a relatively small prediction error for both the training sets and the validation sets while limiting the complexity of the DT classifier.

Table 5-1: Hyperparameter tuning in decision tree using the results of 10-fold cross validation [160]

Hyperparameter	Tested Values	Pearson Corr	First maximum jaccard index ≈ .80	Second maximum jaccard index ≈ .76
min_samples_split	[5, 10, 15, 20, 25]	-0.2	Not more than 5	Not more than 10
max_depth	[5, 10, 15, 20, 25]	0.5	At least 15	At least 10
max_leaf_nodes	[100, 150, 200, 250, 300]	0.7	At least 250	At least 200

5.1.2.5 Support Vector Machine Classifier

Support vector machine (SVM) algorithms classify data by mapping the data into high dimensional feature space and by finding a separating hyperplane (or decision boundary) that partitions the samples into different classes. SVM algorithms use a kernel function such as linear, polynomial, sigmoid, and the radial basis function (RBF) to create a decision boundary. In this study, RBF is used as the kernel function due to its popularity as an effective kernel since it allows any shape associated with a decision boundary to make the distinction between different classes [163]. The two hyperparameters associated with the SVM classifier are (1) the kernel scale parameter denoted by γ (with values between 10^{-6} and 10^{2}), and (2) the regularization parameter denoted by C (with values between 10^{-2} and 10^{6}) [160].

Increasing the value of C increases the risk for overfitting, which means that the classifier model cannot generalize from the training set to previously unseen data (test set). When overfitting, the classification model fits the training set perfectly, but has poor fit on the test set, and this results in an unreliable prediction of class labels for the previously unseen data. On the other hand, lowering the value of C leads to more regularization which means that a more simple

decision function is applied on the training set. When the model is highly regularized, there is a risk of underfitting. When underfitting, the classification model is not fitting well to the training set and performs poorly on the test set [173] [174].

Increasing the value of γ leads to a greater curvature of the decision boundary (or separating hyperplane) and this increases the possibility of overfitting. Whereas, when the decision boundary is nearly linear (low curvature and small values of γ), the SVM model is simplified, resulting in a greater possibility of the underfitting. Thus, when γ is large, the prediction accuracy on a test set can be improved by decreasing the value of C (because decreasing the value of C minimizes the risk of overfitting). Likewise, for very small values of γ, the prediction accuracy on the test set can be improved by increasing the value of C (because increasing the value of C minimizes the risk of underfitting) [173]. The optimization of C and γ can be visualized by examining the results in Figure 5.12.

The optimal value of C and γ are identified using the grid search and 10-fold cross validation methods from the 6×6 optimization grid as illustrated in Figure 5.12. Each cell represents the average of the ten Jaccard indexes achieved on the validation sets (Figure 5.12 (a)) and training sets (Figure 5.12 (b)). The values of C=100 and γ=1 are identified as optimal values to ensure that a decision function is appropriately fitted on the training set and overfitting or underfitting is avoided on the training set. This combination of C and γ results in an accurate classification of data points within both the training and validation sets without adding unnecessary complexity to the SVM classifier.

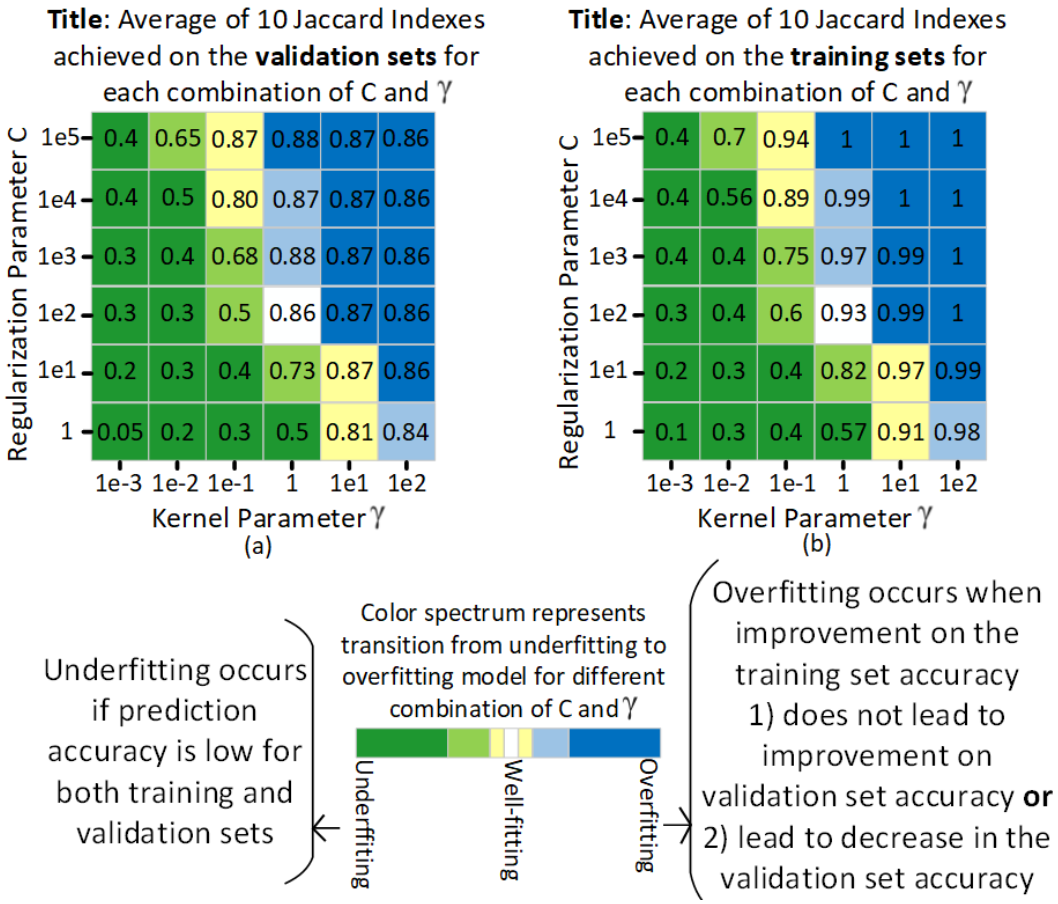


Figure 5.12: Average of ten Jaccard indexes achieved on the (a) validation sets and (b) training sets for each combination of C and γ [160].

5.2 Data Analysis and Modeling for Predicting Glucose Concentration within Phosphate Buffered Saline Solution Supplemented with Albumin

This Section focuses on the statistical and machine learning techniques used to decipher the most accurate glucose concentration from intensity data measured for the glucose in a phosphate buffered saline (PBS) solution supplemented with bovine serum albumin (BSA).

5.2.1 Identifying Optimal Wavelengths for Glucose Prediction

The selection of the optimal wavelengths for glucose prediction is based on the intensity data measured by the Triad sensor using 18 distinct wavelengths. The intensity data for each of the 15 experiments (3 different albumin concentrations repeated 5 times) at each wavelength is

imported into MATLAB™, and is filtered using a finite impulse response low pass filter with an order of 34 and a bandwidth of 100 mHz. The remainder of this Section explains how the optimal wavelengths are identified by analyzing the five 5 g/dL albumin data sets for each of the 18 wavelengths as the glucose concentration is diluted from 250 mg/dL to 40 mg/dL.

Each of the measured intensity data points associated with one of the glucose/albumin/PBS experiments and a specific wavelength are divided by the measured intensity data point associated with the albumin/PBS solution (without glucose). A repeat of this calculation for each of the 18 wavelengths, results in normalized data, in the form of a matrix with 360 rows (number of data points) and 18 columns (number of wavelengths). The logarithmic of the inverse of the normalized data is then calculated for each of the 5 matrices (one for each of the 5 experiments). Then, the data within each matrix is classified into 21 classes, within the glucose concentration range between 40 mg/dL and 250 mg/dL. Each class includes data that corresponds to a glucose concentration within a discrete 10 mg/dL range. For example, class 1 includes data associated with glucose concentration between 40 and 50 mg/dL. The data points are then averaged for each class for each wavelength. This results in a reduced matrix with 21 rows (each row include the average values for a class) and 18 columns (corresponding to each wavelength). Since this is done for each of the five experiments, this results in 5 matrices, and each with 21 rows and 18 columns. The average of all 5 matrices resulting from 5 experiments is calculated. This results in a single matrix with 21 rows and 18 columns. The data associated with this matrix is plotted as average of the normalized intensity vs wavelengths as illustrated in Figure 5.13. The data in Figure 5.13 includes 21 lines, one for each class, and illustrates the glucose detection measurements as a function of wavelength when the albumin concentration is fixed to 5 g/dL throughout the 250 mg/dL – 40 mg/dL range.

The Pearson correlation coefficient is used to measure the linear association between the known glucose concentration values and the average values of the normalized intensity. The intensity of light measured at five distinct wavelengths (435, 485, 645, 860 and 940 nm) is associated with a correlation coefficient of more than 90% as labeled in Figure 5.13. The remaining paragraphs in this Section describe how the albumin concentration (at 4 g/mL, 4.5 g/mL and 5 g/mL) affects the measured intensity for the 5 wavelength values identified so far as the optimum wavelengths for sources and detectors.

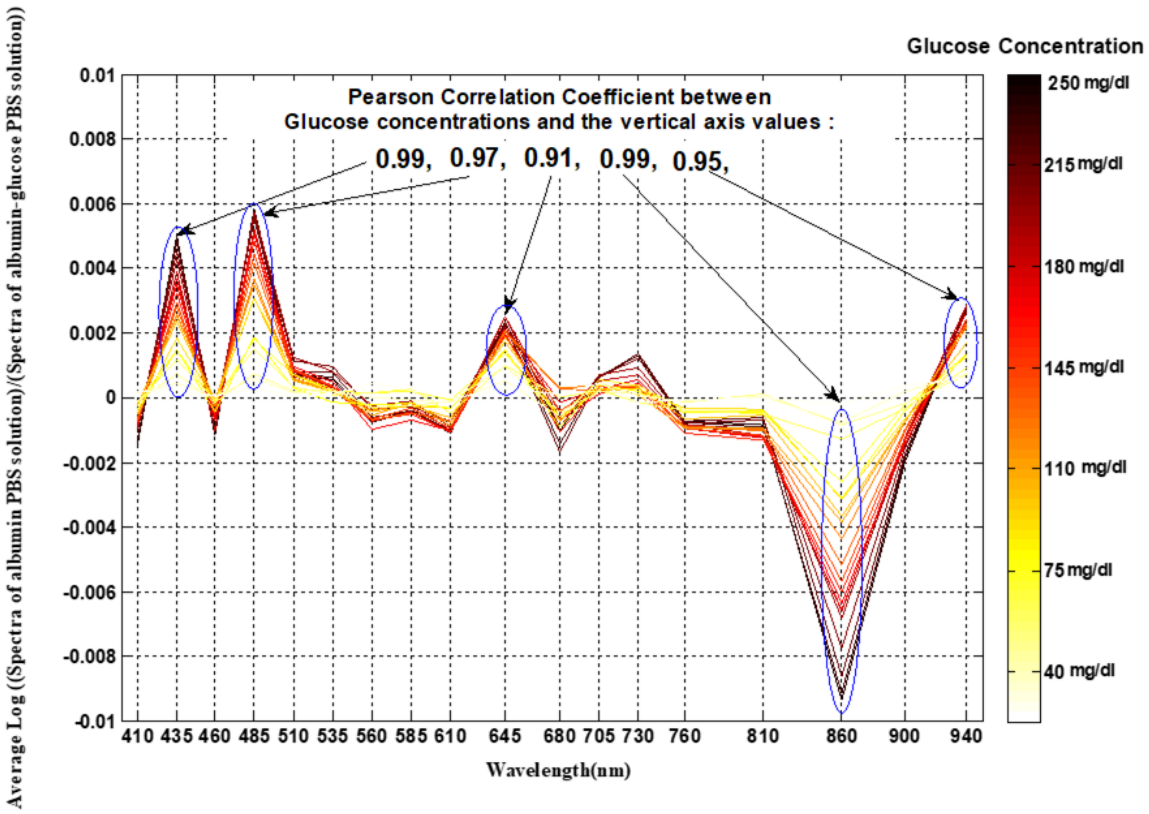


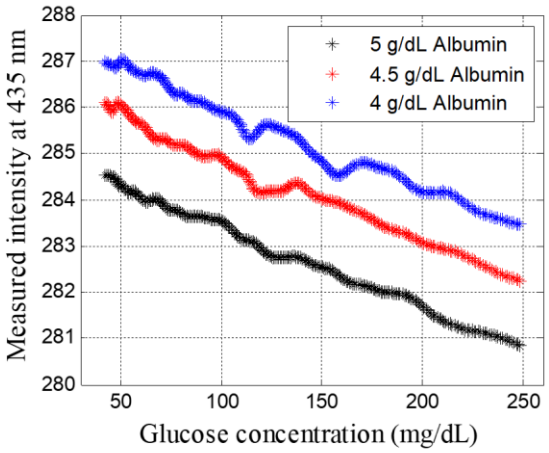
Figure 5.13: Average of normalized spectra for 40 mg/dL to 250 mg/dL glucose range for the 5 g/dL albumin PBS solution.

As mentioned before, the intensity data points measured for one out of five individual experiments consist of a matrix with 360 rows and 18 columns. Each row includes intensity data measured at each of the 18 wavelengths and each column includes 360 concentration values

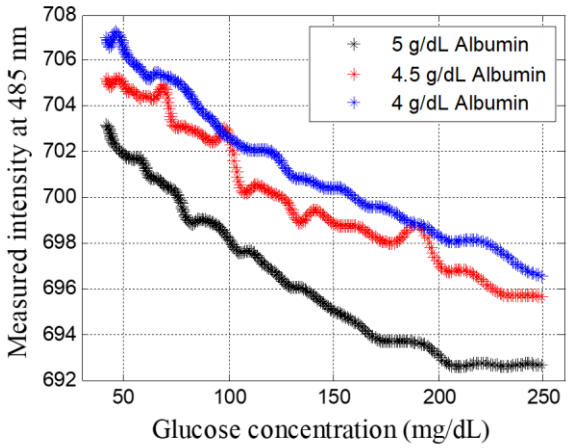
between 250 and 40 mg/dL. The average of 5 matrices resulting from 5 experiments for a given albumin level is calculated. The average of the 5 matrices for the 5 mg/dL albumin experiments results in a single matrix (denoted by matrix A) with 360 rows and 18 columns. The average of 5 matrices from the 5 experiments in which albumin concentration is set to 4.5 g/dL is also calculated and results in matrix B, and the 5 matrices from the 5 experiments in which albumin concentration is 4 g/dL are also averaged to create matrix C. Figure 5.14 compares the intensity data as a function of glucose concentration for each matrix (A, B and C) for each of the 5 wavelengths identified as having the highest Pearson coefficients. Figure 5.14 (a) includes the intensity data measured at 435 nm for matrix A (5 mg/dL) in black, matrix B (4 g/dL) in blue, and matrix C (4.5 g/dL) in red. Figure 5.14 (b) – 5.13 (e) include the intensity data as a function of glucose concentration for the remaining four wavelengths of 485 nm, 645 nm, 860 nm, and 940 nm.

The Pearson correlation coefficient between the intensity data for each albumin value at each glucose concentration value is calculated. The average of the computed coefficients for each glucose concentration is then calculated for each wavelength, and the results of these calculations are 0.99, -0.96, -0.82, 0.73, and -0.86 for wavelengths of 438 nm, 485 nm, 645 nm, 860 nm, and 940 nm, respectively. A high correlation between intensity data and albumin concentration indicates that the measured light intensity are significantly affected by the albumin concentration even if there is no change in glucose concentration. In the other words, the measured light intensity for 2 solutions with same glucose concentration can be different due to different albumin levels, and this leads to errors in glucose concentration readings. Among the 5 identified wavelengths, the intensity data at 435 nm exhibits the highest correlation coefficient with albumin. Using data collected with 435 nm wavelength may lead to more error compared to other wavelengths. Thus, the intensity data measured using the 435 nm wavelength is not included as an appropriate source

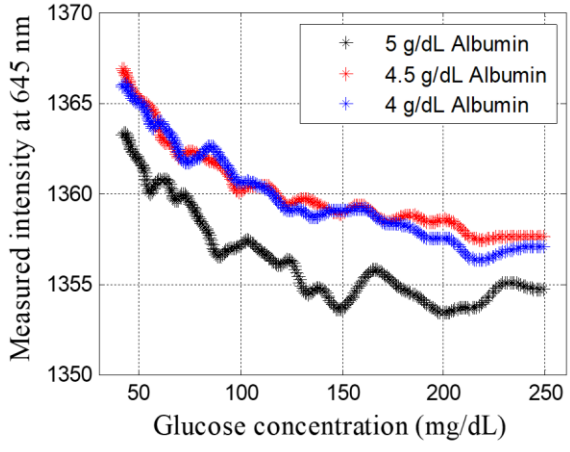
and detector for glucose measurements because the intensity data at this wavelength is much affected by albumin compared to the other four wavelengths. Based on this data, the four wavelengths of 485 nm, 645 nm, 860 nm, and 940 nm are selected as optimal wavelengths for which the measured intensity data can be used to extract accurate information about the glucose concentration. These four wavelengths are the same optimal wavelengths that were used for the predictions of glucose concentration in distilled water, as was described in Subsection 5.1.1 Identifying Optimal Wavelengths for Glucose Prediction.



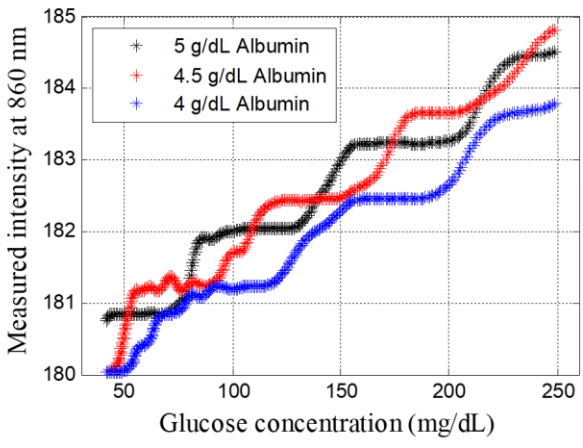
(a)



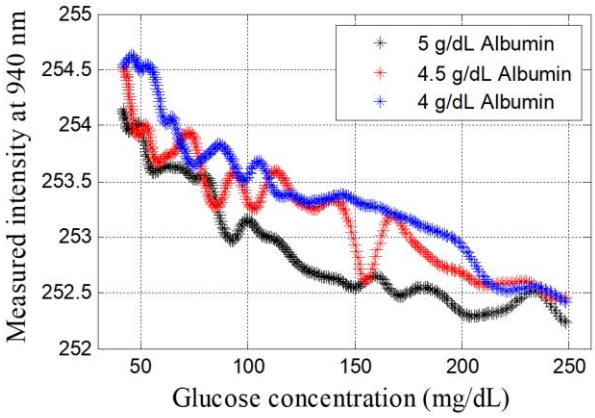
(b)



(c)



(d)



(e)

Figure 5.14: Measured intensity data vs glucose concentration for PBS solutions with albumin concentrations of 5 g/dL (black line), 4.5 g/dL (red line), and 4 g/dL (blue line), measured using wavelengths of (a) 435 nm, (b) 485 nm, (c) 645 nm, (d) 860 nm, and (e) 940 nm.

5.2.2 Glucose Prediction Using Machine Learning Techniques

Figure 5.15 illustrates the data frame with 5,233 rows (sample size) and four columns. The first column represents the index, the next four columns include the measured intensity at the four optimal wavelengths, the fifth column represents the glucose concentration associated with the individual measured data point or index, and the last column is an assigned class label. The dataset is arranged into three classes: class with label 1, 2, and 3 which includes data points that fall within the hypoglycemic (glucose cons. <70 mg/dL), normal (glucose cons. between 70 and 180 mg/dL), and hyperglycemic (glucose cons. >180 mg/dL) range, respectively.

The main Steps (1-4) for predicting glucose concentrations using the machine learning techniques are illustrated in the flow chart in Figure 5.16. The strategy is to optimize and train a support vector machine (SVM) classifier to accurately predict a class (hypoglycemic, normal or hyperglycemic range) for a given test data point, and then according to the predicted class, a feed-forward neural network (FFNN) regression model predicts the glucose concentration for the test

data point. There are three FFNN models, and each one is used to predicting the glucose concentration within one of the three classes.

Index	Wave1: 485 nm	Wave2: 645 nm	Wave3: 860 nm	Wave4: 940 nm	Glucose Cons (mg/dL)	Class Lable	
1	2.1	-0.19	-0.76	1.6	40	1	Hypoglycemic range
2	-0.41	-0.37	-1.79	-0.73	40	1	
⋮	⋮	⋮	⋮	⋮	⋮	⋮	
1,621	-0.23	-0.81	-1.34	1.02	69.9	1	
1,622	0.36	1.63	-0.73	0.17	70	2	Normal range
⋮	⋮	⋮	⋮	⋮	⋮	⋮	
4,304	-1.99	-1.48	0.75	-0.95	179.9	2	
4,305	-1.05	-0.91	1.14	-1.16	180.02	3	Hyperglycemic range
⋮	⋮	⋮	⋮	⋮	⋮	⋮	
5,233	-1.08	0.01	2.07	-0.54	250	3	

Figure 5.15: Standardized data-frame illustrating the indexing associated with the measured light intensity data for the range of glucose concentrations for each of the four wavelengths.

As part of the first Step, the "StandardScaler" function in Python is used to transform the data so that each feature (i.e. measured intensity data points at each of the four wavelengths) is centered on zero (has mean of zero) and has a standard deviation of 1. Data standardization improves the classification performance of each model and helps to increase the convergence speed of the gradients in neural networks. After data standardization, the data points are split up in a random fashion into two sets: dataset1 (85% of total) and dataset2 (15% of total).

In Step 2, the optimal values of hyperparameters in the SVM and FFNN models are identified by performing the 5-fold cross validation method on dataset1 (85% of total).

In Step 3, dataset 1 serves as the training set and is used to create a SVM classification model and 3 FFNN regression models based on the best hyperparameters identified in Step 2.

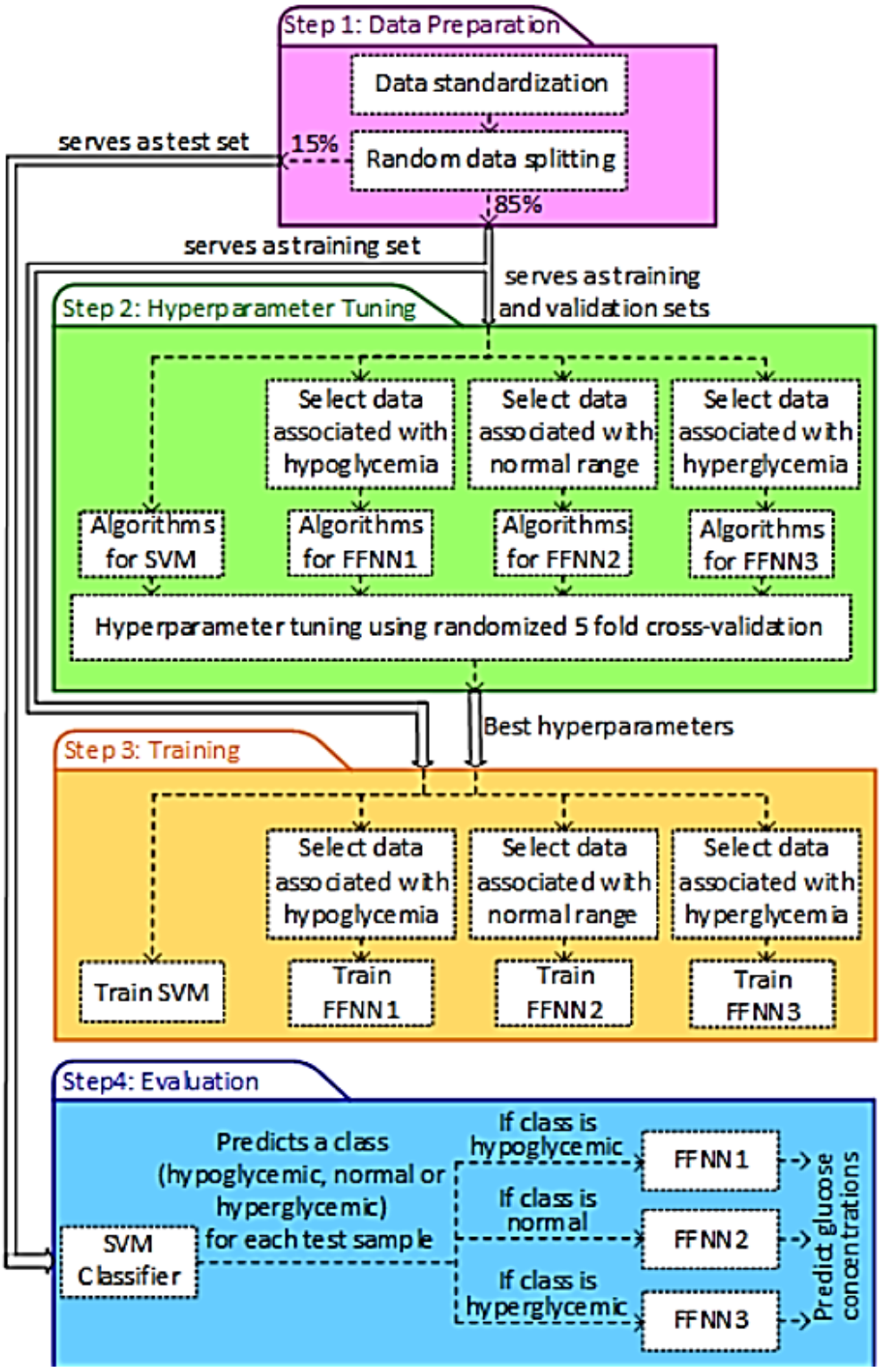


Figure 5.16: Machine learning approach to predict glucose concentrations based on the intensity data measured from glucose containing PBS solution supplemented with albumin.

In the last Step, Step 4, the glucose prediction accuracy of the models that are created in Step 3 are determined using dataset 2 (serves as test set and 15% of total data).

The next Chapter focuses on Step 2, and the results from Steps 3 and 4 are included in Section 6.2 Glucose Prediction Accuracy Results for Phosphate Buffered Saline Supplemented with Glucose and Albuminof Chapter 6.

5.2.2.1 Support Vector Machine Classifier for Classification of Samples into Hypoglycemic, Normal, and Hyperglycemic Ranges

The concept behind the SVM classifier is explained in Subsection 5.1.2.5 Support Vector Machine Classifier. The accuracy of the SVM classifier is dependent on the values of two hyperparameters, the regularization parameter (C) and the kernel scale parameter (γ). The radial basis (RBF) function is once again used as the kernel function since it is effective at making a distinction between different classes. The optimal values of the two hyperparameters are determined by performing the grid search and 5-fold CV methods using a 4x5 optimization grid as illustrated in Figure 5.17. First a set of values are given to each hyperparameter (e.g. n values for C, and m values for gamma) and then for each possible combinations of the two hyperparameters (total possible combination= n×m), a 5-fold CV method is performed on the dataset. Performing the 5-fold CV calculation on the dataset involves a random splitting of the dataset into 5 groups (known as folds). A model training and validation is repeated 5 times, and for each calculation, 4 of the 5 folds serve as the training set to train a model, and the fifth fold serves as the test set to evaluate the model. It means that during the 5-fold CV iterations, each of the 5 folds (or groups) has served as both a validation set and as a training set.

Figure 5.17 illustrates a 4×5 optimization grid in which a cell includes the average of the accuracy scores resulting from the 5-fold CV method. The 5-fold CV method for each of the n×m

combinations/iterations of C and γ results in n×m average accuracy scores, and these averaged results are then interpreted to select the best combinations of the two hyperparameters (C and γ). The optimal values for C and γ are identified as C = γ = 10 based on the 5-fold CV method. These values are selected to balance the effects of overfitting and underfitting. An increase in the value of C and γ leads to an increase in the prediction accuracy. However, when large values of C and γ do not significantly improve the accuracy score, the model is getting overfitted. Thus, the lowest possible values of the C and γ associated with the highest accuracy score are C = γ =10. These values result in the highest prediction accuracy while avoiding overfitting of the model, and result in an accuracy score of 0.99 as shown in Figure 5.17.

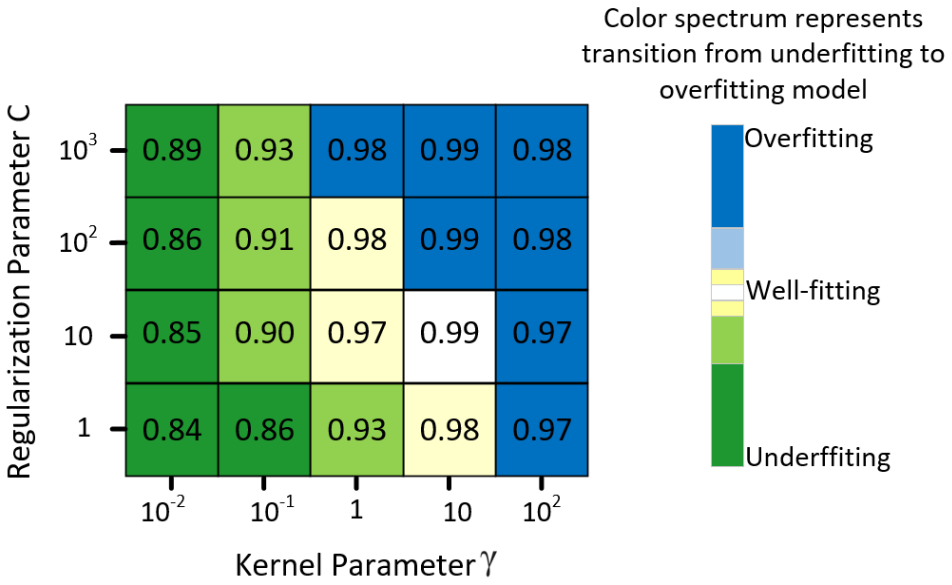


Figure 5.17: Average of accuracy scores resulting from 5-fold CV method for different combinations of C and γ.

The 5 folds or groups in the 5-fold CV method are generated in a random manner. Thus, repeating the 5-fold CV method may result in different accuracy scores for a given classifier. Thus, the prediction accuracy of the SVM classifier is evaluated by repeating the 5-fold CV method n

times (n=1 to 13). A boxplot of the accuracy scores is shown in Figure 5.18, and demonstrates that the SVM model with C=10 and γ=10 has the prediction accuracy of not less than 0.98 for each time that the 5-fold CV is performed, and for each validation set that is used to evaluate the model each time it is repeated. The green triangle shows the average of accuracy scores achieved on the 5-fold CV method after n iterations. The resulting average accuracy score is almost 0.99 for n times the CV method is repeated.

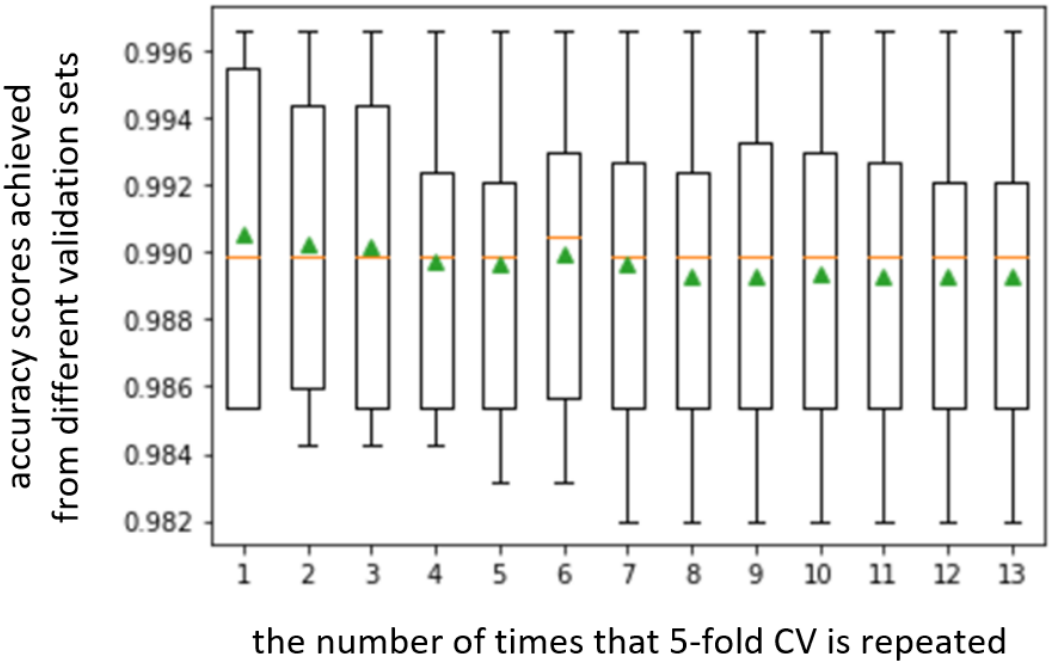


Figure 5.18: Box-plot of accuracy scores for the 5-fold CV method after it is repeated n times (n=1 to 13).

5.2.2.2 Three Feed-forward Neural Network Models Predicting Glucose Concentration within the Three Ranges: Hypoglycemic, Normal, and Hyperglycemic

The use of the feed-forward neural network (FFNN) method to create a regression model is explained in Subsection 5.1.2.2 Feed-forward Neural Network for Regression. To create a FFNN, the back-propagation training principle is used in conjunction with an optimization method called "mini-batch gradient descent (MBGD) with momentum".

The measured data points are grouped into three classes: hypoglycemic, normal, and hyperglycemic as demonstrated above in Figure 5.15. This Subsection focuses on training and optimizing three FFNN models by using data points within each of the three classes. The data points in the hypoglycemic range (class 1) are used to create a network (denoted by FFNN1), the second network is trained using the data points in the normal range (class 2) and this network is denoted by FFNN2, and the third network (denoted by FFNN3) is trained using data points in the hyperglycemic range (class 3). Thus, there are a total of 3 networks, and each network is used to predict the glucose concentration within one the glucose ranges: hypoglycemic, normal, or hyperglycemic.

The FFNN models are trained using two hyperparameters, the learning rate and the hidden unit size. The goal is to select a combination of hyperparameter values that result in neural network models with the lowest possible RMSE value along with the smallest hidden unit size. These values first need to be optimized in order to improve the performance of the neural network's ability to accurately predict the glucose concentration. Values are specified to each of the two hyperparameters for each neural network model as shown in Figure 5.19 for the FFNN1 model. A 5-fold CV method is then performed for each combination of the two hyperparameters to determine the average of accuracy scores for 5 different validation sets. The lowest RMSE will help identify the optimal hyperparameters for each of the FFNN models (FFNN1, FFNN2, and FFNN3). The 4 × 5 optimization grid in Figure 5.19 includes 20 cells, and each cell shows the average accuracy scores (here RMSE is used as an accuracy metric) that are obtained from 5 different validation sets for a combination of the two hyperparameters. RMSE has the same units as the quantity being predicted, so the units for RMSE are mg/dL for glucose concentration. The optimum combination in this case is a learning rate of 0.05 and a hidden unit size of 14, which

results in the lowest RMSE (5.7 mg/dL) without adding complexity to the network with an unnecessary large hidden unit size.

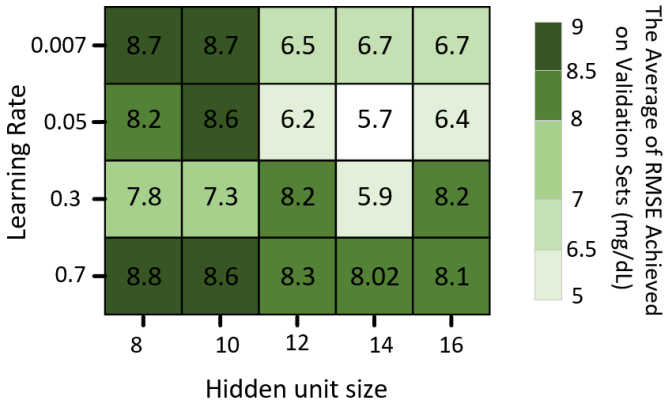


Figure 5.19: Tuning of the hyperparameters for the FFNN1 model using a 5-fold cross validation and grid search method.

The optimization grids for the FFNN2 and FFNN3 models are shown in Figure 5.20 (a) and (b). The optimal values for the learning rate and the hidden unit size are 0.005 and 14, for the FFNN2 model, respectively, and 0.0007 and 18 for the FFNN3 model, respectively.

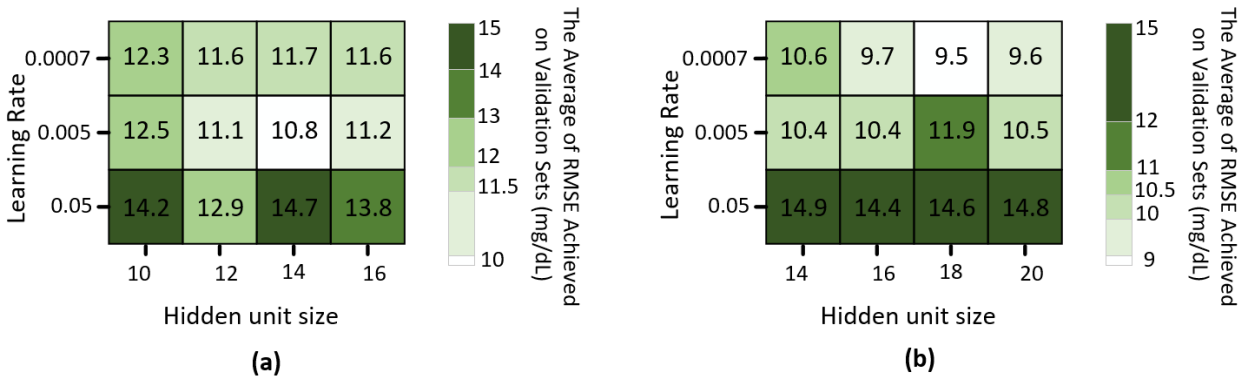


Figure 5.20: Tuning of hyperparameters for (a) FFNN2, and (b) FFNN3 using a 5-fold cross validation and grid search method.

Table 5-2 includes the values of the hyperparameters (learning rate and hidden unit size), along with two other parameters (mini batch size and number of training epoch) for each of the three networks (FFNN1, FFNN2, and FFNN3). The number of epochs can be different for different

networks as specified in Table 5-2. A network is trained during multiple epochs until the network converges to an acceptable RMSE. The training of a network is stopped when there is no significant improvement in the RMSE.

Table 5-2: Values of the parameters in FFNN1, FFNN2, and FFNN3

Range / Info	Hypoglycemic	Normal	Hyperglycemic
Prediction Model	Feed-forward neural network (FFNN1)	Feed-forward neural network (FFNN2)	Feed-forward neural network (FFNN3)
Optimizer	MBGD with momentum	MBGD with momentum	MBGD with momentum
Learning Rate	0.05	0.005	0.0007
Hidden Unit Size	14	14	18
Mini batch size	512	512	64
Number of Training Epoch	~1000	~6000	~8000
RMSE based on 5-fold CV	5.7 mg/dL	10.8 mg/dL	9.5 mg/dL

6. Results

In the previous Chapter, different machine learning techniques are used to create different prediction models that predict glucose concentration based on the intensity data measured at the optimal wavelengths for different glucose concentration within aqueous solutions. For each prediction model, the best hyperparameters are identified to have the maximum prediction accuracy. This Chapter presents the results obtained by evaluating and comparing different prediction models using evaluation criteria such as root mean square error (RMSE), coefficient of determination, Jaccard index, and F1 score. The clinical accuracy of the prediction models are also investigated and compared using Clarke error grid analysis. Based on the results in this Chapter, the model which is best suited for the prediction of glucose concentrations is selected.

6.1 Glucose Prediction Accuracy Results for Solution of Glucose and Water

This Subsection discusses the results obtained by evaluating the performance of five prediction models that are trained and optimized as described and conceptualized in Section 5.1 Data Analysis and Modeling for Predicting Glucose Concentration within . The five prediction models are trained based on the intensity data measured for different glucose concentrations within distilled water. The focus of this Section is to select the model that is best suited for the prediction of glucose concentrations by evaluating the predictive performance of each model using the 5-fold CV method.

6.1.1 Evaluating Glucose Prediction Models Using F1-Score and Confusion Matrix

As Figure 6.1 (a) illustrates the evaluation of the SVM classifier for one out of the five validation sets using the confusion matrix. The confusion matrix represents the actual class labels vs the predicted labels and is used to visualize the performance of the classifier. If an element in row i and column j of a confusion matrix is denoted by n_{ij}, then n_{ij} is the number of data samples

that belong to i^{th} class and predicted to be in j^{th} class. Thus, each element on the main diagonal of the confusion matrix counts the number of data samples that are correctly classified.

The confusion matrix in Figure 6.1 (a) with 21 classes is transformed into a smaller matrix and reclassified into 3 classes as illustrated in Figure 6.1 (b). Each of the three classes contains data samples that fall into one of the glycemic ranges: hypoglycemic, normal or hyperglycemic. As illustrated in Figure 6.1 (a), the data samples within the first three classes are associated with the glucose concentrations between 40 and 70 mg/dL, and this range of glucose serves as an alert for hypoglycemia. By merging the data samples within the first three classes, we create a single class containing data samples that fall in the hypoglycemic range (the yellow cell in Figure 6.1 (b)). Merging the last 7 classes in Figure 6.1 (a) results in a single class that contains data samples associated with high levels of glucose concentrations between 180 and 250 mg/dL which are representative of cases with hyperglycemia (the red cell in Figure 6.1 (b)). A class for the normal glucose range between 70 and 180 mg/dL consists of the reclassification of 11 classes into one (the green cell in Figure 6.1 (b)) [160].

The matrix in Figure 6.1 (b) helps visualize the accuracy of data classification for the hypoglycemic, normal and hyperglycemic ranges. Since there are 5 different validation sets, the average of 5 matrices resulting from the 5 validation sets is computed to create a single matrix labeled as Figure 6.1 (c). The average value for each cell of this matrix is rounded to the nearest integer. The F1-score is used to measure the accuracy of glucose predictions for each of the three classes (*class (i), i= 1, 2 or 3*), and is defined as the weighted average of the precision and recall for each class using the formula in Eq. (6.1).

$$f1\_score(class(i)) = 2 \times \frac{\frac{n_{ii}}{\sum_{j=1}^{c} n_{ij}} \times \frac{n_{ii}}{\sum_{j=1}^{c} n_{ji}}}{\frac{n_{ii}}{\sum_{j=1}^{c} n_{ij}} + \frac{n_{ii}}{\sum_{j=1}^{c} n_{ji}}} \qquad (6.1)$$

where n_{ij} is the number of samples that belong to i^{th} class and are predicted to be in j^{th} class, and c is the total number of classes. Based on Figure 6.1 (c), the value of c is 3, which is representative of the three glucose ranges (or classes): hypoglycemic, normal, and hyperglycemic.

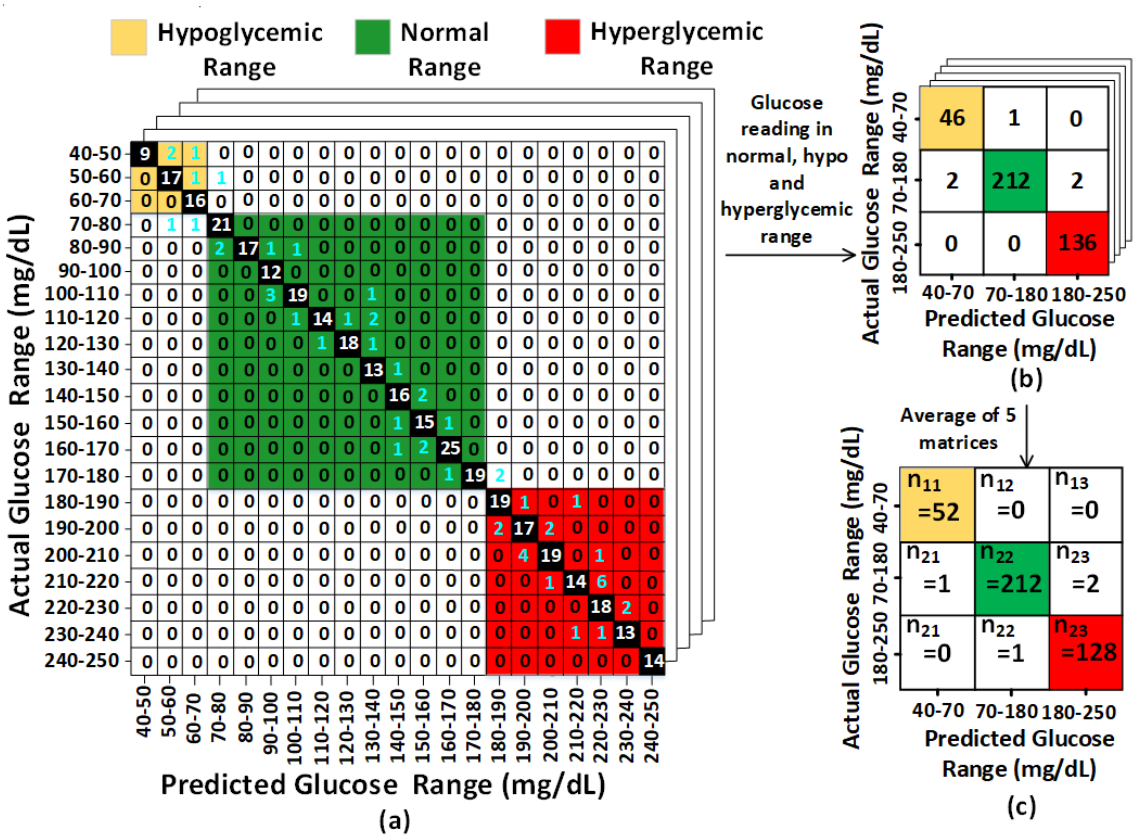


Figure 6.1: (a) The confusion matrix resulting from SVM classifier that is evaluated using one of the validation sets (b) Actual vs predicted glucose within hypoglycemic, normal and hyperglycemic ranges, and (c) average of 5 matrices resulting from the 5 validation sets [160].

The other two classifiers, DT and KNN, are evaluated in a similar fashion, resulting in a single matrix for each classifier, similar to the one in Figure 6.1 (c). The classifier predicts a

discrete range of glucose concentrations, whereas the regression model predicts continuous values of glucose concentrations. Thus, the comparison of the classifier models to the regression models requires that every continuous prediction value from the regression models be converted to a discrete range (hypoglycemic, normal or hyperglycemic). A matrix similar to the one in Figure 6.1 (c) is created for each regression model, as shown in Figure 6.2.

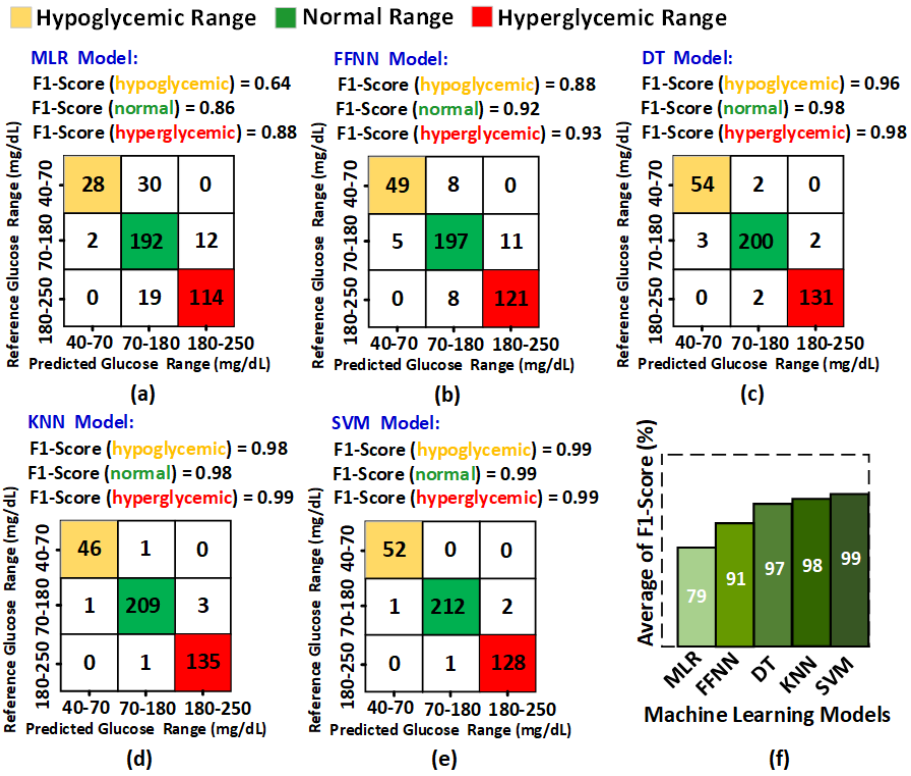


Figure 6.2: Visualization of the accuracy of glucose predictions in the hypoglycemic, normal, and hyperglycemic ranges when using (a) MLR, (b) FFNN, (c) DT, (d) KNN, (e) SVM based models, and (f) a bar graph comparing the different models based on the average of the F1-scores [160].

The confusion matrices resulting for the five different models along with an F1-score for each of the three classes are shown in Figure 6.2 (a) through (e). A bar graph in Figure 6.2 (f) compares the different models based on the average F1-scores. Classifiers demonstrate a higher capacity for accuracy and the ability to identify the hypoglycemia, normal and hyperglycemia

conditions. Among the classifiers, SVM, with an average F1-score of 0.99, yields a better performance compared to the other prediction models. This indicates that the SVM classifier is able to predict the hypoglycemic and hyperglycemic ranges with the highest degree of accuracy [160].

6.1.2 Evaluating Glucose Prediction Models Using Clarke Error Grid Analysis

The clinical accuracy of the glucose predictions is usually evaluated using Clarke error grid analysis (EGA). Clarke EGA evaluates continuous prediction values against reference values and places them into zones for clinical treatment.

The prediction outputs from the regression models (multiple linear regression and feed-forward neural network) are a series of continuous values for glucose concentration, whereas the prediction outputs from the classification models (k-nearest neighbor, decision tree and support vector machine) consist of class labels that indicate a particular range of glucose concentrations. Thus, the class labels for the latter must first be converted to continuous values. A glucose value must be assigned to an actual class label and to a predicted class label for each sample in a validation set. The conversion of class labels to continuous values is carried out in a way that results in a maximum distance between the actual and predicted glucose concentration values based on the maximum and minimum concentrations associated with each class. For example, a data point that consists of an actual class label of 2 and a predicted class label of 1 is converted to 60 mg/dL concentration for the actual label and 40 mg/dL for the predicted label, respectively. This method re-classifies each data point to a concentration value based on a worst-case scenario [160].

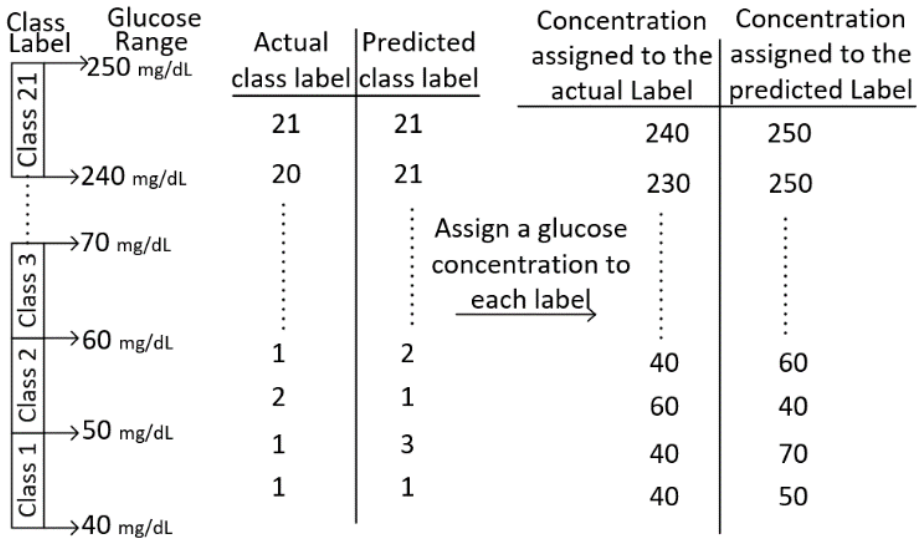


Figure 6.3: Conversion of class labels to continuous glucose concentration values [160].

The Clarke error grid includes 5 zones, identified as zones A-E, depending on glucose concentration prediction accuracy. A glucose prediction point will fall in zone A if a predicted value of glucose concentration is within 20% of the actual glucose level. Thus, the recommended treatment will be appropriate for the patient. Glucose predictions falling in zone B differs from the glucose reference level by more than 20%, but the error would result in a clinically acceptable treatment for the patient. Glucose readings that fall in zone C are associated with a level of error that leads to unnecessary treatment. Glucose readings that fall in zone D include a level of error that may result in a failure in the detection and treatment of hypoglycemia and/or hyperglycemia. The worst case is when a prediction value falls in zone E since this level of error leads to a treatment that is opposite to what is actually required for the patient [4] [175]. The results of the Clarke EGA is included in Figure 6.4 and quantifies the accuracy of each of the five prediction models.

The percentage of predictions falling in zone D (the zone of failure to detect) is lowest (0.25%) for the SVM and KNN classifier models (see Figure 6.4 (e)). Among these two classifiers,

SVM results in the highest agreement between the reference and predicted glucose level since 97.5% of glucose predictions fall within zone A (clinically accurate) and 2.25% fall within zone B (clinically acceptable) [160].

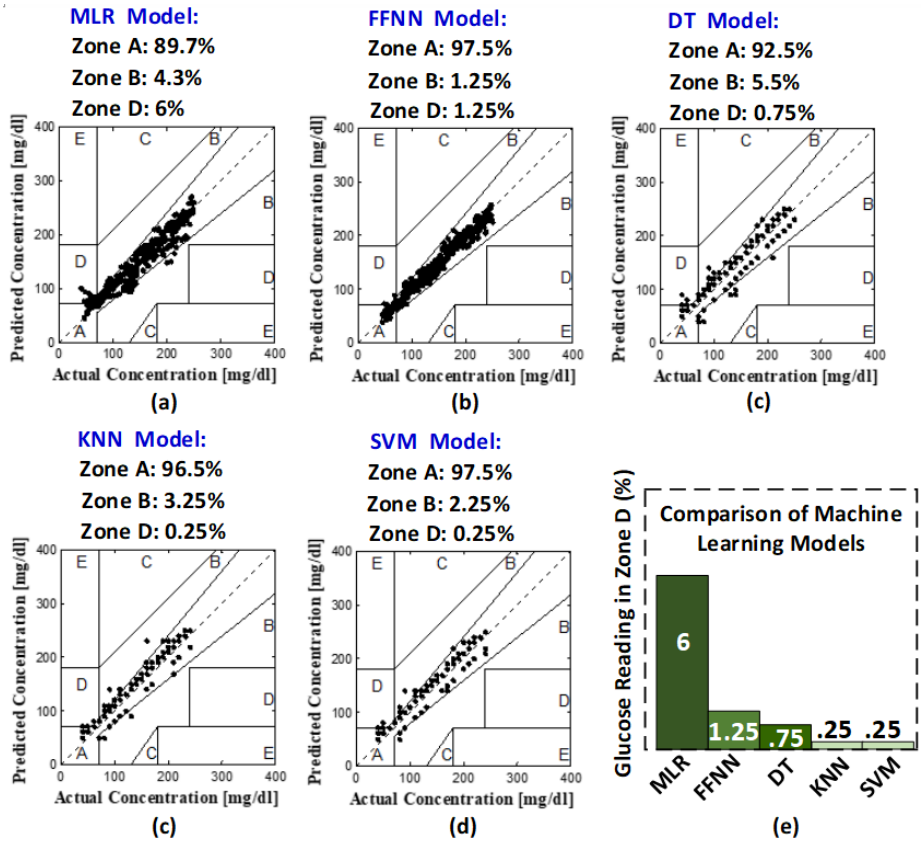


Figure 6.4: Clarke EGA of the glucose predictions when using (a) MLR, (b) FFNN, (c) DT, (d) KNN, (e) SVM based model, and (e) the Zone D bar graph representation of all 5 models [160].

6.1.3 Comparison between Prediction Models and Discussion

The developed VIS-NIR optical device is used to measure the intensity of light transmitted through glucose within distilled water solutions in the range between 410 and 940 nm. The use of multi-wavelength measurements and analysis extracts more quantitative information about the glucose concentration in solution, and improves the accuracy of non-invasive glucose predictions

values. Four distinct wavelengths, 485, 645, 860 and 940 nm, result in a higher correlation between predicted and actual glucose concentration values.

Glucose prediction models using regression and classification models yield promising results. The regression models, MLR and FFNN, predict continuous glucose concentration values based on measured data, whereas classification models, KNN, SVM and DT, predict a class label. A class label encompasses a discrete 10 mg/dL glucose range. Identifying the best values of hyperparameters helps to optimize each model until it yields the most accurate prediction score. The performance of each model is evaluated using the 5-fold CV method, and the accuracy results obtained from 5 validation sets are compared in Table 6-1 [160].

The FFNN regression model with an R-squared value of 0.96 and a RMSE of 11.1 mg/dL outperforms the MLR model. This is not surprising since the neural network model is able to learn complex non-linear relationships between its input and output values. Regression algorithms aim to minimize the error between actual and predicted glucose concentration values, but are not directly involved in minimizing the error in the prediction of hypo or hyperglycemia. Thus, it is possible for a regression model to make a prediction that erroneously identifies a value within the hypo or hyperglycemic ranges. On the other hand, classification algorithms learn how to classify data samples into the normal, hypo and hyperglycemic ranges. The prediction output of a classification model (or classifier) is a discrete glucose range, in comparison to the regression model that predicts continuous glucose concentration values. When the classifier predicts a correct glucose concentration range, the actual glucose concentration value lies somewhere within the predicted range. The difference between the highest and lowest value is the width of the range. The width of the range will depend on the number of classes selected for the entire range of data samples. In this study, data samples are grouped into 21 classes. Each class includes samples that

correspond to a discrete 10 mg/dL glucose range, and each class falls into one of the glycemic ranges (normal, hypo or hyperglycemic). For example, the first class (class1) includes samples associated with a glucose concentration range between 40 and 50 mg/dL, and this range is associated with the hypoglycemic range. Likewise, the last class (class 21) includes samples that are associated with a glucose concentration range between 240 and 250 mg/dL, and this range lies within the hyperglycemic range. The partitioning of data samples into 21 classes enables the use of classification algorithms that have the ability to classify a new data sample into the correct glucose range. Thus, there is a high probability that a glucose prediction will fall within the glycemic range that contains the actual glucose value [160]. Table 6-1 includes a summary of the analysis of all five models.

Table 6-1: The results obtained from 5-Fold cross validation for different glucose prediction Models [160]

Evaluation Metrics	Regression Models		Classification Model		
	MLR	FFNN	DT	KNN	SVM
Jaccard Index, when there are 21 classes	N/A	N/A	0.75	0.84	0.86
R-Squared	0.91	0.96	N/A	N/A	N/A
RMSE (mg/dL)	16.2	11.1	15	13.5	12.3
F1-Score (hypoglycemic)	0.64	0.88	0.96	0.98	0.99
F1-Score (normal)	0.86	0.92	0.98	0.98	0.99
F1-Score (hyperglycemic)	0.88	0.93	0.98	0.99	0.99
Averaged F1-Score	0.79	0.91	0.97	0.98	0.99
Data percentage in the zone D of the Clarke error grid	6%	1.25%	75%	0.25%	0.25%

The classification models are associated with the highest F1-scores compared to regression models, and outperform the MLR model, which is the model commonly used for glucose predictions. A higher F1-score in the case of classifiers indicates an ability to correctly distinguish between the three glycemic ranges: hypoglycemic, normal and hyperglycemic. Classifiers result in lower percentage of glucose reading in zone D of the Clarke error grid, and this indicates a lower expectation of incorrectly detecting and treating hypoglycemia or hyperglycemia. It is very crucial to predict glucose values in the correct glycemic range, so diabetes patients can take appropriate actions to prevent life threating consequences.

The best glucose model is the one with (1) the lowest RMSE and (2) the lowest error in the prediction of hypo and hyperglycemia. The FFNN model satisfies the first requirement since it has the lowest RMSE value of 11.1 mg/dL, but this model does not have the highest F1-score. The SVM model has the highest F1 score with a value of 0.99, it has the lowest percentage of glucose readings within zone D of the Clarke error grid (0.25%), and an RMSE value of 12.3 mg/dL, which is comparable to the RMSE of the FFNN model (11.1 mg/dL). It makes sense to sacrifice the RMSE slightly in order to achieve this level of accuracy in the diagnosis of hypoglycemia and hyperglycemia. The KNN classifier has the second best F1-score and it coincides with the SVM model in terms of the percentage of readings that fall in zone D of the Clarke error grid (0.25%). The DT classifier rates higher than the FFNN model in terms of its ability to distinguish between different glycemic ranges with an F1-score of 0.97 vs 0.91, although the RMSE of the DT model (15 mg/dL) is much greater than the value for the FFNN model (11.1 mg/dL). Among all five models, the MLR model has the lowest F1-score, the highest percentage of data falling in zone D of the Clarke error grid (6%), and the highest RMSE (16.2%). The MLR model makes predictions

based on a linear relationship between the input parameters and the glucose concentration, whereas the other models rely on non-linear relationships [160].

6.2 Glucose Prediction Accuracy Results for Phosphate Buffered Saline Supplemented with Glucose and Albumin

The optimal hyperparameters identified in Subsection 5.2.2 Glucose Prediction Using Machine Learning Techniques, are used to create SVM and FFNN models. The SVM model is used to accurately predict a class (hypoglycemic, normal or hyperglycemic range) for a given test data point, and then according to the predicted class, a FFNN regression model predicts the glucose concentration for the test data point. Our dataset includes data points within three ranges: hypoglycemic, normal and hyperglycemic. FFNN1 is trained using data points that belong to the hypoglycemic range, and FFNN2 and FFNN3 are trained using data points that are associated with the normal and hyperglycemic ranges, respectively. In this Section, the prediction accuracy of the models created in Subsections 5.2.2 Glucose Prediction Using Machine Learning Techniques are evaluated using the test set (15% of total data).

6.2.1 Evaluating Glucose Prediction Models Using F1-Score and Confusion Matrix

The data used to test the SVM classifier model is 15% of the total data, and is denoted by Dataset 2. Dataset 2 serves as the test set and is input into the SVM classifier. The SVM classifier arranges the test data points into three classes (or ranges) based on the glucose concentration associated with each data point. The accuracy of classification by the SVM model can be visualized using the confusion matrix in Figure 6.5. The values along the main diagonal of the confusion matrix indicate the number of data points that are correctly classified, and the remaining cells include the rest of the data points that are misclassified into the wrong glucose range. Since there are only 6 misclassified data points among a total of 785 data points, the Jaccard index is equal to

0.99. The F1-score is used as an evaluation metric to determine the prediction accuracy for each individual class. The equation for the F1 score (Eq. (6.1), introduced in Subsection 6.1.1 Evaluating Glucose Prediction Models Using F1-Score and Confusion Matrix) is used to measure the accuracy of the glucose predictions for each range. The F1 score is calculated to be 0.99 for each of the three classes based on the SVM model. A high F1 score indicates the generalizability of the SVM classifier to perform well on the test set (previously unseen data) with a high prediction accuracy when categorizing the test data points into the three ranges.

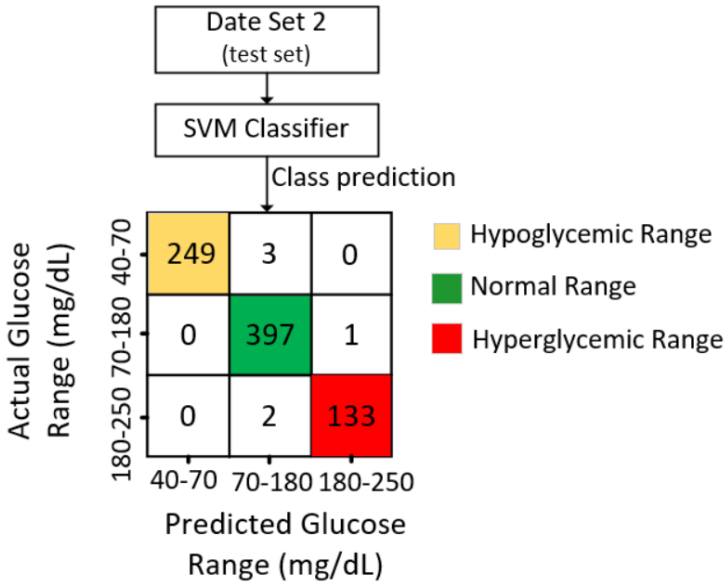


Figure 6.5: Confusion matrix used to characterize the SVM classifier ability to accurately classify the test set data points.

6.2.2 Evaluating Glucose Prediction Models Using Clarke Error Grid Analysis

The SVM classifier is able to accurately predict a class for a given test data point as demonstrated in Section 6.2.1 Evaluating Glucose Prediction Models Using F1-Score and Confusion Matrix. Based on the predicted class, one of the FFNN models is used to predict the glucose concentration for that data point. Figure 5.15 illustrates the classification for each range (hypoglycemic, normal and hyperglycemic). If the predicted class is hypoglycemic (class1), then

the test data point is input into the FFNN1 model to predict the glucose concentration within hypoglycemic range. If the prediction made by the SVM classifier is normal (class 2), then the FFNN2 model is used to predict the glucose concentration, and in the case that the SVM classifier predicts class 3 (hyperglycemic) for a test data point, then the FFNN3 is used to predict the glucose concentration for that data point.

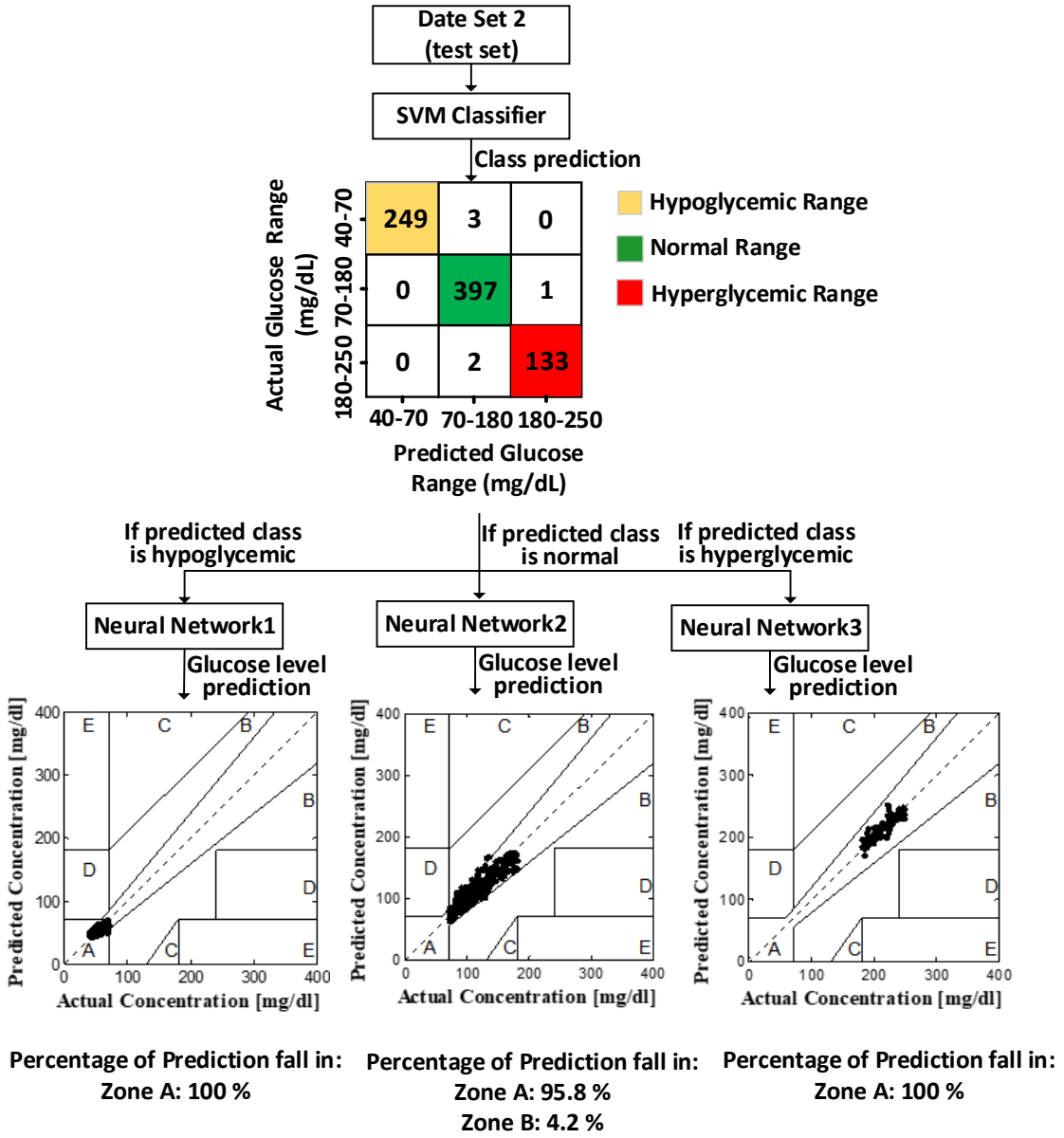


Figure 6.6: Prediction accuracy of test set data points for the SVM and FFNN models based on the Clarke error grid analysis.

After the FFNN1, FFNN2 and FFNN3 models are finished predicting the glucose concentration within the hypoglycemic, normal, and hyperglycemic ranges, respectively, the Clarke error grid analysis (EGA) is used to evaluate the clinical prediction accuracy for each of the three FFNN regression models. Figure 6.6 illustrates how the FFNN models are evaluated using the Clarke error grid analysis. The Clarke error grid includes 5 zones (A to E). These zones are described as zone A (clinically accurate glucose reading), zone B (clinically acceptable glucose reading), zone C (leads to unnecessary treatment), zone D (failure to detect and treat hypoglycemia and/or hyperglycemia), and zone E (leads to treatment that is opposite to what is actually required for the patient). For the FFNN1 model, 100% of the predictions fall within zone A of the Clarke error grid. For the FFNN2 model, 95.8% and 4.2% of the predictions fall within zone A and B of the Clarke error grid, respectively. And for the FFNN3 model, 100% of the predictions fall within zone A. Overall, 100% of the predictions fall within either zone A or B of the Clarke error grid. This indicates that although 6 data points are misclassified by the SVM classifier, the error in glucose prediction is at a clinically acceptable level.

6.2.3 Discussion

The VIS-NIR optical device designed and developed here is used to measure the intensity of light transmitted through a glucose and albumin PBS solution using light sources in the range between 410 and 940 nm. It is known that changes in the albumin level result in changes in the intensity of light entering the detector, and this can negatively impact the accuracy of glucose predictions. The use of a multi-wavelength source and/or detector helps to compensate the error in the glucose readings caused by changes in the albumin level, and this has the potential to improve the accuracy of non-invasive glucose sensors. We demonstrate that the use of 485, 645, 860 and

940 nm wavelengths result in a higher correlation between the predicted and actual glucose concentration values.

The measured light intensity data associated with these four wavelengths are classified as hypoglycemic (class 1), normal (class 2), and hyperglycemic (class 3) ranges by the SVM classifier model with a 99% accuracy based on the F1-score. Depending on the predicted range, one of the 3 feed-forward neural networks (FFNN1, FFNN2, or FFNN3) is used to predict the glucose concentration. A summary of the accuracy of the FFNN models is included in Table 6-2: The prediction accuracy results obtained by evaluation of models using test set.

Table 6-2: The prediction accuracy results obtained by evaluation of models using test set when combination of both SVM classifier and FFNN models are used for glucose predictions

Range / Metrics	Hypoglycemic, Class 1	Euglycemic, Class 2	Hyperglycemic, Class 3
Number of data points classified into each class by the SVM classifier	n_1=249	n_2=402	n_3=134
Number of data points that are correctly classified by the SVM classifier	249	397	133
SVM classifier F1-score	0.99	0.99	0.99
SVM classifier jaccard index	0.99		
RMSE (mg/dL)	5.54	11.2	8.6
Overall RMSE (mg/dL)	Sqrt (($(RMSE_1)^2 \times n_1 + (RMSE_2)^2 \times n_2 + (RMSE_3)^2 \times n_3$)) / ($n_1 + n_2 + n_3$))= Sqrt ((30.7×249+125.4×402+74×134)/785)) = 9.3		
Percentage of predictions fall in zone A of Clarke grid	100%	95.8%	100%
Overall percentage of predictions fall in zones A+B of Clarke grid	($n_1^{(A+B)} + n_2^{(A+B)} + n_3^{(A+B)}$) / ($n_1 + n_2 + n_3$) = (249 + 402 + 134) / (249 + 402 + 134) = 100%		

The RMSE scores range between 5.54 mg/dL and 11.2 mg/dL. The highest accuracy and lowest RMSE score (5.54 mg/dL) is associated with the FFNN1 model within the hypoglycemic range. In this case, 100% of the predictions fall within zone A, which indicates that all the predictions are clinically accurate. The highest RMSE value is associated with predictions in the normal range by the FFNN2 prediction model. In this case, 95.8% of predictions fall within zone A, and 4.2% fall within zone B of the Clarke error grid. Although 4.2% predictions fall within zone B, these values are clinically acceptable. The FFNN3 model resulted in prediction with a RMSE of 8.6 mg/dL in hyperglycemic range, and all the predictions made by the FFNN3 model fall within zone A of the Clarke error grid.

Using the SVM classifier to predict class distribution of the test set data points, followed by a regression model that predicts a glucose concentration value within each class results in an overall RMSE equal to 9.3 mg/dL (see calculations in Table 6-2: The prediction accuracy results obtained by evaluation of models using test set), and 100% clinical accuracy based on Clarke error grid analysis. The use of this modeling methodology to predict glucose values outperforms the methodology of using a single regression model over the entire glucose range. To demonstrate that the proposed modeling methodology outperforms the use of a single regression model that is commonly used in the literature to predict the glucose concentration, a single FFNN model was optimized and trained using all the data points within the training set (denoted by dataset 1 in Subsection 5.2.2 Glucose Prediction Using Machine Learning Techniques), and evaluated using the test set (denoted by dataset 2 in Subsection 5.2.2 Glucose Prediction Using Machine Learning Techniques). Note that in the case of a single FFNN, there is no need to distinguish between different ranges (or classes), and thus the SVM classifier model was not used. The results of this study are included in Table 6-3. In this case, the single FFNN model resulted in a RMSE of 12.7

mg/dL and 96.6% of the predictions were within zone A, 2.5% within zone B, and 0.9% within zone D of the Clarke error grid. The glucose reading in zone D indicates a potential failure to detect and treat hypoglycemia and/or hyperglycemia, making the patient unaware of the hypo or hyperglycemia condition.

This study demonstrates that when a SVM classification model is used followed by 3 FFNN models versus the case where a single FFNN model is used (comparison of results from Tables 6-2 and 6-3), the RMSE is lower. The multiple FFNN machine learning technique results in a RMSE of 9.3 mg/dL versus 12.7 mg/dL for the single FFNN model. In addition, 100% of the predictions fell within zone A and B of the Clarke grid when both SVM classification and FFNN regression models are used.

Table 6-3: The prediction accuracy results obtained by evaluation of models using test set when a single FFNN model is used for glucose predictions

Range / Metrics	Entire Range
Number of data points	n=785
RMSE (mg/dL)	12.7
Percentage of predictions fall in zones of Clarke grid	Zone A= 96.6 Zone B = 2.5 Zone D = 0.9

6.3 ACCURACY COMPARISON BETWEEN THIS WORK AND THE LITERATURE

Table 6-4 summarizes the results from this research in comparison to the performance of non-invasive glucose sensors identified from published literature sources [11], [12], [176]–[180]. The accuracy of glucose readings is evaluated using three metrics, mean average relative difference (MARD), root mean squared error (RMSE) in mg/dL, and Clarke/Consensus grid

analysis. The lower values of MARD and RMSE is preferable and means more accurate glucose monitoring. This work results in MARD = 7% and RMSE = 9.3 mg/dL, which are lower or comparable with the findings in the published studies. Also, the proposed approach in this work ensures that 100% of glucose readings fall within the clinically acceptable zones (A and B) of the Clarke error grid.

Table 6-4: Comparison between results obtained in this study and literature

Ref	Exp.	Measurement Method	Prediction Method	Component Range (mg/dL)	MARD (%) or Error (mg/dL)	Error Grid Analysis
[176]	In-vitro solution	NIR-MIR (2100-8000nm) spectroscopy	Principal component and linear discriminant analysis	Glucose (5-500) Albumin (5 g/dL)	21.8-43.1 mg/dL	A+B =100%
[12]	In-vitro solution	Polarimetry range (380-680 nm)	Partial least square regression	Glucose (0-500) Albumin (0-1 g/dL)	16 mg/dL	-
[177]	In-vitro solution	NIR (905, 1550 nm) and Photoacoustic Spectroscopy	Kernel based regression	Glucose (0-500)	8.3%	-
[178]	In-vitro solution	NIR single wavelength (1550 nm)	Linear regression	Glucose (50-400)	-	A+B = 85.6, C= 3.75, D=7.5
[179]	In-vitro solution	MIR Spectroscopy	Partial least square regression	Glucose (50-400) within Serum	7.1-35.53 mg/dL	A+B= 100%
[11]	In-vivo human	Raman Spectroscopy	Partial least square regression	Glucose (40-500)	27.3%	A= 54.1% B= 39.8% C= 5.4, D= 0.7%
[180]	In-vivo human	Dexcom G4TM CGM	-	Glucose (40-350)	11%	A= 76.3% B= 22.3%, D= 1.3%
This Work	**In-vitro solution**	**VIS-NIR Spectroscopy (4 wavelengths)**	**SVM Classification and FFNNs Regression**	**Glucose (40-250) Albumin (4-5 g/dL)**	**7%, 9.3 mg/dL**	**A + B =100%**

The typical accuracy of the commercialized invasive continuous-time glucose monitoring devices are 10-30 % for MARD, and 90-98% for percentage of glucose readings within zones A and B. Although it is not fair to compare the accuracy results for human testing by commercialized invasive sensors to the in-vitro experiments performed in this study, the accuracy results for this study are within the acceptable accuracy range that exist for the commercial invasive sensors.

7. Conclusions and Future Work

This Chapter concludes this research by summarizing the key arguments, reviewing main points, and highlighting accomplishments. Also, it includes a "Future Work" Section to explain what questions arose from this work, where researchers need to be looking next, and a recommended direction for future experiments.

7.1 Conclusions

This research includes the design and development of a multiple wavelength VIS-NIR optical based sensor to non-invasively and accurately measure in-vitro glucose concentrations. The goal of the research is to design a sensor that considers the different species affecting the non-invasive penetration of light sources into the body and to study the effect of machine learning or data analysis on the accuracy of glucose measurements on in-vitro samples before such techniques are applied to non-invasive glucose sensors. The design of the sensor encompasses the identification of the optimal wavelengths for the sensor light sources, signal processing of data collected using the sensor, and machine learning prediction methodologies. Based on this study, the following conclusions are summarized below:

1. Based on the preliminary study using the commercial spectrometer (Chapter 3) and the current literature, the NIR and visible ranges are the preferable light sources to measure glucose in solution compared to MIR. This is due to a lower absorption of light by water molecules within these ranges.
2. A commercialized spectrometer is used to identify the wavelengths with the highest absorption of light by glucose molecules. These turned out to be 950 nm, 1,025 nm, 1,200 nm and 1,600 nm.

3. A single wavelength glucose sensor is designed and tested using commercial available light sources close to the four values identified using the commercial spectrometer. These light sources included wavelengths of 940 nm, 1,050 nm, 1,200 nm and 1,600 nm. The most accurate glucose measurements are associated with an incident light source with a wavelength of 940 nm.
4. The accuracy of the sensor is improved further by designing a sensor with multiple light sources and detectors. The intensity of the signal transmitted through the sample is dependent on the absorption and scattering behavior of light within the sample, which depends on the glucose concentration in solution, in addition to other influential components, thus the identification of the most effective wavelengths for detection is critical for a non-invasive glucose sensor. The multiple wavelength sensor includes 3 LEDs and a Triad sensor with 3 detector devices that are able to measure the intensity of light at 18 wavelengths within the VIS-NIR range between 410 and 940 nm. This is referred to in the document as the VIS-NIR optical sensor. Of the 18 wavelengths examined, 485 nm, 645 nm, 860 nm, and 940 nm have the highest sensitivity and selectivity in relation to glucose concentration measurements.
5. Multi-wavelength analysis helps to minimize the error associated with multiple high absorbing species associated with non-invasive glucose measurements, such as water, proteins (ex. albumin). In this study, the intensity data measured at multiple wavelengths (485 nm, 645 nm, 860 nm, and 940 nm) helps to minimize errors resulting from species that affect the absorption and scattering of the intensity of light and thus affect the detection signal. The use of multiple wavelengths here is tested using an aqueous glucose with albumin PBS solution, and is shown to enhance the sensitivity and selectivity to glucose

concentration. The light attenuation coefficient is unique for each species and is dependent on the wavelength of the incident light interacting with each molecule. The intensity data measured at these optimal wavelengths helps to compensate for the presence of water and albumin, and shows a higher correlation with the known glucose concentration levels based on the Pearson correlation coefficient.

6. Classification based models (k-neareset neighbor (KNN), decision tree (DT) and support vector machine (SVM)) demonstrate a greater ability to correctly distinguish between the hypoglycemic, normal and hyperglycemic ranges compared to the regression models (multiple linear regression (MLR) and feed-forward neural network (FFNN)). Here, five machine learning techniques and prediction models are tested and compared using an aqueous glucose solution with glucose concentrations between 40 - 250 mg/dL to identify the model with the lowest RMSE and the lowest error in prediction of hypoglycemia and hyperglycemia. Among the five models, the SVM classifier results in the highest F1 score with a value of 0.99, and results in the lowest percentage of glucose readings within zone D of the Clarke error grid (0.25%), and a corresponding RMSE value of 12.3 mg/dL, which is comparable to the lowest RMSE (11.1 mg/dL) resulting from the FFNN model.
7. The combination of both the classification and regression models to predict glucose concentrations for a given dataset results in more accurate and reliable predictions compared to using a single regression model for the entire glucose range. The use of hybrid models improves the RMSE from 12.7 mg/dL (in the case of a single regression model) to 9.3 mg/dL (in the case of hybrid models as was done in this study), and ensures 100% of glucose readings fall within clinically acceptable zones (A and B) of the Clarke error grid. This was concluded from glucose measurements within a phosphate buffered saline

solution with different levels of bovine serum albumin. A variation in the albumin concentration can affect the measured intensity, and thus affect the accuracy of glucose readings. In this case, there is need for more powerful prediction approaches to accurately estimate the glucose concentrations within the albumin containing PBS solution. A regression method in conjunction with a classification method are used to improve the accuracy of the glucose prediction, as compared to using a regression method alone or a classification method alone. First, the intensity data collected using the four optimal wavelengths is input into a SVM classifier to discriminate between the hypoglycemic, normal and hyperglycemic ranges with 99% accuracy based on F1-scores. Then, for each of the three glycemic ranges, individualized feed-forward neural network models are developed and applied to a test dataset to predict the glucose concentration within each range separately. Although the presence of albumin affects the accuracy of the glucose readings, the use of multiple wavelength analysis and SVM classification, along with 3 FFNN models for each glycemic range, helps to minimize the error so that the accuracy is similar when compared to the results from aqueous glucose solutions without the presence of albumin.

The activities and concepts used in this research are not limited to glucose sensing, and can also be applied to improve the accuracy of other non-invasive sensing techniques. In-vitro experiments performed in this study can be considered as a first and important step to demonstrate the feasibility of non-invasive measurements of glucose in the presence of PBS solutions with albumin. The next step is to test the sensor and methodologies on a body measurement site (e.g. finger). In-vivo experiments will require a change in the design of the VIS-NIR sensor so that it is miniaturized enough to be compatible as a wearable type sensor. Especially important in this work

is to examine the relationship between in-vitro and in-vivo glucose detection by performing clinical studies and comparing the data from in-vivo to corresponding in-vitro solutions.

7.2 FUTURE WORK

The in-vitro experiments (using glucose and albumin within PBS solution) performed in this study is an important step to examine the feasibility and accuracy of non-invasive glucose sensor measurements, to identify the theoretical and methodological challenges, and to establish proof of concept for the non-invasive approach proposed in this study.

In this study, it is demonstrated that collecting intensity data at 4 wavelengths, 485, 645, 860 and 940 nm, (compared to single wavelength measurement) and glucose predictions based on hybrid techniques, SVM classification and FFNN regression, (compared to using single regression over the entire glycemic range) help to minimize the error in glucose readings that is induced by water and albumin content within samples. In addition to water and albumin, there are other physiological factors that may affect the accuracy of glucose readings. As a future phase of this study, there is need to validate effectiveness of the proposed experimental and theoretical methods by performing new experiments using blood samples and even more complex experiments on living organisms. The following are recommendations to continue to expand on the current research introduced in this work.

1. The VIS-NIR based optical glucose sensor using multiple wavelength light sources included a UV, visible and NIR light source. The amplitude of the measured signal coming from the visible light source is much higher compared to the UV and NIR light sources. This can be balanced out by using light sources with a higher emitting intensity. Another way to increase the measured light intensity is to optimize the placement of the light source with respect to the photodiode sensor device.

2. The preliminary study identified four potentially optimal wavelengths (950 nm, 1,025 nm, 1,200 nm and 1,600 nm). The glucose absorption peaks are observed at these four wavelengths. The peak at 950 nm shows a higher correlation between intensity of light and changes in glucose concentration. The LED light sources commercially available with the closest corresponding wavelengths include 940 nm, 1,050 nm, 1,200 nm and 1,600 nm. Therefore, these values are used in place of the potentially optimal wavelengths. Therefore, the 940 nm light source was used for both the single and multiple wavelength sensors. However, other wavelengths can be tested for possible light source candidates in future work.
3. Perform in-vitro measurements using human blood plasma samples: This study would require a sufficient number of blood samples to provide proper validation. In addition to water (~92%), albumin (~60% of total plasma protein), and salt (~0.5%) that we examined in this study, plasma includes other components such as immunoglobulin G (~20% of total plasma protein); and transferrin, fibrinogen, and lipids (0.4-0.7%). Experiments with human plasma samples provides an opportunity to determine the effect of the physiological factors in blood plasma samples on the accuracy of glucose concentration readings, and to examine if multiple wavelengths measurements and hybrid based prediction models can help to compensate for the error associated with the plasma readings.
4. In-vitro evaluation of mixtures of red blood cells and plasma: Hemoglobin is a protein that is found within red blood cells. The absorption peaks for hemoglobin within the VIS-NIR range are placed at 540 nm, 576 nm, and between 380 and 420 nm [181]. The four optimal wavelengths for glucose measurements identified in this research are 485, 645, 860, and 940 nm. There is no overlap between the absorption peaks for hemoglobin and the four

wavelength values used in this study. Therefore, it can be concluded that analyzing intensity data using light sources with these four wavelengths is a reliable approach for glucose prediction within red blood cells and plasma.

5. In-vitro experiments using a tissue-mimicking phantom that emulates properties of biological tissue. There are different commercial tissue phantoms that are suitable for diverse applications. Optical tissue phantoms need to be used to mimics light propagation and interactions through human tissue. For example, agar, gelatin, or collagen based optical phantom that allow easy inclusion of cellular constituents such as blood or fat can be used for spectroscopy measurements [182]. This will provide more realistic and pre-clinical evaluation of proposed approaches.
6. Clinical trials by conducting experiments on human volunteers. For this kind of measurement, it is necessary to apply a sensor on a human body site. Selecting a measurement site with a high density of blood vessels minimizes the time lag between actual blood glucose concentrations vs. the values read by the sensor. One of the suitable human body locations for sensor placement is the fingertip as a result of the high density of blood vessels. If a sensor is intended for the finger, the physical shape of the sensor needs to be changed and miniaturized so that it is a wearable type sensor. This type of validation study would quantify the correlation and agreement between in-vitro experiments such as those performed in this study and the clinical reference methods used to measure human blood glucose.

References

[1] N. H. Cho *et al.*, "IDF Diabetes Atlas: Global estimates of diabetes prevalence for 2017 and projections for 2045," *Diabetes Res. Clin. Pract.*, vol. 138, pp. 271–281, Apr. 2018, doi: 10.1016/j.diabres.2018.02.023.

[2] W. R. Rowley, C. Bezold, Y. Arikan, E. Byrne, and S. Krohe, "Diabetes 2030: Insights from Yesterday, Today, and Future Trends," *Popul. Health Manag.*, vol. 20, no. 1, pp. 6–12, Feb. 2017, doi: 10.1089/pop.2015.0181.

[3] American Diabetes Association, "2. Classification and Diagnosis of Diabetes: *Standards of Medical Care in Diabetes—2018*," *Diabetes Care*, vol. 41, no. Supplement 1, pp. S13–S27, Jan. 2018, doi: 10.2337/dc18-S002.

[4] M. Shokrekhodaei and S. Quinones, "Review of Non-Invasive Glucose Sensing Techniques: Optical, Electrical and Breath Acetone," *Sensors*, vol. 20, no. 5, p. 1251, Feb. 2020, doi: 10.3390/s20051251.

[5] T. Lin, "Non-Invasive Glucose Monitoring: A Review of Challenges and Recent Advances," *Curr. Trends Biomed. Eng. Biosci.*, vol. 6, no. 5, Jul. 2017, doi: 10.19080/CTBEB.2017.06.555696.

[6] S. A. Siddiqui, Y. Zhang, J. Lloret, H. Song, and Z. Obradovic, "Pain-Free Blood Glucose Monitoring Using Wearable Sensors: Recent Advancements and Future Prospects," *IEEE Rev. Biomed. Eng.*, vol. 11, pp. 21–35, 2018, doi: 10.1109/RBME.2018.2822301.

[7] S. Liakat, K. A. Bors, L. Xu, C. M. Woods, J. Doyle, and C. F. Gmachl, "Noninvasive in vivo glucose sensing on human subjects using mid-infrared light," *Biomed. Opt. Express*, vol. 5, no. 7, p. 2397, Jul. 2014, doi: 10.1364/BOE.5.002397.

[8] J. Y. Chen *et al.*, "Non-invasive blood glucose measurement of 95% certainty by pressure regulated Mid-IR," *Talanta*, vol. 197, pp. 211–217, May 2019, doi: 10.1016/j.talanta.2019.01.034.

[9] P. Jain, R. Maddila, and A. M. Joshi, "A precise non-invasive blood glucose measurement system using NIR spectroscopy and Huber's regression model," *Opt. Quantum Electron.*, vol. 51, no. 2, p. 51, Feb. 2019, doi: 10.1007/s11082-019-1766-3.

[10] E. Guevara, J. C. Torres-Galván, M. G. Ramírez-Elías, C. Luevano-Contreras, and F. J. González, "Use of Raman spectroscopy to screen diabetes mellitus with machine learning tools," *Biomed. Opt. Express*, vol. 9, no. 10, p. 4998, Oct. 2018, doi: 10.1364/BOE.9.004998.

[11] S. M. Lundsgaard-Nielsen, A. Pors, S. O. Banke, J. E. Henriksen, D. K. Hepp, and A. Weber, "Critical-depth Raman spectroscopy enables home-use non-invasive glucose monitoring," *PLOS ONE*, vol. 13, no. 5, p. e0197134, May 2018, doi: 10.1371/journal.pone.0197134.

[12] C. Stark, C. A. Carvajal Arrieta, R. Behroozian, B. Redmer, F. Fiedler, and S. Müller, "Broadband polarimetric glucose determination in protein containing media using characteristic optical rotatory dispersion," *Biomed. Opt. Express*, vol. 10, no. 12, p. 6340, Dec. 2019, doi: 10.1364/BOE.10.006340.

[13] Y. Tanaka, T. Tajima, M. Seyama, and K. Waki, "Differential Continuous Wave Photoacoustic Spectroscopy for Non-Invasive Glucose Monitoring," *IEEE Sens. J.*, vol. 20, no. 8, pp. 4453–4458, Apr. 2020, doi: 10.1109/JSEN.2019.2962251.

[14] S. Saha *et al.*, "A Glucose Sensing System Based on Transmission Measurements at Millimetre Waves using Micro strip Patch Antennas," *Sci. Rep.*, vol. 7, no. 1, p. 6855, Dec. 2017, doi: 10.1038/s41598-017-06926-1.

[15] M. Baghelani, Z. Abbasi, M. Daneshmand, and P. E. Light, "Non-invasive continuous-time glucose monitoring system using a chipless printable sensor based on split ring microwave resonators," *Sci. Rep.*, vol. 10, no. 1, p. 12980, Dec. 2020, doi: 10.1038/s41598-020-69547-1.

[16] X. Xiao and Q. Li, "A Noninvasive Measurement of Blood Glucose Concentration by UWB Microwave Spectrum," *IEEE Antennas Wirel. Propag. Lett.*, vol. 16, pp. 1040–1043, 2017, doi: 10.1109/LAWP.2016.2618946.

[17] T.-L. Chen, Y.-L. Lo, C.-C. Liao, and Q.-H. Phan, "Noninvasive measurement of glucose concentration on human fingertip by optical coherence tomography," *J. Biomed. Opt.*, vol. 23, no. 04, p. 1, Apr. 2018, doi: 10.1117/1.JBO.23.4.047001.

[18] T. Tajima, Y. Tanaka, M. Nakamura, and M. Seyama, "Multi-modality analysis of glucose aqueous solution using photoacoustic and dielectric spectroscopy for non-invasive glucose monitoring," San Francisco, California, United States, Mar. 2017, p. 1006445. doi: 10.1117/12.2251049.

[19] K. D. Pathirage, P. Roopasinghe, J. J. Sooriyaarachchi, R. Weththasinghe, and N. D. Nanayakkara, "Removing subject dependencies on Non-Invasive Blood Glucose Measurement using Hybrid Techniques," in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Berlin, Germany, Jul. 2019, pp. 7197–7200. doi: 10.1109/EMBC.2019.8856391.

[20] A. Tura, A. Maran, and G. Pacini, "Non-invasive glucose monitoring: Assessment of technologies and devices according to quantitative criteria," *Diabetes Res. Clin. Pract.*, vol. 77, no. 1, pp. 16–40, Jul. 2007, doi: 10.1016/j.diabres.2006.10.027.

[21] D. Bruen, C. Delaney, L. Florea, and D. Diamond, "Glucose Sensing for Diabetes Monitoring: Recent Developments," *Sensors*, vol. 17, no. 8, p. 1866, Aug. 2017, doi: 10.3390/s17081866.

[22] R. Kasahara, S. Kino, S. Soyama, and Y. Matsuura, "Noninvasive glucose monitoring using mid-infrared absorption spectroscopy based on a few wavenumbers," *Biomed. Opt. Express*, vol. 9, no. 1, p. 289, Jan. 2018, doi: 10.1364/BOE.9.000289.

[23] J. Y. Sim, C.-G. Ahn, E.-J. Jeong, and B. K. Kim, "In vivo Microscopic Photoacoustic Spectroscopy for Non-Invasive Glucose Monitoring Invulnerable to Skin Secretion Products," *Sci. Rep.*, vol. 8, no. 1, p. 1059, Dec. 2018, doi: 10.1038/s41598-018-19340-y.

[24] V. P. Rachim and W.-Y. Chung, "Wearable-band type visible-near infrared optical biosensor for non-invasive blood glucose monitoring," *Sens. Actuators B Chem.*, vol. 286, pp. 173–180, May 2019, doi: 10.1016/j.snb.2019.01.121.

[25] J. Yadav, A. Rani, V. Singh, and B. Mohan Murari, "Investigations on Multisensor-Based Noninvasive Blood Glucose Measurement System," *J. Med. Devices*, vol. 11, no. 3, p. 031006, Sep. 2017, doi: 10.1115/1.4036580.

[26] X. Xiao, Q. Yu, Q. Li, H. Song, and T. Kikkawa, "Precise Noninvasive Estimation of Glucose Using UWB Microwave With Improved Neural Networks and Hybrid Optimization," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–10, 2020, doi: 10.1109/TIM.2020.3010680.

[27] S. Hu, S. Nagae, and A. Hirose, "Millimeter-Wave Adaptive Glucose Concentration Estimation With Complex-Valued Neural Networks," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 7, pp. 2065–2071, Jul. 2019, doi: 10.1109/TBME.2018.2883085.

[28] M. Islam, M. S. Ali, N. J. Shoumy, S. Khatun, M. S. A. Karim, and B. S. Bari, "Non-invasive blood glucose concentration level estimation accuracy using ultra-wide band and

artificial intelligence," *SN Appl. Sci.*, vol. 2, no. 2, p. 278, Feb. 2020, doi: 10.1007/s42452-019-1884-3.

[29] S. Shobitha, P. M. Amita, K. B. Niranjana, and Mohd. A. Mohd. Ali, "Noninvasive Blood Glucose Prediction from Photoplethysmogram Using Relevance Vector Machine," in *2018 3rd International Conference for Convergence in Technology (I2CT)*, Pune, Apr. 2018, pp. 1–4. doi: 10.1109/I2CT.2018.8529481.

[30] "Extracellular fluid." Sep. 30, 2019. [Online]. Available: https://en.wikipedia.org/wiki/Extracellular_fluid

[31] K. Zirk and H. Poetzschke, "On the suitability of refractometry for the analysis of glucose in blood-derived fluids," *Med. Eng. Phys.*, vol. 26, no. 6, pp. 473–481, Jul. 2004, doi: 10.1016/j.medengphy.2004.03.008.

[32] H. F. Bunn, D. N. Haney, S. Kamin, K. H. Gabbay, and P. M. Gallop, "The biosynthesis of human hemoglobin A1c. Slow glycosylation of hemoglobin in vivo.," *J. Clin. Invest.*, vol. 57, no. 6, pp. 1652–1659, Jun. 1976, doi: 10.1172/JCI108436.

[33] E. Cengiz and W. V. Tamborlane, "A Tale of Two Compartments: Interstitial Versus Blood Glucose Monitoring," *Diabetes Technol. Ther.*, vol. 11, no. S1, p. S-11-S-16, Jun. 2009, doi: 10.1089/dia.2009.0002.

[34] P. A. J. Kolarsick, M. A. Kolarsick, and C. Goodwin, "Anatomy and Physiology of the Skin:," *J. Dermatol. Nurses Assoc.*, vol. 3, no. 4, pp. 203–213, Jul. 2011, doi: 10.1097/JDN.0b013e3182274a98.

[35] W. Groenendaal, G. von Basum, K. A. Schmidt, P. A. J. Hilbers, and N. A. W. van Riel, "Quantifying the Composition of Human Skin for Glucose Sensor Development," *J. Diabetes Sci. Technol.*, vol. 4, no. 5, pp. 1032–1040, Sep. 2010, doi: 10.1177/193229681000400502.

[36] W. G. John, “Glycated Haemoglobin Analysis,” *Ann. Clin. Biochem. Int. J. Biochem. Lab. Med.*, vol. 34, no. 1, pp. 17–31, Jan. 1997, doi: 10.1177/000456329703400105.

[37] I. Goodall, “HbA1c standardisation destination--global IFCC Standardisation. How, why, where and when--a tortuous pathway from kit manufacturers, via inter-laboratory lyophilized and whole blood comparisons to designated national comparison schemes,” *Clin. Biochem. Rev.*, vol. 26, no. 1, pp. 5–19, Feb. 2005.

[38] R. C. Hardison, “Evolution of hemoglobin and its genes,” *Cold Spring Harb. Perspect. Med.*, vol. 2, no. 12, p. a011627, Dec. 2012, doi: 10.1101/cshperspect.a011627.

[39] K. J. Welsh, M. S. Kirkman, and D. B. Sacks, “Role of Glycated Proteins in the Diagnosis and Management of Diabetes: Research Gaps and Future Directions,” *Diabetes Care*, vol. 39, no. 8, pp. 1299–1306, Aug. 2016, doi: 10.2337/dc15-2727.

[40] A. Stirban, T. Gawlowski, and M. Roden, “Vascular effects of advanced glycation endproducts: Clinical effects and molecular mechanisms,” *Mol. Metab.*, vol. 3, no. 2, pp. 94–108, Apr. 2014, doi: 10.1016/j.molmet.2013.11.006.

[41] R. J. Koenig, C. M. Peterson, R. L. Jones, C. Saudek, M. Lehrman, and A. Cerami, “Correlation of Glucose Regulation and Hemoglobin A $_{Ic}$ in Diabetes Mellitus,” *N. Engl. J. Med.*, vol. 295, no. 8, pp. 417–420, Aug. 1976, doi: 10.1056/NEJM197608192950804.

[42] “Introduction: *Standards of Medical Care in Diabetes—2018*,” *Diabetes Care*, vol. 41, no. Supplement 1, pp. S1–S2, Jan. 2018, doi: 10.2337/dc18-Sint01.

[43] D. M. Nathan *et al.*, “Translating the A1C assay into estimated average glucose values,” *Diabetes Care*, vol. 31, no. 8, pp. 1473–1478, Aug. 2008, doi: 10.2337/dc08-0545.

[44] C. L. Rohlfing, H.-M. Wiedmeyer, R. R. Little, J. D. England, A. Tennill, and D. E. Goldstein, “Defining the Relationship Between Plasma Glucose and HbA1c: Analysis of

glucose profiles and HbA1c in the Diabetes Control and Complications Trial," *Diabetes Care*, vol. 25, no. 2, pp. 275–278, Feb. 2002, doi: 10.2337/diacare.25.2.275.

[45] H. V. Roohk and A. R. Zaidi, "A review of glycated albumin as an intermediate glycation index for controlling diabetes," *J. Diabetes Sci. Technol.*, vol. 2, no. 6, pp. 1114–1121, Nov. 2008, doi: 10.1177/193229680800200620.

[46] T. Shafi *et al.*, "Serum Fructosamine and Glycated Albumin and Risk of Mortality and Clinical Outcomes in Hemodialysis Patients," *Diabetes Care*, vol. 36, no. 6, pp. 1522–1533, Jun. 2013, doi: 10.2337/dc12-1896.

[47] R. Reddavide *et al.*, "Tissue expression of glycated apolipoprotein B in colorectal adenoma and cancer," *Anticancer Res.*, vol. 31, no. 2, pp. 555–559, Feb. 2011.

[48] A. Sharif and K. Baboolal, "Diagnostic application of the A(1c) assay in renal disease," *J. Am. Soc. Nephrol. JASN*, vol. 21, no. 3, pp. 383–385, Mar. 2010, doi: 10.1681/ASN.2010010031.

[49] F. E. Vos, J. B. Schollum, C. V. Coulter, T. C. A. Doyle, S. B. Duffull, and R. J. Walker, "Red blood cell survival in long-term dialysis patients," *Am. J. Kidney Dis. Off. J. Natl. Kidney Found.*, vol. 58, no. 4, pp. 591–598, Oct. 2011, doi: 10.1053/j.ajkd.2011.03.031.

[50] S. M. Attard *et al.*, "Implications of iron deficiency/anemia on the classification of diabetes using HbA1c," *Nutr. Diabetes*, vol. 5, no. 6, pp. e166–e166, Jun. 2015, doi: 10.1038/nutd.2015.16.

[51] Alap L. Christy, Poornima A. Manjrekar, R. P. Babu, A. Hegde, and M. S. Rukmini, "Influence of Iron Deficiency Anemia on Hemoglobin A1C Levels in Diabetic Individuals with Controlled Plasma Glucose Levels," *Iran. Biomed. J. IBJ ISSN 1028-852X*, 1996, doi: 10.6091/ibj.1257.2014.

[52] M. B. Davidson and D. L. Schriger, “Effect of age and race/ethnicity on HbA1c levels in people without known diabetes mellitus: implications for the diagnosis of diabetes,” *Diabetes Res. Clin. Pract.*, vol. 87, no. 3, pp. 415–421, Mar. 2010, doi: 10.1016/j.diabres.2009.12.013.

[53] G. Cavagnolli, A. L. Pimentel, P. A. C. Freitas, J. L. Gross, and J. L. Camargo, “Effect of ethnicity on HbA1c levels in individuals without diabetes: Systematic review and meta-analysis,” *PloS One*, vol. 12, no. 2, p. e0171315, 2017, doi: 10.1371/journal.pone.0171315.

[54] K.-J. Tien, C.-Y. Yang, S.-F. Weng, S.-Y. Liu, M.-C. Hsieh, and C.-W. Chou, “The impact of ambient temperature on HbA1c in Taiwanese type 2 diabetic patients: The most vulnerable subgroup,” *J. Formos. Med. Assoc.*, vol. 115, no. 5, pp. 343–349, May 2016, doi: 10.1016/j.jfma.2015.03.010.

[55] F. R. J. Hinde, P. J. Standen, N. P. Mann, and D. I. Johnston, “Seasonal variation of haemoglobin A1 in children with insulin-dependent diabetes mellitus,” *Eur. J. Pediatr.*, vol. 148, no. 7, pp. 597–599, Jun. 1989, doi: 10.1007/BF00441507.

[56] S. K. Jain, R. McVie, J. J. Jaramillo, M. Palmer, and T. Smith, “Effect of modest vitamin E supplementation on blood glycated hemoglobin and triglyceride levels and red cell indices in type I diabetic patients,” *J. Am. Coll. Nutr.*, vol. 15, no. 5, pp. 458–461, Oct. 1996, doi: 10.1080/07315724.1996.10718624.

[57] G. N. Dakhale, H. V. Chaudhari, and M. Shrivastava, “Supplementation of Vitamin C Reduces Blood Glucose and Improves Glycosylated Hemoglobin in Type 2 Diabetes Mellitus: A Randomized, Double-Blind Study,” *Adv. Pharmacol. Sci.*, vol. 2011, pp. 1–5, 2011, doi: 10.1155/2011/195271.

[58] R. N. Pittman, “Regulation of Tissue Oxygenation Second Edition,” *Colloq. Ser. Integr. Syst. Physiol. Mol. Funct.*, vol. 8, no. 2, pp. i–99, Aug. 2016, doi: 10.4199/C00140ED2V01Y201606ISP065.

[59] M. Nitzan and H. Taitelbaum, “The measurement of oxygen saturation in arterial and venous blood,” *IEEE Instrum. Meas. Mag.*, vol. 11, no. 3, pp. 9–15, Jun. 2008, doi: 10.1109/MIM.2008.4534373.

[60] “understanding pulse oximetry spo2 concepts.” Sep. 30, 2019. Accessed: Sep. 30, 2019. [Online]. Available: http://incenter.medical.philips.com/doclib/enc/fetch/586262/586457/Understanding_Pulse_Oximetry.pdf%3Fnodeid%3D586458%26vernum%3D2

[61] T.-T. Wei, H.-Y. Tsai, C.-C. Yang, W.-T. Hsiao, and K.-C. Huang, “Noninvasive glucose evaluation by human skin oxygen saturation level,” in *2016 IEEE International Instrumentation and Measurement Technology Conference Proceedings*, Taipei, Taiwan, May 2016, pp. 1–5. doi: 10.1109/I2MTC.2016.7520571.

[62] Sandra L. Schutz, “Oxygen Saturation Monitoring by Pulse Oximetry.” 2011. [Online]. Available: https://docplayer.net/20722410-Oxygen-saturation-monitoring-by-pulse-oximetry.html

[63] E. D. Chan, M. M. Chan, and M. M. Chan, “Pulse oximetry: Understanding its basic principles facilitates appreciation of its limitations,” *Respir. Med.*, vol. 107, no. 6, pp. 789–799, Jun. 2013, doi: 10.1016/j.rmed.2013.02.004.

[64] J. Allen, “Photoplethysmography and its application in clinical physiological measurement,” *Physiol. Meas.*, vol. 28, no. 3, pp. R1–R39, Mar. 2007, doi: 10.1088/0967-3334/28/3/R01.

[65] O.-H. Lin, J.-Y. Lai, and H.-Y. Tsai, “Preventing Diabetes Extremity Vascular Disease with Blood Oxygen Saturation Images,” *Int. J. Instrum. Sci.*, p. 7, 2013.

[66] K. Ito *et al.*, “Factors Affecting Cerebral Oxygenation in Hemodialysis Patients: Cerebral Oxygenation Associates with pH, Hemodialysis Duration, Serum Albumin Concentration, and Diabetes Mellitus,” *PLOS ONE*, vol. 10, no. 2, p. e0117474, Feb. 2015, doi: 10.1371/journal.pone.0117474.

[67] E. Selvin, J. Coresh, S. H. Golden, L. L. Boland, F. L. Brancati, and M. W. Steffes, “Glycemic Control, Atherosclerosis, and Risk Factors for Cardiovascular Disease in Individuals With Diabetes: The Atherosclerosis Risk in Communities study,” *Diabetes Care*, vol. 28, no. 8, pp. 1965–1973, Aug. 2005, doi: 10.2337/diacare.28.8.1965.

[68] R. Norouzirad, P. González-Muniesa, and A. Ghasemi, “Hypoxia in Obesity and Diabetes: Potential Therapeutic Effects of Hyperoxia and Nitrate,” *Oxid. Med. Cell. Longev.*, vol. 2017, pp. 1–14, 2017, doi: 10.1155/2017/5350267.

[69] I. H. A. Heinonen, R. Boushel, and K. K. Kalliokoski, “The Circulatory and Metabolic Responses to Hypoxia in Humans – With Special Reference to Adipose Tissue Physiology and Obesity,” *Front. Endocrinol.*, vol. 7, Aug. 2016, doi: 10.3389/fendo.2016.00116.

[70] R. H. Eckel *et al.*, “Obesity and Type 2 Diabetes: What Can Be Unified and What Needs to Be Individualized?,” *Diabetes Care*, vol. 34, no. 6, pp. 1424–1430, Jun. 2011, doi: 10.2337/dc11-0447.

[71] K. M. Oltmanns *et al.*, “Hypoxia Causes Glucose Intolerance in Humans,” *Am. J. Respir. Crit. Care Med.*, vol. 169, no. 11, pp. 1231–1237, Jun. 2004, doi: 10.1164/rccm.200308-1200OC.

[72] K. Sada *et al.*, "Hyperglycemia Induces Cellular Hypoxia through Production of Mitochondrial ROS Followed by Suppression of Aquaporin-1," *PLOS ONE*, vol. 11, no. 7, p. e0158619, Jul. 2016, doi: 10.1371/journal.pone.0158619.

[73] L. Pu, Y. Shen, L. Lu, R. Zhang, Q. Zhang, and W. Shen, "Increased blood glycohemoglobin A1c levels lead to overestimation of arterial oxygen saturation by pulse oximetry in patients with type 2 diabetes," *Cardiovasc. Diabetol.*, vol. 11, no. 1, p. 110, 2012, doi: 10.1186/1475-2840-11-110.

[74] K. Maruo and Y. Yamada, "Near-infrared noninvasive blood glucose prediction without using multivariate analyses: introduction of imaginary spectra due to scattering change in the skin," *J. Biomed. Opt.*, vol. 20, no. 4, p. 047003, Apr. 2015, doi: 10.1117/1.JBO.20.4.047003.

[75] O. Mitsuhiro, Y. Takehiro, M. Kosuke, N. Koji, Y. Yasuhiro, and Y. Ken-Ichi, "Determination of Concentrations of Glucose and Human Serum Albumin in Mixtures in Phosphate-Buffered Solution by Near-Infrared Spectroscopy," *Sens. Mater.*, p. 323, 2012, doi: 10.18494/SAM.2012.794.

[76] L.-H. Lin, Y.-L. Lo, C.-C. Liao, and J.-X. Lin, "Optical detection of glucose concentration in samples with scattering particles," *Appl. Opt.*, vol. 54, no. 35, p. 10425, Dec. 2015, doi: 10.1364/AO.54.010425.

[77] J. W. Kang *et al.*, "Direct observation of glucose fingerprint using in vivo Raman spectroscopy," *Sci. Adv.*, vol. 6, no. 4, p. eaay5206, Jan. 2020, doi: 10.1126/sciadv.aay5206.

[78] T. Yilmaz, R. Foster, and Y. Hao, "Radio-Frequency and Microwave Techniques for Non-Invasive Measurement of Blood Glucose Levels," *Diagnostics*, vol. 9, no. 1, p. 6, Jan. 2019, doi: 10.3390/diagnostics9010006.

[79] K. Tonyushkina and J. H. Nichols, “Glucose Meters: A Review of Technical Challenges to Obtaining Accurate Results,” *J. Diabetes Sci. Technol.*, vol. 3, no. 4, pp. 971–980, Jul. 2009, doi: 10.1177/193229680900300446.

[80] M. Shokrekhodaei, S. Quinones, R. Martinek, and H. Nazeran, “A Robust PPG-based Heart Rate Monitor for Fitness and eHealth Applications,” in *2018 IEEE 20th International Conference on e-Health Networking, Applications and Services (Healthcom)*, Ostrava, Sep. 2018, pp. 1–5. doi: 10.1109/HealthCom.2018.8531082.

[81] S. L. Jacques, “Optical properties of biological tissues: a review,” *Phys. Med. Biol.*, vol. 58, no. 11, pp. R37–R61, Jun. 2013, doi: 10.1088/0031-9155/58/11/R37.

[82] O. Amir *et al.*, “Continuous Noninvasive Glucose Monitoring Technology Based on ‘Occlusion Spectroscopy,’” *J. Diabetes Sci. Technol.*, vol. 1, no. 4, pp. 463–469, Jul. 2007, doi: 10.1177/193229680700100403.

[83] J. Li *et al.*, “An Approach for Noninvasive Blood Glucose Monitoring Based on Bioimpedance Difference Considering Blood Volume Pulsation,” *IEEE Access*, vol. 6, pp. 51119–51129, 2018, doi: 10.1109/ACCESS.2018.2866601.

[84] J. Chung, H. So, Choi, and T. K. S. Wong, “Recent advances in noninvasive glucose monitoring,” *Med. Devices Evid. Res.*, p. 45, Jun. 2012, doi: 10.2147/MDER.S28134.

[85] R. Poddar, J. T. Andrews, P. Shukla, and P. Sen, “Non-Invasive Glucose Monitoring Techniques: A review and current trends,” *ArXiv08105755 Phys.*, Oct. 2008, Accessed: Oct. 01, 2019. [Online]. Available: http://arxiv.org/abs/0810.5755

[86] K. Singh, G. Sandhu, B. Lark, and S. Sud, “Molar extinction coefficients of some carbohydrates in aqueous solutions,” *Pramana*, vol. 58, no. 3, pp. 521–528, Mar. 2002, doi: 10.1007/s12043-002-0061-0.

[87] B. Kozma, L. Párta, D. Zalai, S. Gergely, and A. Salgó, "A Model System and Chemometrics to Develop near Infrared Spectroscopic Monitoring for Chinese Hamster Ovary Cell Cultivations," *J. Infrared Spectrosc.*, vol. 22, no. 6, pp. 401–410, Dec. 2014, doi: 10.1255/jnirs.1133.

[88] D. M. Haaland, M. R. Robinson, G. W. Koepp, E. V. Thomas, and R. P. Eaton, "Reagentless Near-Infrared Determination of Glucose in Whole Blood Using Multivariate Calibration," *Appl. Spectrosc.*, vol. 46, no. 10, pp. 1575–1578, Oct. 1992, doi: 10.1366/000370292789619232.

[89] X. Guo, A. Mandelis, A. Matvienko, K. Sivagurunathan, and B. Zinman, "Wavelength-modulated differential laser photothermal radiometry for blood glucose measurements," *J. Phys. Conf. Ser.*, vol. 214, p. 012025, Mar. 2010, doi: 10.1088/1742-6596/214/1/012025.

[90] A. Pfützner *et al.*, "Evaluation of a New Noninvasive Glucose Monitoring Device by Means of Standardized Meal Experiments," *J. Diabetes Sci. Technol.*, vol. 12, no. 6, pp. 1178–1183, Nov. 2018, doi: 10.1177/1932296818758769.

[91] Y. (Joseph) Segman, "Device and Method for Noninvasive Glucose Assessment," *J. Diabetes Sci. Technol.*, vol. 12, no. 6, pp. 1159–1168, Nov. 2018, doi: 10.1177/1932296818763457.

[92] N. S. Oliver, C. Toumazou, A. E. G. Cass, and D. G. Johnston, "Glucose sensors: a review of current and emerging technology," *Diabet. Med.*, vol. 26, no. 3, pp. 197–210, Mar. 2009, doi: 10.1111/j.1464-5491.2008.02642.x.

[93] J. D. Rogers, A. J. Radosevich, Ji Yi, and V. Backman, "Modeling Light Scattering in Tissue as Continuous Random Media Using a Versatile Refractive Index Correlation

Function," *IEEE J. Sel. Top. Quantum Electron.*, vol. 20, no. 2, pp. 173–186, Mar. 2014, doi: 10.1109/JSTQE.2013.2280999.

[94] R. J. McNichols and G. L. Coté, "Optical glucose sensing in biological fluids: an overview," *J. Biomed. Opt.*, vol. 5, no. 1, p. 5, 2000, doi: 10.1117/1.429962.

[95] C. D. Malchoff, K. Shoukri, J. I. Landau, and J. M. Buchert, "A Novel Noninvasive Blood Glucose Monitor," *Diabetes Care*, vol. 25, no. 12, pp. 2268–2275, Dec. 2002, doi: 10.2337/diacare.25.12.2268.

[96] M. H. Al-Hafidh, A. Glidle, R. Wilson, A. E. Kelly, J. Reboud, and J. M. Cooper, "Multireflection Polarimetry in Microfluidics," *IEEE Sens. Lett.*, vol. 3, no. 10, pp. 1–4, Oct. 2019, doi: 10.1109/LSENS.2019.2943688.

[97] G. L. Cote, M. D. Fox, and R. B. Northrop, "Noninvasive optical polarimetric glucose sensing using a true phase measurement technique," *IEEE Trans. Biomed. Eng.*, vol. 39, no. 7, pp. 752–756, Jul. 1992, doi: 10.1109/10.142650.

[98] R. W. Knighton and X.-R. Huang, "Linear birefringence of the central human cornea," *Invest. Ophthalmol. Vis. Sci.*, vol. 43, no. 1, pp. 82–86, Jan. 2002.

[99] V. Tuchin, Ed., *Handbook of Optical Sensing of Glucose in Biological Fluids and Tissues*, vol. 20082359. Taylor & Francis, 2008. doi: 10.1201/9781584889755.

[100] J. S. Baba, B. D. Cameron, S. Theru, and G. L. Coté, "Effect of temperature, pH, and corneal birefringence on polarimetric glucose monitoring in the eye," *J. Biomed. Opt.*, vol. 7, no. 3, p. 321, 2002, doi: 10.1117/1.1484163.

[101] P. Westphal, J.-M. Kaltenbach, and K. Wicker, "Corneal birefringence measured by spectrally resolved Mueller matrix ellipsometry and implications for non-invasive glucose

monitoring," *Biomed. Opt. Express*, vol. 7, no. 4, p. 1160, Apr. 2016, doi: 10.1364/BOE.7.001160.

[102] G. Purvinis, B. D. Cameron, and D. M. Altrogge, "Noninvasive Polarimetric-Based Glucose Monitoring: An *in Vivo* Study," *J. Diabetes Sci. Technol.*, vol. 5, no. 2, pp. 380–387, Mar. 2011, doi: 10.1177/193229681100500227.

[103] R. Pandey, "Raman Spectroscopy-Based Sensing of Glycated Hemoglobin: Critical Analysis and Future Outlook," *Postdoc J.*, vol. 3, no. 2, Feb. 2015, doi: 10.14304/SURYA.JPR.V3N2.2.

[104] J. Shao *et al.*, "In Vivo Blood Glucose Quantification Using Raman Spectroscopy," *PLoS ONE*, vol. 7, no. 10, p. e48127, Oct. 2012, doi: 10.1371/journal.pone.0048127.

[105] E. Wiercigroch *et al.*, "Raman and infrared spectroscopy of carbohydrates: A review," *Spectrochim. Acta. A. Mol. Biomol. Spectrosc.*, vol. 185, pp. 317–335, Oct. 2017, doi: 10.1016/j.saa.2017.05.045.

[106] J. Lipson *et al.*, "Requirements for Calibration in Noninvasive Glucose Monitoring by Raman Spectroscopy," *J. Diabetes Sci. Technol.*, vol. 3, no. 2, pp. 233–241, Mar. 2009, doi: 10.1177/193229680900300203.

[107] M. Teresa Montero, J. Hernández, and J. Estelrich, "Fluorescence quenching of albumin. A spectrofluorimetric experiment," *Biochem. Educ.*, vol. 18, no. 2, pp. 99–101, Apr. 1990, doi: 10.1016/0307-4412(90)90188-T.

[108] J. S. Maier, S. A. Walker, S. Fantini, M. A. Franceschini, and E. Gratton, "Possible correlation between blood glucose concentration and the reduced scattering coefficient of tissues in the near infrared," *Opt. Lett.*, vol. 19, no. 24, p. 2062, Dec. 1994, doi: 10.1364/OL.19.002062.

[109] V. Tuchin, *Tissue Optics*. 1000 20th Street, Bellingham, WA 98227-0010 USA: SPIE, 2007. doi: 10.1117/3.684093.

[110] O. S. Abdalsalam, A. M. Osman, R. M. Abd-Alhadi, and S. D. Alshmaa, "Design of simple noninvasive glucose measuring device," in *2013 INTERNATIONAL CONFERENCE ON COMPUTING, ELECTRICAL AND ELECTRONIC ENGINEERING (ICCEEE)*, Khartoum, Sudan, Aug. 2013, pp. 216–219. doi: 10.1109/ICCEEE.2013.6633935.

[111] K. V. Larin, M. S. Eledrisi, M. Motamedi, and R. O. Esenaliev, "Noninvasive Blood Glucose Monitoring With Optical Coherence Tomography: A pilot study in human subjects," *Diabetes Care*, vol. 25, no. 12, pp. 2263–2267, Dec. 2002, doi: 10.2337/diacare.25.12.2263.

[112] K. V. Larin, M. Motamedi, T. V. Ashitkov, and R. O. Esenaliev, "Specificity of noninvasive blood glucose sensing using optical coherence tomography technique: a pilot study," *Phys. Med. Biol.*, vol. 48, no. 10, pp. 1371–1390, May 2003, doi: 10.1088/0031-9155/48/10/310.

[113] A. I. Kholodnykh, I. Y. Petrova, K. V. Larin, M. Motamedi, and R. O. Esenaliev, "Precision of measurement of tissue optical properties with optical coherence tomography," *Appl. Opt.*, vol. 42, no. 16, p. 3027, Jun. 2003, doi: 10.1364/AO.42.003027.

[114] J. Yi and V. Backman, "Imaging a full set of optical scattering properties of biological tissue by inverse spectroscopic optical coherence tomography," *Opt. Lett.*, vol. 37, no. 21, p. 4443, Nov. 2012, doi: 10.1364/OL.37.004443.

[115] R. V. Kuranov, V. V. Sapozhnikova, D. S. Prough, I. Cicenaite, and R. O. Esenaliev, "In vivo study of glucose-induced changes in skin properties assessed with optical coherence tomography," *Phys. Med. Biol.*, vol. 51, no. 16, pp. 3885–3900, Aug. 2006, doi: 10.1088/0031-9155/51/16/001.

[116] V. V. Sapozhnikova, R. V. Kuranov, I. Cicenaite, R. O. Esenaliev, and D. S. Prough, "Effect on blood glucose monitoring of skin pressure exerted by an optical coherence tomography probe," *J. Biomed. Opt.*, vol. 13, no. 2, p. 021112, 2008, doi: 10.1117/1.2909671.

[117] J. Solanki, P. Sen, J. T. Andrews, and K. K. Thareja, "Blood glucose monitoring in human subjects using optical coherence tomography," *J. Opt.*, vol. 41, no. 3, pp. 127–133, Sep. 2012, doi: 10.1007/s12596-012-0067-z.

[118] D. K. Kamat, D. Bagul, and P. M. Patil, "Blood Glucose Measurement Using Bioimpedance Technique," *Adv. Electron.*, vol. 2014, pp. 1–5, Dec. 2014, doi: 10.1155/2014/406257.

[119] D. A. Dean, T. Ramanathan, D. Machado, and R. Sundararajan, "Electrical impedance spectroscopy study of biological tissues," *J. Electrost.*, vol. 66, no. 3–4, pp. 165–177, Mar. 2008, doi: 10.1016/j.elstat.2007.11.005.

[120] K. Chinen, I. Kinjo, A. Zamami, K. Irei, and K. Nagayama, "New equivalent-electrical circuit model and a practical measurement method for human body impedance," *Biomed. Mater. Eng.*, vol. 26, no. s1, pp. S779–S786, Aug. 2015, doi: 10.3233/BME-151369.

[121] E. Hernández-Balaguera, E. López-Dolado, and J. L. Polo, "Obtaining electrical equivalent circuits of biological tissues using the current interruption method, circuit theory and fractional calculus," *RSC Adv.*, vol. 6, no. 27, pp. 22312–22319, 2016, doi: 10.1039/C5RA24535D.

[122] D. Miklavčič, N. Pavšelj, and F. X. Hart, "Electric Properties of Tissues," in *Wiley Encyclopedia of Biomedical Engineering*, M. Akay, Ed. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2006, p. ebs0403. doi: 10.1002/9780471740360.ebs0403.

[123] T. Saito *et al.*, "Inverse distribution of serum sodium and potassium in uncontrolled inpatients with diabetes mellitus," *Endocr. J.*, vol. 46, no. 1, pp. 75–80, Feb. 1999, doi: 10.1507/endocrj.46.75.

[124] G. Liamis, "Diabetes mellitus and electrolyte disorders," *World J. Clin. Cases*, vol. 2, no. 10, p. 488, 2014, doi: 10.12998/wjcc.v2.i10.488.

[125] Y. Hayashi, L. Livshits, A. Caduff, and Y. Feldman, "Dielectric spectroscopy study of specific glucose influence on human erythrocyte membranes," *J. Phys. Appl. Phys.*, vol. 36, no. 4, pp. 369–374, Feb. 2003, doi: 10.1088/0022-3727/36/4/307.

[126] K. Song, U. Ha, S. Park, J. Bae, and H.-J. Yoo, "An Impedance and Multi-Wavelength Near-Infrared Spectroscopy IC for Non-Invasive Blood Glucose Estimation," *IEEE J. Solid-State Circuits*, vol. 50, no. 4, pp. 1025–1037, Apr. 2015, doi: 10.1109/JSSC.2014.2384037.

[127] I. Harman-Boehm, A. Gal, A. M. Raykhman, E. Naidis, and Y. Mayzel, "Noninvasive Glucose Monitoring: Increasing Accuracy by Combination of Multi-Technology and Multi-Sensors," *J. Diabetes Sci. Technol.*, vol. 4, no. 3, pp. 583–595, May 2010, doi: 10.1177/193229681000400312.

[128] P. Mehrotra, B. Chatterjee, and S. Sen, "EM-Wave Biosensors: A Review of RF, Microwave, mm-Wave and Optical Sensing," *Sensors*, vol. 19, no. 5, p. 1013, Feb. 2019, doi: 10.3390/s19051013.

[129] C. Jang, J.-K. Park, H.-J. Lee, G.-H. Yun, and J.-G. Yook, "Temperature-Corrected Fluidic Glucose Sensor Based on Microwave Resonator," *Sensors*, vol. 18, no. 11, p. 3850, Nov. 2018, doi: 10.3390/s18113850.

[130] T. Yilmaz, R. Foster, and Y. Hao, "Broadband Tissue Mimicking Phantoms and a Patch Resonator for Evaluating Noninvasive Monitoring of Blood Glucose Levels," *IEEE Trans. Antennas Propag.*, vol. 62, no. 6, pp. 3064–3075, Jun. 2014, doi: 10.1109/TAP.2014.2313139.

[131] E. Topsakal, T. Karacolak, and E. C. Moreland, "Glucose-dependent dielectric properties of blood plasma," in *2011 XXXth URSI General Assembly and Scientific Symposium*, Istanbul, Aug. 2011, pp. 1–4. doi: 10.1109/URSIGASS.2011.6051324.

[132] S. Gabriel, R. W. Lau, and C. Gabriel, "The dielectric properties of biological tissues: II. Measurements in the frequency range 10 Hz to 20 GHz," *Phys. Med. Biol.*, vol. 41, no. 11, pp. 2251–2269, Nov. 1996, doi: 10.1088/0031-9155/41/11/002.

[133] E. Reyes-Vera, G. Acevedo-Osorio, M. Arias-Correa, and D. E. Senior, "A Submersible Printed Sensor Based on a Monopole-Coupled Split Ring Resonator for Permittivity Characterization," *Sensors*, vol. 19, no. 8, p. 1936, Apr. 2019, doi: 10.3390/s19081936.

[134] H. Choi, S. Luzio, J. Beutler, and A. Porch, "Microwave noninvasive blood glucose monitoring sensor: Human clinical trial results," in *2017 IEEE MTT-S International Microwave Symposium (IMS)*, Honololu, HI, USA, Jun. 2017, pp. 876–879. doi: 10.1109/MWSYM.2017.8058721.

[135] H. Choi *et al.*, "Design and In Vitro Interference Test of Microwave Noninvasive Blood Glucose Monitoring Sensor," *IEEE Trans. Microw. Theory Tech.*, vol. 63, no. 10, pp. 3016–3025, Oct. 2015, doi: 10.1109/TMTT.2015.2472019.

[136] A. La Gioia *et al.*, "Open-Ended Coaxial Probe Technique for Dielectric Measurement of Biological Tissues: Challenges and Common Practices," *Diagnostics*, vol. 8, no. 2, p. 40, Jun. 2018, doi: 10.3390/diagnostics8020040.

[137] F. Costa, M. Borgese, M. Degiorgi, and A. Monorchio, “Electromagnetic Characterisation of Materials by Using Transmission/Reflection (T/R) Devices,” *Electronics*, vol. 6, no. 4, p. 95, Nov. 2017, doi: 10.3390/electronics6040095.

[138] M. Hofmann, G. Fischer, R. Weigel, and D. Kissinger, “Microwave-Based Noninvasive Concentration Measurements for Biomedical Applications,” *IEEE Trans. Microw. Theory Tech.*, vol. 61, no. 5, pp. 2195–2204, May 2013, doi: 10.1109/TMTT.2013.2250516.

[139] V. Turgul and I. Kale, “Characterization of the complex permittivity of glucose/water solutions for noninvasive RF/Microwave blood glucose sensing,” in *2016 IEEE International Instrumentation and Measurement Technology Conference Proceedings*, Taipei, Taiwan, May 2016, pp. 1–5. doi: 10.1109/I2MTC.2016.7520546.

[140] P. H. Siegel, Y. Lee, and V. Pikov, “Millimeter-wave non-invasive monitoring of glucose in anesthetized rats,” in *2014 39th International Conference on Infrared, Millimeter, and Terahertz waves (IRMMW-THz)*, Tucson, AZ, USA, Sep. 2014, pp. 1–2. doi: 10.1109/IRMMW-THz.2014.6956294.

[141] E. J. Rothwell, J. L. Frasch, S. M. Ellison, P. Chahal, and R. O. Ouedraogo, “ANALYSIS OF THE NICOLSON-ROSS-WEIR METHOD FOR CHARACTERIZING THE ELECTROMAGNETIC PROPERTIES OF ENGINEERED MATERIALS,” *Prog. Electromagn. Res.*, vol. 157, pp. 31–47, 2016, doi: 10.2528/PIER16071706.

[142] K. B. Beć and C. W. Huck, “Breakthrough Potential in Near-Infrared Spectroscopy: Spectra Simulation. A Review of Recent Developments,” *Front. Chem.*, vol. 7, p. 48, Feb. 2019, doi: 10.3389/fchem.2019.00048.

[143] J. Chung, H. So, Choi, and T. K. S. Wong, “Recent advances in noninvasive glucose monitoring,” *Med. Devices Evid. Res.*, p. 45, Jun. 2012, doi: 10.2147/MDER.S28134.

[144] W. Villena Gonzales, A. Mobashsher, and A. Abbosh, "The Progress of Glucose Monitoring—A Review of Invasive to Minimally and Non-Invasive Techniques, Devices and Sensors," *Sensors*, vol. 19, no. 4, p. 800, Feb. 2019, doi: 10.3390/s19040800.

[145] J. Dai, Z. Ji, Y. Du, and S. Chen, "In vivo noninvasive blood glucose detection using near-infrared spectrum based on the PSO-2ANN model," *Technol. Health Care*, vol. 26, pp. 229–239, May 2018, doi: 10.3233/THC-174592.

[146] M. F. G. Wood, D. Co^té, and I. A. Vitkin, "Combined optical intensity and polarization methodology for analyte concentration determination in simulated optically clear and turbid biological media," *J. Biomed. Opt.*, vol. 13, no. 4, p. 044037, 2008, doi: 10.1117/1.2968198.

[147] L.-N. Li, Q.-B. Li, and G.-J. Zhang, "A Weak Signal Extraction Method for Human Blood Glucose Noninvasive Measurement using Near Infrared Spectroscopy," *J. Infrared Millim. Terahertz Waves*, vol. 30, no. 11, pp. 1191–1204, Nov. 2009, doi: 10.1007/s10762-009-9544-0.

[148] "SparkFun Triad Spectroscopy Sensor." Mar. 07, 2021. [Online]. Available: https://www.sparkfun.com/products/15050

[149] "AMS AG website to purchase AS7265x photodetectors." May 17, 2021. [Online]. Available: https://ams.com/spectral-sensing

[150] "Spectral Triad (AS7265x) Hookup Guide." Mar. 07, 2021. [Online]. Available: https://learn.sparkfun.com/tutorials/spectral-triad-as7265x-hookup-guide?_ga=2.209320475.821052045.1585464539-1826701728.1577990555

[151] "Datasheet for SparkFun Triad Spectroscopy Sensor." Mar. 07, 2021. [Online]. Available: https://cdn.sparkfun.com/assets/learn_tutorials/8/3/0/AS7265x_Datasheet.pdf

[152] “SparkFun RedBoard Qwiic.” Mar. 07, 2021. [Online]. Available: https://www.sparkfun.com/products/15123?_ga=2.36951712.1851648704.1577990555-1826701728.1577990555

[153] “Datasheet for UV LED.” Mar. 07, 2021. [Online]. Available: http://www.vishay.com/docs/82556/vlmu3100.pdf

[154] “Datasheet for WHITE LED.” Mar. 07, 2021. [Online]. Available: https://www.lumileds.com/uploads/459/DS208-pdf

[155] “Datasheet for NIR LED.” Mar. 07, 2021. [Online]. Available: http://www.everlight.com/file/ProductFile/201407052116051910.pdf

[156] “EasyEDA - Online PCB design & circuit simulator.” Mar. 07, 2021. [Online]. Available: https://easyeda.com/

[157] “Spectrophotometer Cells.” Mar. 07, 2021. [Online]. Available: https://www.starna.com/fluorimeter/macro

[158] “Peristaltic Pumps | Bio-Chem Fluidics.” Mar. 07, 2021. [Online]. Available: https://biochemfluidics.com/products/peristaltic-pumps

[159] “Onshape | Product Development Platform.” Mar. 07, 2021. [Online]. Available: https://www.onshape.com/

[160] M. Shokrekhodaei, D. P. Cistola, R. C. Roberts, and S. Quinones, “Non-Invasive Glucose Monitoring Using Optical Sensor and Machine Learning Techniques for Diabetes Applications,” *IEEE Access*, vol. 9, pp. 73029–73045, 2021, doi: 10.1109/ACCESS.2021.3079182.

[161] V. B. S. Prasath *et al.*, "Distance and Similarity Measures Effect on the Performance of K-Nearest Neighbor Classifier -- A Review," *Big Data*, vol. 7, no. 4, pp. 221–248, Dec. 2019, doi: 10.1089/big.2018.0175.

[162] K. Boonchuay, K. Sinapiromsaran, and C. Lursinsap, "Decision tree induction based on minority entropy for the class imbalance problem," *Pattern Anal. Appl.*, vol. 20, no. 3, pp. 769–782, Aug. 2017, doi: 10.1007/s10044-016-0533-3.

[163] O. Devos, C. Ruckebusch, A. Durand, L. Duponchel, and J.-P. Huvenne, "Support vector machines (SVM) in near infrared (NIR) spectroscopy: Focus on parameters optimization and model interpretation," *Chemom. Intell. Lab. Syst.*, vol. 96, no. 1, pp. 27–33, Mar. 2009, doi: 10.1016/j.chemolab.2008.11.005.

[164] E. C. Alexopoulos, "Introduction to multivariate regression analysis," *Hippokratia*, vol. 14, no. Suppl 1, pp. 23–28, Dec. 2010.

[165] R. J. Casson and L. D. Farmer, "Understanding and checking the assumptions of linear regression: a primer for medical researchers: Assumptions of linear regression," *Clin. Experiment. Ophthalmol.*, vol. 42, no. 6, pp. 590–596, Aug. 2014, doi: 10.1111/ceo.12358.

[166] J. Schmidt, M. R. G. Marques, S. Botti, and M. A. L. Marques, "Recent advances and applications of machine learning in solid-state materials science," *Npj Comput. Mater.*, vol. 5, no. 1, p. 83, Dec. 2019, doi: 10.1038/s41524-019-0221-0.

[167] O. I. Abiodun *et al.*, "Comprehensive Review of Artificial Neural Network Applications to Pattern Recognition," *IEEE Access*, vol. 7, pp. 158820–158846, 2019, doi: 10.1109/ACCESS.2019.2945545.

[168] K. G. Sheela and S. N. Deepa, "Review on Methods to Fix Number of Hidden Neurons in Neural Networks," *Math. Probl. Eng.*, vol. 2013, pp. 1–11, 2013, doi: 10.1155/2013/425740.

[169] T. K. Gupta and K. Raza, "Optimization of ANN Architecture: A Review on Nature-Inspired Techniques," in *Machine Learning in Bio-Signal Analysis and Diagnostic Imaging*, Elsevier, 2019, pp. 159–182. doi: 10.1016/B978-0-12-816086-2.00007-2.

[170] N. Ali, D. Neagu, and P. Trundle, "Evaluation of k-nearest neighbour classifier performance for heterogeneous data sets," *SN Appl. Sci.*, vol. 1, no. 12, p. 1559, Dec. 2019, doi: 10.1007/s42452-019-1356-9.

[171] Z. Geler, V. Kurbalija, M. Radovanović, and M. Ivanović, "Comparison of different weighting schemes for the kNN classifier on time-series data," *Knowl. Inf. Syst.*, vol. 48, no. 2, pp. 331–378, Aug. 2016, doi: 10.1007/s10115-015-0881-0.

[172] L. Yang and A. Shami, "On hyperparameter optimization of machine learning algorithms: Theory and practice," *Neurocomputing*, vol. 415, pp. 295–316, Nov. 2020, doi: 10.1016/j.neucom.2020.07.061.

[173] T. Hastie, S. Rosset, R. Tibshirani, and J. Zhu, "The Entire Regularization Path for the Support Vector Machine," *J Mach Learn Res*, vol. 5, pp. 1391–1415, Dec. 2004.

[174] C. Saunders, M. O. Stitson, and J. Weston, "Support Vector Machine Reference Manual." Mar. 07, 2021. [Online]. Available: https://eprints.soton.ac.uk/258959/1/SVM_Reference.pdf

[175] J. L. Parkes, S. L. Slatin, S. Pardo, and B. H. Ginsberg, "A new consensus error grid to evaluate the clinical significance of inaccuracies in the measurement of blood glucose," *Diabetes Care*, vol. 23, no. 8, pp. 1143–1148, Aug. 2000, doi: 10.2337/diacare.23.8.1143.

[176] H. Khadem, M. R. Eissa, H. Nemat, O. Alrezj, and M. Benaissa, "Classification before regression for improving the accuracy of glucose quantification using absorption spectroscopy," *Talanta*, vol. 211, p. 120740, May 2020, doi: 10.1016/j.talanta.2020.120740.

[177] P. P. Pai, P. K. Sanki, S. K. Sahoo, A. De, S. Bhattacharya, and S. Banerjee, "Cloud Computing-Based Non-Invasive Glucose Monitoring for Diabetic Care," *IEEE Trans. Circuits Syst. Regul. Pap.*, vol. 65, no. 2, pp. 663–676, Feb. 2018, doi: 10.1109/TCSI.2017.2724012.

[178] V. Turgul and I. Kale, "All-digital 1550 nm optical aqueous glucose solution measurement system," in *2017 24th IEEE International Conference on Electronics, Circuits and Systems (ICECS)*, Batumi, Dec. 2017, pp. 462–465. doi: 10.1109/ICECS.2017.8292032.

[179] S. Liakat, K. A. Bors, T.-Y. Huang, A. P. M. Michel, E. Zanghi, and C. F. Gmachl, "In vitro measurements of physiological glucose concentrations in biological fluids using mid-infrared light," *Biomed. Opt. Express*, vol. 4, no. 7, p. 1083, Jul. 2013, doi: 10.1364/BOE.4.001083.

[180] S. Pleus *et al.*, "Performance Evaluation of a Continuous Glucose Monitoring System under Conditions Similar to Daily Life," *J. Diabetes Sci. Technol.*, vol. 7, no. 4, pp. 833–841, Jul. 2013, doi: 10.1177/193229681300700405.

[181] P. Liu, Z. Zhu, C. Zeng, and G. Nie, "Specific absorption spectra of hemoglobin at different PO2 levels: potential noninvasive method to detect PO2 in tissues," *J. Biomed. Opt.*, vol. 17, no. 12, p. 125002, Dec. 2012, doi: 10.1117/1.JBO.17.12.125002.

[182] B. W. Pogue and M. S. Patterson, "Review of tissue simulating phantoms for optical spectroscopy, imaging and dosimetry," *J. Biomed. Opt.*, vol. 11, no. 4, p. 041102, 2006, doi: 10.1117/1.2335429.

Vita

Maryamsadat Shokrekhodaei received the M.S. degree in electrical engineering from Sharif University of Technology, Tehran, Iran, in 2016. In 2017, she entered the doctoral program in electrical and computer engineering at the University of Texas at El Paso (UTEP). Her research interests include developing of non-invasive biomedical sensors, artificial intelligence, and RFIC circuits and systems design. Her Ph.D. research focuses on accuracy improvement of optical based in-vitro glucose sensing system for non-invasive monitoring of diabetes. She was awarded Dodson Research Grant as partial funding for her research in 2018. While at UTEP, she had the privilege of teaching hundreds of undergraduate students as a teaching assistant, and contributing to the development of two online based courses: "Circuit, Electronic Materials and Devices", and "Machine Learning with Python".

Maryamsadat's work has been recognized in two separate publications, the first "Review of Non-invasive Glucose Sensing Techniques: Optical, Electrical and Breath Acetone" was published in *Sensors* in 2020. Her second publication, "Non-Invasive Glucose Monitoring Using Optical Sensor and Machine Learning Techniques for Diabetes Applications," was published in *IEEE Access* in 2021. Maryamsadat is noted as the first author and corresponding author in both of these works. Furthermore, she has coauthored articles such as, "The Conceptual Design of a Novel Workstation for Seizure Prediction Using Machine Learning With Potential eHealth Applications" published in IEEE Journal of Translational Engineering in Health and Medicine in 2019.

Contact Information: mshokrekhod@miners.utep.edu