Non-invasive Glucose Monitoring using Microwave Sensor with Machine Learning

by

Saeed Mohammed Bamatraf

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Electrical and Computer Engineering

Waterloo, Ontario, Canada, 2022

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner: Tayeb A. Denidni
Professor, INRS-EMT,
University of Quebec (Universite de Quebec)

Supervisor: Omar M. Ramahi
Professor, Dept. of Electrical and Computer Engineering,
University of Waterloo

Internal Member: Zhou Wang
Professor, Dept. of Electrical and Computer Engineering,
University of Waterloo

Mark Crowley
Assistant Professor, Dept. of Electrical and Computer Engineering,
University of Waterloo

Internal-External Member: Daniel Stashuk
Professor, Dept. of Systems Design Engineering,
University of Waterloo

ii

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Diabetes is a chronic condition that occurs when the levels of glucose are high in the blood because the body cannot produce any or enough insulin or use insulin effectively. According to the International Diabetes Federation (IDF), 537 million people are currently living with diabetes. It is very important for people with diabetes to regularly check their blood glucose levels to keep track of any increase or decrease in these levels, and adjust the amount of medication accordingly. This process is called Continuous Glucose Monitoring (CGM). CGM techniques can be categorized based on invasiveness as invasive, minimally invasive, and non-invasive. In the non-invasive there is no need for any blood sample extraction or any implantation of electrodes in the body.

First, a review of the non-invasive CGM techniques in the last ten years is conducted in order to understand the current status of the CGM and highlight the challenges that face every technique in order to come up with a better solution. The techniques used for non-invasive CGM can be classified into six major categories: optical, microwave, thermal, transdermal, hybrid and other. In order to overcome the shortcomings of the invasive and minimally-invasive methods of CGM, such as pain, discomfort, and risk of infection, non-invasive CGM is needed. However, due to the multiple challenges such as accuracy, usability and applicability, contemporary non-invasive glucose monitors are still not sufficiently reliable.

In this thesis, a non-invasive glucose monitoring system is developed using microwave sensor with machine learning techniques. The system has two parts: hardware, which is the microwave sensor, and software, which is the machine learning algorithms. The physical sensor is microwaves-based using inexpensive printed circuit board technology. Electrically-small dipole and another spiral microwave sensor were designed and used in this thesis taking into account different factors like frequency range, penetration and safety of the human. Machine learning techniques were used to select the most distinguish features in order to predict the actual glucose level in the human. Different feature engineering types were used to extract the discriminate features that will be inputted to different regression algorithms to predict the glucose levels.

The main idea of the thesis is based on studying dielectric properties (permittivity and conductivity) of the human body tissues in order to find a relation with the corresponding glucose level in those tissues. This is done using CST simulation along with experiments.

Experimental results on aqueous solutions (water-glucose solutions) used as a proof of concept and to check the ability of the microwave sensors to detect the different concentrations of these simple water- glucose solutions.

iv

In simulation, a hand model system was designed with different tissues/layers to simulate the effects of the microwave sensor with respect to changing in dielectric properties (permittivity and conductivity) of those tissues/layers. Different systems (corresponding to different hand layers/tissues) were trained and tested using cross validation, and the Root Mean Square Error (RMSE) was acceptable.

Acknowledgements

Dedication

This is dedicated to:

- My great beloved parents, who never stop giving of themselves in countless ways.

- My dearest wife, who stayed with me in ease and difficulty.

- My beloved brothers and sisters, whom I can not find a replacement.

- My beloved sweet daughters: Lama, and Nemah, whom I can't force myself to stop loving.

- All my family, the symbol of love and giving.

- My friends who encourage and support me.

- All the people in my life who touch my heart.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

CGM Continuous Glucose Monitoring iv, xi, 2–5, 8–17, 23, 28, 30, 40–42, 53, 99, 100

DAK Dielectric Assessment Kit 20

ECG Electrocardiography viii, 3, 14

GDM Gestational Diabetes 1

IDF International Diabetes Federation iv, 1

IR Infrared 4, 5, 13, 14

ISF Interstitial Fluid 2, 12

KNN K-Nearest-Neighbour 35

MARD Mean Absolute Relative Difference 40, 41

MHC Metabolic Heat Confirmation 10

MIR Mid Infrared 4, 5, 15, 16

NIR Near Infrared 4, 5, 15, 16

OCT Optical Coherence Tomography 7, 8, 15, 16

PAI Photoacoustic Imaging 5, 6, 15, 16

PCA Principle Component Analysis 46, 57, 76

Chapter 1

Introduction

1.1 Overview

Glucose ($C_6H_{12}O_6$) is a simple sugar and one of the most abundant carbon-based molecules in nature. It is a primary source of energy in the human body. Glucose is occurred naturally in the form of D-Glucose and also called dextrose.

Diabetes mellitus, or diabetes for simplicity, is a chronic condition that occurs when the levels of glucose are high in the blood because the body cannot produce any or enough insulin or cannot use insulin effectively [4].

There are three main types of diabetes, Type-1 Diabetes (T1D), Type-2 Diabetes (T2D) and Gestational Diabetes (GDM). In T1D, the body does not produce insulin as the immune system attacks and destroys the cells in the pancreas that secrete insulin. This type of diabetes may be appeared since birth or developed later in life. With the more common T2D, the body does not make or use insulin well because the cells' glucose receptors are malfunctioned. Combination of factors may cause T2D such as lack of exercises, eating habit and genetic factors. GDM is a form of diabetes that occurs during pregnancy and most women no longer have it after the baby is born. According to the IDF, 537 million people are currently living with diabetes [5]. Countries at all income levels suffer from high human, social and economic costs for diabetes in all forms. Energy is very important to the human body and it comes in the form of glucose. When glucose transported via blood stream, cells process and absorb the glucose via a process called Glycolysis.

The normal glucose levels in the human blood are between 4.0 and 7.0 mmol/L (72 and 126 mg/dL) on an empty stomach, and between 5.0 and 10.0 mmol/L (90 and 180 mg/dL) two hours after a meal [6]. Hyperglycemia occurs when the blood glucose level

is above normal values. Whereas Hypoglycemia occurs when the blood glucose level is below normal values. It is very important for people with diabetes to regularly check their blood glucose levels to keep track of any increase or decrease in these levels, and adjust the amount of medication accordingly. This process is called CGM.

CGM techniques can be categorized as invasive, minimally invasive, and non-invasive. Invasive techniques are the most widely used blood glucose measurement because they have the highest accuracy. In the invasive, a blood sample is needed to be extracted using a lancet and the fingertip is the most widely used human body part. The blood sample is given to a glucometer to measure the glucose level. Due to the pain and inconvenience, caused to the people with diabetes, by the frequent finger prick, minimally invasive techniques are introduced to help in reducing the pain. In the minimally invasive, measuring the glucose level is done by subcutaneously needle-type electrodes implanted in the body. However, these techniques are still not recommended for most of the people with diabetes because it involves direct interaction with tissues and needed to be replaced from time to time. In addition, minimally invasive devices' results are not as accurate as the invasive techniques ones. The last and most recent category is the non-invasive CGM techniques in which there is no need for any blood sample extraction or any implantation of electrodes in the body.

There is a desperate need for a simple, non-invasive and pain-free CGM method, which will make a major improvement in the field of diabetes care. Many research groups and companies are nowadays working on developing non-invasive CGM methods using different body fluids. The main fluid is the blood; however, other fluids can be used as an alternative to measure glucose levels instead of extracting blood. Those fluids include saliva [7, 8], sweat [9, 10, 11, 12], urine [13], tears [14, 15, 16, 17, 18], breath [19] and Interstitial Fluid (ISF) [20]. In addition, some research used skin oxygen saturation [21].

However, on the other hand, there are a lot of challenges facing the development of non-invasive CGM devices such as accuracy, usability, and applicability. Clearly accuracy is a primary consideration since the insulin dosage directly depends on the measured glucose value.

This research will shed the lights on some of the past and current non-invasive CGM techniques (in the period 2009-2021) and the major challenges facing these techniques, to build a solid understanding of the nature of the problem in order to have good non-invasive techniques for CGM which will help improving the field of diabetes technology. First of all, in this thesis what is considered to be non-invasive is that any techniques that doesn't involve any drawing of any human body sample or any electrode implantation in the body.

1.2   CGM Classifications

Based on the nature of non-invasive CGM techniques, we can classify them into five major categories: optical [22, 23], microwave [24], thermal [25], transdermal [26], and hybrid [27, 28]. In addition, some research use ECG as a measure for glucose in T1D [29, 30]. The classification in this thesis is different from other research where they classified based on the properties of glucose, blood-tissues, and breath acetone [31].

Multivariate calibration methods for spectral analysis are used to establish a regression model that will map or relate the measured input signal (depending on the technique used) with the corresponding glucose level. This quantitative model will be used later to predict the glucose levels based on the reading signals. There are many methods such as principal component regression (PCR), partial least squares (PLS), multiple linear regression (MLR), support vector machine (SVM) and artificial neural network (ANN).

1.3   Non-invasive Optical-based Glucose Monitoring:

It is known that light interacts with molecules based on the molecule's chemical structure, which makes light a good choice for detection. When light is incident on a sample, one of four scenarios take place: the light is either absorbed, transmitted, reflected or scattered. The advantages of using optical-based methods are that influence of pH and temperature fluctuations are not critical for the measurement. In addition, the optical measurements are flexible and, in general, the optical signals' power used is harmless to the human body when taking into account both wavelength and radiant power [32].

In this section, I will discuss different types of optical-based techniques used for CGM and any associated devices using these techniques:

1.3.1   Vibrational Spectroscopy:

Vibrational Spectroscopy can characterize and identify the molecules. It gives a dynamic picture of the molecule while X-Ray gives a static picture [33]. In glucose detection, glucose anomers can be clearly identified and distinguished using vibrational spectroscopy due to small differences in the spectra [32]. There are two types of vibrational spectroscopy: Infrared and Raman.

i. Infrared Spectroscopy: Infrared (IR) is the most widely used vibrational spectroscopy. In IR spectroscopy, the absorption of electromagnetic radiation between the wavelengths of 700 nm to 1 mm is investigated [32]. Near Infrared (NIR) (750nm – 2500 nm) allows the measurement of glucose in the blood under the skin to a depth of few mm in range. NIR has the advantage of high sensitivity of the photoconductive detectors, and there is a reasonable transparency of water to the signal bandwidth used by NIR. In addition, NIR has a low cost which makes it a widely used technique in the field of non-invasive monitoring [34].

Despite the fact that NIR is simple and harmless, there are many limitations when using it for glucose detection. Factors like body temperature, skin thickness and hydration, blood pressure and fatty tissues affect the accuracy of the results because they affect light absorption.

Dantu et al [35] developed a non-invasive NIR CGM technique using a smartphone. A laser beam is directed to the fingertip and the smartphone camera captures the light from the other side. MATLAB is used to analyze the result and by using a modified Beer–Lambert law, a linear proportionality between the results and the actual glucose concentration is extracted. The device is still not sufficiently reliable and is affected by light conditions.

In [36], a hybrid algorithm using NIR spectroscopy is proposed to improve the prediction performance of blood glucose concentration. The algorithm is a combination of wavelet prism modified uninformative variable elimination approach (WP-mUVE) and least squares support vector machine (LSSVM). Data is collected from the human tongue, which eliminates the effects of skin thickness and fat and provides high Signal to Noise Ratio (SNR). Results provided improvement in accuracy and adaptability of the prediction. The system may be affected by environmental parameters such as temperature, humidity and pressure.

Another NIR technique is the one with aqueous solution experiment in [37] which used a light beam of 1310 nm wavelength. The results showed a direct proportionality between the output power and glucose concentration. However, to verify the accuracy of the method, it needs to be tested using real human blood instead of an aqueous solution only.

The second type of IR is Mid Infrared (MIR) spectroscopy, which is based on light in the 2500-10000 nm spectrum. MIR has the physical principle of NIR; however, due to higher wavelengths, MIR increased the absorption. On the other hand, penetration of light in tissues can reach a few micrometers [38]. As a result, the reflected light can only be considered because there is no light transmitted through the body [34]. Due to this penetration problem, MIR is less studied for glucose detection compared to NIR.

The biggest challenges for glucose detection with IR are the high scattering and low glucose concentrations. Moreover, the glucose signal is masked by a large peak caused by water [32]. Researcher at CADIPT [39] used a pair of laser beams instead of a needle with wavelength-modulated differential photothermal radiometry (WM-DPTR). One of the beams is absorbed by both water and glucose, while the other only by water so that the water absorptions cancel each other out. Unfortunately, there is no update about the current status of this technique.

One of the devices developed using NIR technology is Combo Glucometer [40, 41]. Cnoga Medical Ltd. developed a device with LCD display for non-invasive capillary blood glucose measurement. The Combo Glucometer can be connected using a USB interfaced with PC, smart-phone or web, it has rechargeable battery and it weighs less than 100 gr. However, for good accuracy, the Combo Glucometer requires measuring glucose levels invasively and non-invasively simultaneously usually for 3 days calibration to personalize the device. The device is currently CE approved while FDA approval is pending.

ii. Raman Spectroscopy: Raman spectroscopy is another vibrational technique that uses a single wavelength laser source (from visible to MIR) to excite the electrons and detecting the Raman shifted (inelastic scattered) light as shown in Figure 1.1. Raman is an analogous technique to IR, but with a less intuitive physical background [32, 38]. Compared to NIR spectroscopy, Raman spectroscopy has the advantage of providing sharper and less overlapped spectra. In addition, there is a proportionality between the intensity of spectral features and the concentration of particular species. In addition, the temperature change does not have much effect on the spectra because of the small sensitivity.

The main drawbacks of Raman spectroscopy are that the laser wavelength and intensity are instable, and the spectral acquisition times are too long. Moreover, SNR ratio is significantly reduced because of the low power of the light source. Finally, as with NIR, there is the problem of interference from other compounds [34].

C8 MediSensors was a non-invasive CGM device developed using Raman spectroscopy technology. The device received a CE-approval, but it is no longer available and the company went out of business in 2013 [42], which rises a lot of questions about CE-approval.

1.3.2 Photoacoustic Spectroscopy:

Photoacoustic Imaging (PAI), or photoacoustic spectroscopy, is based on the principle of thermal expansion of an object caused by the absorption of light (Figure 1.2). An oscillating movement in the tissue is induced by the emitting light, which results in pressure
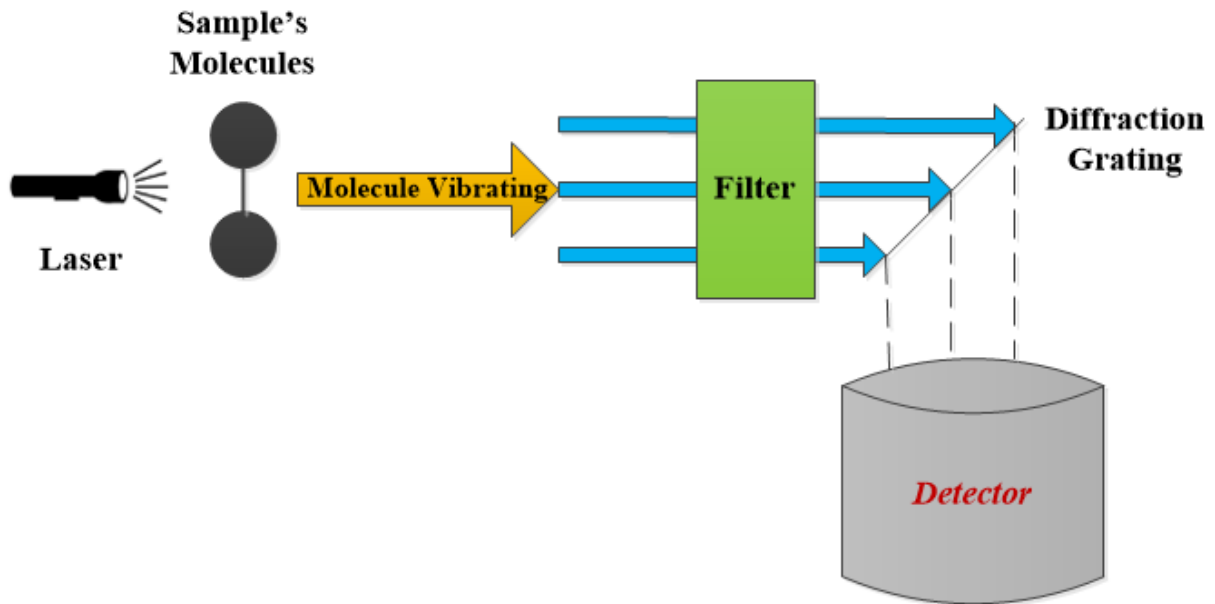
Figure 1.1: Raman Spectroscopy

waves that can be interpreted as a sound signal. This principle is called the photothermal or photoacoustic effect [43]. PAI benefits from the advantages of both optical and ultrasound imaging, as it combines high contrast of light absorption with the high resolution of ultrasound imaging. In addition, PAI has a high SNR. The drawbacks of PAI are the limit on the imaging depth and the relatively long time needed for data collection [43].

In [44], photoacoustic spectroscopy aqueous glucose monitoring was used. The glucose concentration in aqueous solution was estimated by the change in the optical absorption coefficient and the change in pressure. The system used a 1550-nm laser light and a microphone in a photoacoustic cell to acquire the photoacoustic signal samples. Different glucose concentrations were used to collect photoacoustic signals, and these signals showed a linear proportionality with the glucose concentrations. There is no dependency between the photoacoustic signal acquisition process and the light scattering in tissues. The drawback of the system is the high cost in terms of energy and time and the sensitivity of the photoacoustic signal acquisition process to pressure and temperature variations.
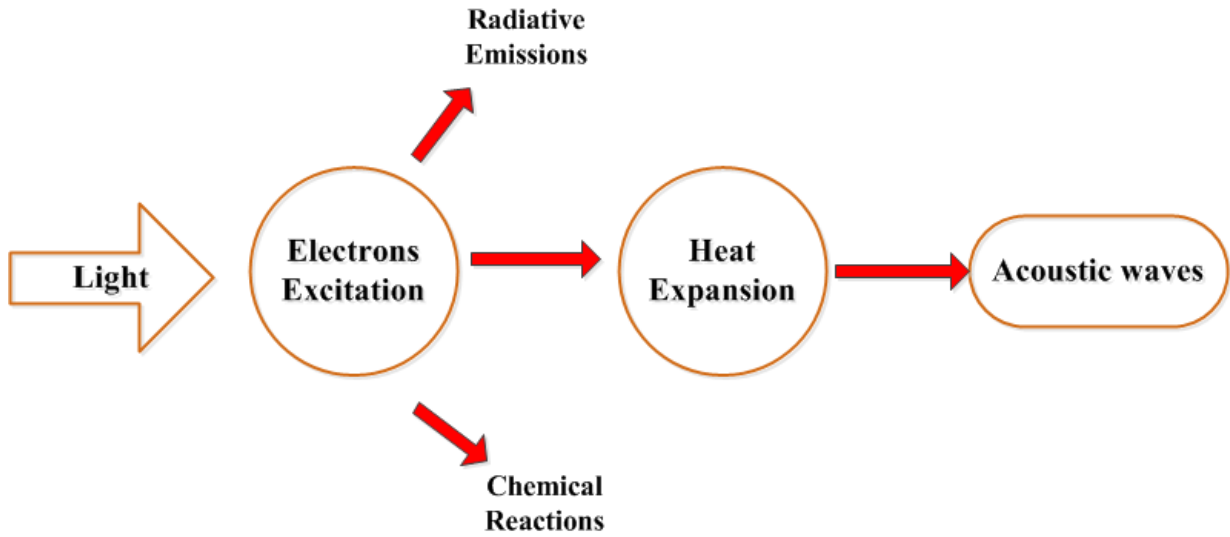
Figure 1.2: Simple diagram for acoustic waves

1.3.3 Optical Coherence Tomography (OCT):

Optical Coherence Tomography (OCT) is an optical signal acquisition system consisting of a low coherence light, an interferometer, and a photo detector to measure the interferometric signal [34]. The interferometer consists of a sample arm and a reference arm with moving mirror. The interferometric signal is the result from the combination of light backscattered from tissues and light returned from the reference arm of the interferometer. The signal is detected by the photo detector [38].

OCT is used to measure the glucose concentration by the correlation of the delay between the backscattered light in the sample arm and the reflected light in the reference arm. A time-domain OCT system was used in [45]. The system images the diffusion of glucose in the skin of anesthetized monkeys. After calculating the permeability rate, the results suggest that OCT might be utilized for non-invasive monitoring of blood glucose in the multilayered biological tissues in vivo.

In another study [23], glucose concentration is measured by an OCT signal amplitude using tissue phantom and human blood samples. Results showed a cyclic correlation between signal intensity and glucose concentration, however, the research concluded that using OCT in glucose measurement is unfaithful because some parameters may affect the OCT signal intensity such as the size of RBC and the presence of creatinine and other materials in the blood.

7

One recent study used OCT for non-invasive glucose monitoring [46]. In this study, changes induced by glucose in the optical properties of the dermis were measured. The study showed a correlation between the blood glucose concentration in the dermis and the OCT signal slope.

Another recent study explored the changes in OCT signal slope (OCTSS) with the variation of blood glucose concentration (BGC) [47]. The results calculated the correlation coefficient R between OCTSS and BGC in both volunteers with diabetes (R= 0.91) and healthy volunteers (R= 0.78).

The advantages of using OCT for CGM include higher SNR, good penetration, and high resolution. However, there are some limitations to OCT such as sensitivity to the motion of individuals and skin temperature. Moreover, there is no clear comparison of the OCT method with other similar methods for non-invasive CGM [34].

1.3.4   Fluorescence Spectroscopy:

Fluorescence is the emission of absorbed light by a substance. Fluorescent molecules are called fluorophores, which absorb energy of a specific wavelength and reemit energy at a different wavelength (Figure 1.3). A fluorescence glucose-sensing molecule is constructed by changing fluorescence from baseline according to the glucose concentration [22]. Glucose does not fluoresce, therefore, it must be combined with a fluorophore for detection [32].

In [48], a glucose sensor is proposed using the fluorescence labeled i-motif DNA as a probe. The pH value of the system is decreased by the enzymatically generated gluconic acid. The proposed method can detect glucose effectively in human serum and urine. The advantages of using fluorescence-based CGM are higher sensitivity to low glucose concentration, less need for calibration, and no need for implanted transmitter or power source for wireless transfer of glucose data [22]. On the other hand, there are some drawbacks of using fluorescence-based CGM such as: the scattering phenomena, the short life times and biocompatibility of the devices [34].

1.4   Non-invasive Microwave-based Glucose Monitoring:

Microwaves are electromagnetic waves with frequencies between 100's MHz and 100's GHz. The antenna, which is the source of radiation, is affected by the electrical property in its environment. This affect manifest itself by a change in the antenna's input impedance

8
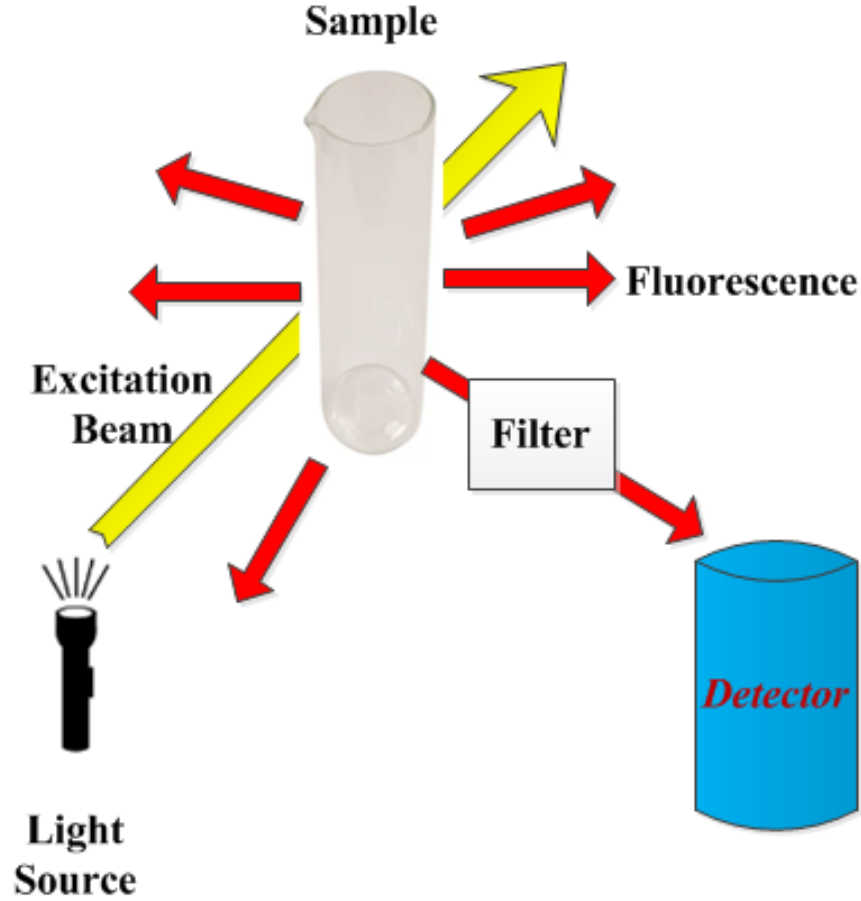
Figure 1.3: Illustration of fluorescence spectroscopy

(which is related to the antenna's resonance). Therefore, studying the change in the electrical properties of blood as a function of glucose concentration is the main idea behind most of the glucose detection modalities that employ microwaves, or microwave spectroscopy in general [49].

Microwave spectroscopy, has the advantage that the molecular structure of the biological samples remains unaltered because microwave sensors (antennas) do not produce ionizing radiation [50]. It is worthy mentioning that microwave is not the only waves that produce non-ionizing radiation.

There are some factors affecting the design of a microwave-based CGM technique. These factors include: the frequency of the sensor, the sensor type and the type of the

signal to be analyzed in order to calculate the glucose level.

Table 1.1 summarizes the different research using microwaves, while highlighting the primary factors in designing any microwave-based CGM technique.

One of the recent techniques proposes adaptability to different measurement conditions through the learning process by applying techniques of artificial intelligence and machine learning [51].

GlucoWise is a device designed using radio wave spectroscopy for CGM. The target area of the device is the skin between the thumb and forefinger or the earlobe. The device is in development and will be available once clinical trials are completed.

1.5   Non-invasive Thermal-based Glucose Monitoring:

Non-Invasive thermal-based CGM includes techniques like thermal emission [52], Metabolic Heat Confirmation (MHC) [42]. Thermal emission spectroscopy utilize the energy emitted naturally from the human body as a heat in the band between 8 μm and 14 μm (far-IR band). Glucose molecules absorb this radiation at 9.4 μm wavelength, so it provides a measure for glucose concentration. However, factors like temperature, motion and thickness of the tissues affects the thermal emission technique.

In MHC, the amount of heat produced by the metabolic oxidation of glucose correlates with the amount of glucose and oxygen levels in the body. It is relatively easy to measure physiological parameters in the body such as the blood oxygen saturation, blood flow velocity and heart rate, however, MHC technique is highly affected by environmental parameters like temperature and sweat.

The method used in [25] combines the conservation of energy method with a sensor integration module to collect physiological parameters, such as the blood oxygen saturation, blood flow velocity and heart rate. After collecting the sample experimentally, a classification model based on decision tree and back propagation neural network was used to classify the glucose level into three categories reaching an accuracy as high as 88%.

Most of the thermal-based research I found in the literature was done before 2009, so they are out of our scope.

Table 1.1: Summary of microwave non-invasive CGM techniques

Reference	Frequency Used	Sensor Type	Signal Type
[49]	500 MHz-20 GHz	Microstrip Patch	relative permittivity and conductivity
[53]	4-5 GHz	Probe-tip	S11
[54]	0.5-2 GHz	Circular Spiral	S21
[55]	23MHz-12GHz	Microstrip	Impulse response
[56]	1-10 GHz	Rectangle antenna	Specific Absorption Rate (SAR)
[57]	0-4 GHz	Microstrip	S11 and S21
[58]	0-40 GHz	Reflectometer	Reflection- and transmission
[59]	0-3 GHz	Spiral	S11 and S21
[60]	500 MHz-8 GHz	Planar dipole	Input Impedance
[61]	0.3-20 GHz	Patch resonator	Input Impedance, S parameters
[62]	1-3 GHz	Ring resonator, single and double spiral resonator	S21
[63]	1.3-1.7 GHz	Split-ring resonator	S-parameters
[64]	300MHz-2GHz	Spiral microstrip resonator	S11
[65]	3 GHz	Double split-ring resonator	S21
[66]	1.4-1.8 GHz	Two spatially separated split-ring resonators	S21
[13]	0.2-50 GHz	Open-ended coaxial probe	Dielectric constant and loss factor
[67]	1-3 GHz	A microstrip antenna	S11
[68]	0-4 GHz	Square-shaped spiral inductor with rounded corners	S-parameters
[69]	1.4-2.2 GHz	Split ring resonator	S21
[70]	3-7 GHz	Four-arm spiral microstrip	S11
[71]	50-65 GHz	Two facing microstrip patch antennas	S-parameters
[72]	0-6 GHz	Planar Ring Resonator	S21
[73]	3.2-12 GHz	A pair of small UWB microstrip patch antennas	UWB pulse signal
[74]	1.4 GHz	Split-ring resonator	Microwave data and temperature data
[75]	5.41 GHz	Complementary split-ring resonator (CSRR)	S21
[76]	1-3 GHz	ENG unit-cell resonator (in the shape of ring and horn)	S11 and S21
[77]	2.4-2.5 GHz	4-cells complementary split ring resonators (CSRRs) hexagonal	S21
[78]	1-6 GHz	Triple-Pole complementary split ring resonators (CSRRs)	S21

1.6 Non-invasive Transdermal-based Glucose Monitoring:

Impedance spectroscopy, Reverse Iontophoresis (RI) and ultrasound techniques can be classified as transdermal-based techniques [42].

Cadduff's group published the first non-optical non-invasive CGM system using impedance spectroscopy [79]. Pendra (Pendragon Medical) was a non-invasive device developed in 2000 using impedance spectroscopy to monitor glucose in the blood continuously. However, the device is no longer available. Frequent calibrations and poor accuracy were two of the major reasons for the failure of the device despite its CE approved in May 2003 [80].

RI is based on the flow of a low electrical current through the skin, between an anode and a cathode positioned on the skin surface [34].

In [81], the transdermal extraction of glucose and potassium using RI was investigate. Two ethanol enhanced skin gel electrodes were developed and used with a small current to extract those molecules from nine healthy volunteers for two separate 60 min periods. The extraction methodology was sensitive to relatively small changes in analyte levels within the blood. The results show a correlation between blood and transdermal glucose measurements. However, there is a need for an alternative calibration approaches for good accuracy needed in clinical application.

In RI, the electrodes are easily applied to the skin. However, there are some drawbacks such as the effect of electrodes on the skin, and the need for the electrodes to be placed for at least one hour before the measurement, which introduces a significant inconvenience. Additionally, the measurement suffers from accuracy especially when there sweat is present. Because of these limitations, GlucoWatch (Cygnus Inc., USA) was withdrawn from the market. SugarBEAT (Nemaura Medical, UK) [82] is expected to overcome these challenges. SugarBEAT is in the first phase of a commercial launch and was expected to be available commercially in 2020, however, it did not.

In order to design a calibration-free non-invasive CGM device using ISF, a path-selective, non-invasive, transdermal CGM system based on a miniaturized pixel array platform was developed [26]. The system is based on the fact that hair follicles have low-resistance and preferential pathways for most of the electroosmotic flow during iontophoresis. The system removed the dependency on inter- and intra- individual fluctuations in skin characteristics on the measurement of glucose.

Ultrasound technology is based on low-frequency sensors which can penetrate the body. The advantages of ultrasound are the good skin and tissues penetration, and the immunity

to skin color variation. However, ultrasound are sensitive to temperature. Recent studies used modulated ultrasound with other optical techniques such as IR [83, 84, 85], therefore, these researches will be explained under the next hybrid section.

1.7   Non-invasive Hybrid-based Glucose Monitoring:

Based on the survey above, not a single technique has succeeded to accomplish non-invasive CGM with an acceptable level of accuracy and reliability. Therefore, researchers developed hybrid techniques to harvest the benefit of multiple techniques while at the same time minimizing their drawbacks.

In general, combining various techniques frees the device from being restricted by the limitations of one technique, on the other hand, it increases the complexity and cost of the proposed solution. However, hybrid non-invasive CGM may become promising strategy if it overcomes the accuracy and selectivity challenge.

In [83], a modulated ultrasound is used with infrared (940-nm wavelength) to monitor blood glucose on human subjects. The performance of the system is based on two tests: oral glucose tolerance test and random blood glucose level test. The results obtained from this technique is paired and compared with readings from invasive method. The mean absolute error for both tests were 15.92 mg/dl and 17.76 mg/dl, and MARD was 0.11 and 0.10 respectively. The advantages of the system include safety and tolerance. Results suggest that the modulated-ultrasound with optical is a promising technique for non-invasive CGM because it utilizes the best features of every technique and avoids the drawbacks. For example, 40 kHz central frequency based ultrasonic transmitters has a very good tissue penetration, while the absorption spectrum of molecules can be distinguished in the range of 900-1000 nm.

Since factors such as temperature, sweat, blood perfusion and body movements affect the contact between the sensor and the skin, a multi-sensor system combining dielectric and optical characterisation of skin was proposed in order to reduce the perturbation of the reading of BGC [86]. In this method, results were derived from only 10 patients with Type 1 diabetes, where a linear regression was used to model the system. Their results were derived from only 10 patients with Type 1 diabetes, where a linear regression was used to model the system. 10 samples only are not enough for generalization of any system, so, they need more samples.

Another study developed a microwaves-based sensor and microfluidics for glucose concentration quantification in aqueous samples [87]. However, such a system needed to be verified using real clinical trials to check its accuracy and reliability.

13

Recently, a non-invasive lab on-a-chip microwave biosensor system for CGM was designed using complementary split-ring resonator (CSRR) sensor integrated with the microfluidic channel [75]. The results indicate a promising low profile, cost-effective and compact system. However, the system uses glucose-in-water solutions.

A multi-sensor non-invasive CGM system using IR sensor and ultrasonic micro-electro-micro mechanical (MEMS) technology is proposed in [28]. According to the simulation results, the proposed system has potential to enhance accuracy, easiness and comfort for better assistance to monitor and manage diabetes. Only multiple Matlab regression models were applied to determine the glucose levels. In addition, there was no explanation about the nature of the samples used (i.e., in-vivo or ex-vivo, and type of diabetes). The system has yet to be tested using real-world clinical data.

Another hybrid system [88] is proposed using impedance, optical multi-sensors and time series analysis for CGM by using physiological parameter changes. Results from three volunteers with diabetes and six healthy volunteers showed that the average correlation coefficient between the estimated and reference glucose profiles was 0.8314, with a normalized root mean squared error (NRMSE) of 14.6064. For the system to be commercially viable, more clinical trials are needed to eliminate any restrictions and for enhanced reliability.

GlucoTrack is a non-invasive CGM device, which can be considered as a hybrid device as it combines thermal, ultrasonic and electromagnetic techniques in one device. The algorithm calculates the weighted average of the three measurements and returns the glucose level. The device is intended for people with T2D and prediabetes [27, 89, 90]. More evaluations for GlucoTrack performance and user acceptance are needed.

1.8   Other Non-Invasive Glucose Monitoring

1.8.1   ECG

ECG is the process of recording the electrical activity of the heart. ECG can be used to monitor the level of hypoglycemia in people with T1D [29, 30]. In these studies, a non-invasive CGM using the physiological parameters of ECG signal along with extreme learning machine (ELM) and Neural Networks (NN) is proposed to recognize the presence of hypoglycemia.

1.9 The Sensitivity and Selectivity

The sensitivity and selectivity of glucose are major factors in every non-invasive CGM technique. Table 1.2 highlights those factors to take into consideration when choosing the appropriate technique.

Table 1.2: Sensitivity and Selectivity of non-invasive CGM techniques [1, 2]

Technique	Sensitivity	Selectivity
NIR	High	Good
MIR	High	Good, better than NIR
Raman	Low	Excellent
PAI	High	Good
OCT	High	Poor
Fluorescence	High	Excellent
Microwave	High	Poor
Thermal	Low	Good
Transdermal	Low	Good

1.10 Body Fluids Considerations

As I have mentioned earlier, non-invasive CGM can use different body fluids such as blood, ISF, tears, saliva, sweat and even breath. There are some considerations need to be addressed in order to have a clear picture of the relation between glucose concentrations and the fluids.

The blood has the highest glucose concentration comparing to all other fluids, then comes ISF followed by the remaining fluids. There is a time lag in measuring glucose concentrations from all the fluids other than the blood. It varies between 5-45 minutes. Saliva can be easily collected, however, it has many interfering impurities and a low correlation and sensitivity to glucose. Tears have less interfering impurities, however, they are of low comfort, low volume, and their pH level is not stable but varies between 6.5 and 7.6. Sweat can be easily collected too, however, it is not suitable for people with diabetes, and its pH level varies between 4.5 and 7. ISF has a high glucose concentration and a high sensitivity, but usually needs micro-needles which may cause skin irritation [32, 91].

Finally, I summarized the advantages and disadvantages of all the techniques for non-invasive CGM in Table 1.3.

Table 1.3: Advantages and Disadvantages of non-invasive CGM techniques

Technique	Advantages	Disadvantages
NIR	1. Harmless to the body. 2. Flexible 3. Low cost 4. Water absorbs less NIR.	1. Affected by Body temperature, Skin thickness, Blood pressure and Fatty tissues. 2. High scattering
MIR	1. Increased absorption (higher wavelength). 2. Decreased scattering phenomena.	1. Tissue penetration: a few micrometers. 2. Water absorbs more MIR.
Raman	1. Providing sharper and less overlapped spectra. 2. Less sensitivity to temperature and water.	1. Laser wavelength and intensity are instable. 2. Spectral acquisition times are too long. 3. Low SNR. 4. Not tested in Human.
PAI	1. High contrast and resolution. 2. High SNR.	1. Sensitive to fluctuation of temperature. 2. Data collection takes too much time. 3. High cost.
OCT	1. High SNR and resolution. 2. Good Penetration. 3. High resolution.	1. Very sensitive to motion. 2. Affected by skin temperature.
Fluorescence	1. High sensitivity. 2. Less need for calibration.	1. Scattering phenomena. 2. Short life times. 3. Not Bio-compatible.
Microwave	1. Harmless to the body. 2. Non-ionizing. 3. Can be easily miniaturized. 4. In-expensive	1. Poor glucose selectivity.
Thermal	1. Good selectivity. 2. Easy to measure.	1. Affected by temperature, motion and thickness of the tissues.
Transdermal	1. Low frequency: good penetration	1. Needs frequent calibrations. 2. Poor accuracy. 3. Sensitive to temperature

1.11 Problem Statement

In order to overcome the shortcomings of the invasive method of CGM, such as pain, discomfort, and risk of infection, non-invasive CGM is needed. However, due to the multiple challenges such as accuracy, usability and applicability, contemporary non-invasive glucose monitors are still not sufficiently reliable. This research is to develop a non-invasive CGM system using microwave sensor with machine learning techniques to improve the accuracy, usability and applicability of non-invasive techniques in predicting glucose levels.

1.12 Objectives and Contributions of the research

The main objective of this research is to develop a non-invasive CGM system using microwave sensor with machine learning techniques. The system will make it easy to regularly check the blood glucose levels and adjust the medication as needed. The system should take into account the following considerations:

- A non-invasive, and low-cost to enable over-the-counter availability.

- The device has to be small and easy to use by a wide population especially children.

- The results have to be accurate and robust.

- The system has to be comfortable to use and does not interfere with daily human functions.

The main contributions of this research are as follows:

- Understanding the problem of non-invasive CGM and its challenges.

- Building the microwave sensors and taking into account factors such as safety to the human body, penetration depth and high sensitivity.

- Starting with aqueous solutions (i.e., water-glucose solutions) as a proof of concept and to check the ability of the sensors to detect the different concentrations of these simple water- glucose solutions.

- Designing a hand model system with different tissues/layers and simulating the effects of the microwave sensor with respect to changing in dielectric properties of those tissues/layers.

- Collecting the experimental and simulation data and analyzing them using machine learning techniques, starting from feature engineering techniques and ending with predicting the glucose levels and the dielectric properties values.

- Calculating the accuracy of the system when predicting glucose levels and dielectric properties using different tissues/layers.

- Finalizing the system setup to be suited for collecting data from humans.

1.13   Research Methodology

In this section, I will show the methodology of the research as shown in Figure 1.4.
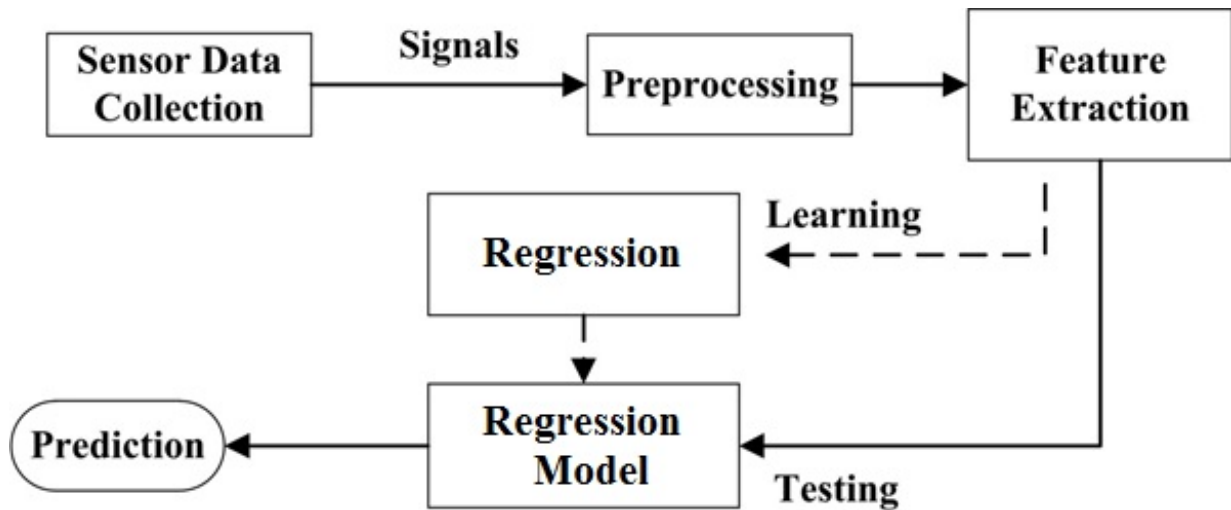
Figure 1.4: The Methodology of The Research

The details of the methodology will be shown in the next chapters. The main idea behind the research is based on measuring the glucose levels based on the dielectric properties of the tissues using the reflected signals (S11) coming from the interaction of the sensors with the tissues, as shown in Figure 1.5.
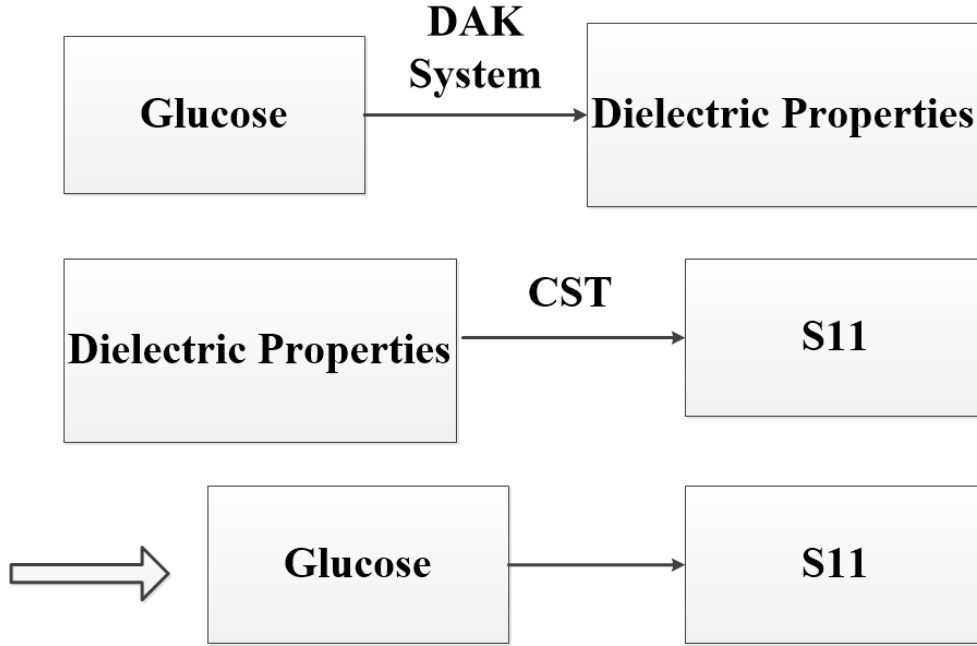
Figure 1.5: The research Idea

1.14 Dielectric Properties of materials

As we know that electrically charged protons and electrons, and non-charged neutrons form the building blocks of any matter. Therefore, any material or substance has electrical properties based on the arrangement of these building blocks.

Understanding how a given material reacts under specific circumstances will helps knowing the properties of the material. Dielectric properties describe the interactions of electric fields with material and defined by Maxwell equations. Permittivity and conductivity are two types of dielectric properties of materials.

Permittivity (ε) [92], is a measure of proportionality that relates the electric field in a material to the electric displacement in that material. It is a measure of electric polarizability of a dielectric, i.e. a material with high permittivity will have more polarization (larger tendency for charge distortion) than a material with low permittivity, thereby more energy will be stored in the material. Permittivity is measured in Farad per meter (F/m) unit and it is a frequency-dependant. The permittivity of free space (vacuum) is denoted by (ε_0) and is equal 8.85x10^{12} F/m. Relative permittivity (ε_0) of a material is a the ratio

19

of material's permittivity to the permittivity of free space

$$\varepsilon_r = \varepsilon/\varepsilon_0 \tag{1.1}$$

The dielectric Properties of the materials can be expressed as complex relative permittivity $(\hat{\varepsilon})$ [93]:

$$\hat{\varepsilon} = \varepsilon^{'} - j\varepsilon^{''} \tag{1.2}$$

where the real part $(\varepsilon^{'})$ is the relative permittivity and is called dielectric constant. The imaginary part $(\varepsilon^{''})$ is the energy loss of the electric field when passing through a material and can be expressed as:

$$\varepsilon^{''} = \sigma/\varepsilon_0\varpi \tag{1.3}$$

where σ is the conductivity, which is the reciprocal of the resistivity. ϖ is the angular frequency of the field. The unit for the conductivity is Siemens per meter (Sm^{-1}) assuming that ε_0 is in (F/m), and ϖ in in radians per second. To conclude, in order to obtain dielectric properties of a material, we can calculate $\varepsilon^{'}$ and $\varepsilon^{''}$ or $\varepsilon^{'}$ and σ values as a function of frequency. Dielectric Assessment Kit (DAK) system is used for dielectric properties measurements over specific frequency ranges.

1.15 Thesis outline

The thesis consists of five chapters. The first chapter contains introduction and motivations of the research. The remaining chapters are organized as follows:

Chapter Two discusses the process of designing the microwave sensors which will be used to collect the glucose data. It represents the hardware part of the system.

Chapter Three explains the machine learning techniques and how to apply those techniques to predict the glucose levels. It shows the performance metrics in order to evaluate the accuracy o the system. It represents the software part of the system.

Chapter Four shows the experimental and simulation results using the microwave sensors with machine learning techniques. It gives the practical details of the steps of the methodology of the system.

Finally, **Chapter Five** concludes the achieved work in the thesis and shows a plan for future work.

Chapter 2

Microwave Sensors Design

2.1 Introduction

In general, microwave sensors (MWS) utilize electromagnetic fields and operate at frequencies in microwave regime starting from 100's MHz up to 100's GHz. There are many types of microwave sensors including planar MWS, resonator MWS, non-resonators MWS etc. Resonator MWS are constructed to have a resonator frequency or oscillation phase based on the measured parameters [94]. Designing of resonator sensors is based on choosing a suitable operating or resonance frequency which is based on the parameters of emission or on electrical length parameters of the proposed sensor. Selecting the operating or the resonance frequency always based on the type of application that need to apply the microwave sensors on it.

The microwave sensor emits electromagnetic waves that significantly formed by the unique vibrations take place between the magnetic field and the electric field. The resonance frequency occurs in electromagnetic waves (EM) as a result of typical oscillating between both electric and magnetic fields. Both the fields are typically perpendicular to each other and with the wave direction too. The microwave sensor emits electromagnetic waves that have their particular frequency and wavelength.

Resonator sensors are of two kinds: passive resonator sensors and active resonator sensors. The sensing configuration of the passive resonator sensors is based on naturally emitted or reflected microwave radiation from the object or material under test (MUT), whereas the sensing configuration of the active resonator sensors is based on the sensor itself where the sensor emits microwave radiation and then senses reflected microwaves from the object or material under test (MUT) [95, 96].

Recently, active resonator sensors have played a fundamental key role in many applications, such as materials characterization, and biomedical applications. There is a growing interest in developing active resonator sensors that lead to high sensitivity and accuracy of measured parameters of material under test (MUT). The active resonator sensors are used to sense the dielectric properties of any materials such as solid, powder and aqueous solutions based on the complex permittivity extraction of the materials [97].

In this chapter, I will discuss the process of designing the microwave sensors which will be used to collect the glucose data in order to analyze and predict the glucose levels. To design a microwave sensor, you need to consider many factors carefully, depending on the type of application. Some of those factors include: the operating or resonance frequency of the sensor, the penetration depth and the size of the sensor.

The size of the active resonator sensors is based on the wavelength of the operating frequency, where the relationship between the wavelength and the operating frequency is based on the following equation:

$$\lambda = \frac{c}{f} \tag{2.1}$$

, where λ is the wavelength, f is the operating or resonance frequency and c is the speed of light $(3 * 10^8 m/s^2)$. As described in the above equation, the wavelength and operating frequency have an inverse relationship; when the operating frequency increases, the wavelength decreases which also decreases the electrical length of the sensor. In addition, the measurement location will place a size constraint on the sensor. Since this sensor may one day be used in an off the-shelf glucometer, choosing the operating or resonance frequency as a low frequency (in terms of microwave frequencies range) is recommended. Penetration depth is a major parameter in microwave sensor design. If the fringing fields of the sensor do not extend deep enough into the tissue before returning to the ground plane, the sensor is useless. Since lower frequencies have longer wavelengths than higher ones, lower frequencies penetrate deeper into tissues. Thus, the selected frequency range also benefits the penetration depth of the sensor. In addition, the relationship between the penetration of the active resonator sensors and the operating frequency of the proposed sensor is based on the following equation:

$$depth = \delta = \sqrt{\frac{\rho}{\pi f \mu}} \tag{2.2}$$

, where ρ is the resistivity of the material under test (ohm-meter), μ is the permeability of the material, and f is the operating frequency. We noticed from the above equation 2.2 that the penetration and the operating frequency are inversely related. The sensitivity and accuracy of the proposed sensor is depending on the penetration of the electromagnetic field or radiation inside the material under test (MUT) [98].

22

2.2 Microwave Sensor Design for CGM

Microwave sensors are critical and fundamental part in designing an intelligent system for CGM. Knowledge of the dielectric properties of human body tissue such as skin, blood or muscle and understanding the interactions between the electromagnetic fields and these tissues is essential for the design of microwave sensors. The dielectric properties of tissues are represented by a complex permittivity where the real part represents the ability of the material to store microwave energy, whereas the imaginary, or the loss factor, indicates the ability of the material to absorb microwave energy [99].

Microwave-based CGM modalities are mainly based on the observations of the variation in the dielectric properties caused by changing in the glucose levels in the tissues. Microwaves, reflected off or scattered from, these tissues are then expected to help in measuring the glucose levels in these tissues. The main keys of designing an effective microwave sensor for non-invasive CGM system are including:

- Fast Measurement

- Accurate Results

- easy to wear and portable

- small sensor size

In this thesis, two microwave sensors were used: spiral and dipole sensors.

2.2.1 Spiral Sensor design and simulation

In this subsection, I will show the design and simulation of the microwave spiral sensor. To improve the efficiency of the proposed system for CGM, a novel microwave spiral sensor is designed. The proposed spiral sensor is an electrically-small enough to confine most of the near-field energy into the human tissue to increase its sensitivity to variation in the dielectric properties of the human tissues, such as blood properties due to changes in the glucose levels. For sufficient electromagnetic power penetration into the human tissues, such as skin tissue, the operating frequency must be chosen carefully to ensure enough penetration level inside these human tissues. Without loss of generality, I select the frequency of operation to be 1000-1500 MHz which ensures sufficient energy penetration in the human hand tissues.

23

Designing the proposed spiral sensor is based on many factors, such as inner and outer radius, spacing between the turns, number of turns, and the angle which is calculated by using the following equations:

$$r = a\omega + b \tag{2.3}$$

Where, r- is the distance from the origin, b- the initial point and a- the growth rate. Each arm of the sensor spiral is linearly proportional to the angle, φ , and is described by the following equation:

$$r = r_0(\varphi) + r_1 \tag{2.4}$$

$$r = r_0(\varphi - \pi) + r_1 \tag{2.5}$$

,where r_1 is the inner radius of the spiral and it is the proportionality constant getting from

$$r_0 = (s + w)/\pi \tag{2.6}$$

,where s is defined as the value of the spacing between each turn N and w is the width of each arm. The width of the arm can be calculated by the following equation:

$$s = (r_2 - r_1)/2N - w \tag{2.7}$$

,where r_2 is the outer radius of the spiral sensor and N is the number of turns.

The sensor was designed as a printed spiral antenna of length and width of 44 mm and trace width w is 0.5 mm hosted on a RO4003 Rogers material with a thickness of 1.52 mm, inner radius r1 is 1.2 mm , outer radius r2 is 3 mm, spacing between turns s is 1 mm, number of turns is 5, and a dielectric substrate of a relative permittivity of ε_r=3.38 as shown in Figure 2.1. The electrical length of the spiral is $\lambda/3$, (where λ is the wavelength in free space) which makes it with high penetration of energy into the human hand tissues, which in turn, makes it highly sensitive to tissue material changes because of glucose. To enable the sensor to be highly sensitive, the sensor must made an efficient radiator which leads to a highly defined ultra-narrow response by using many turns. The sensor was designed using the full-wave numerical simulation tool CST Microwave Studio [100].

The spiral sensor is operated at frequency 1000-1500 MHz, and resonated at 1200 MHz as shown in Figure 2.2 which shows the simulation result of the proposed spiral sensor. The sensor was fabricated and tested and it resonated at 1170 MHz.The experimental results are shown in Figure 2.2. As we can see in Figure 2.2, there is a shift in the resonance frequency due to the fabrication. I attribute the difference due to the non-ideal behavior of the elements and particularly the dispersive nature of all material involved in the fabrication of the sensor (for instance, the inductance of real inductors is frequency dependent whereas the simulated ones have frequency independent inductance).
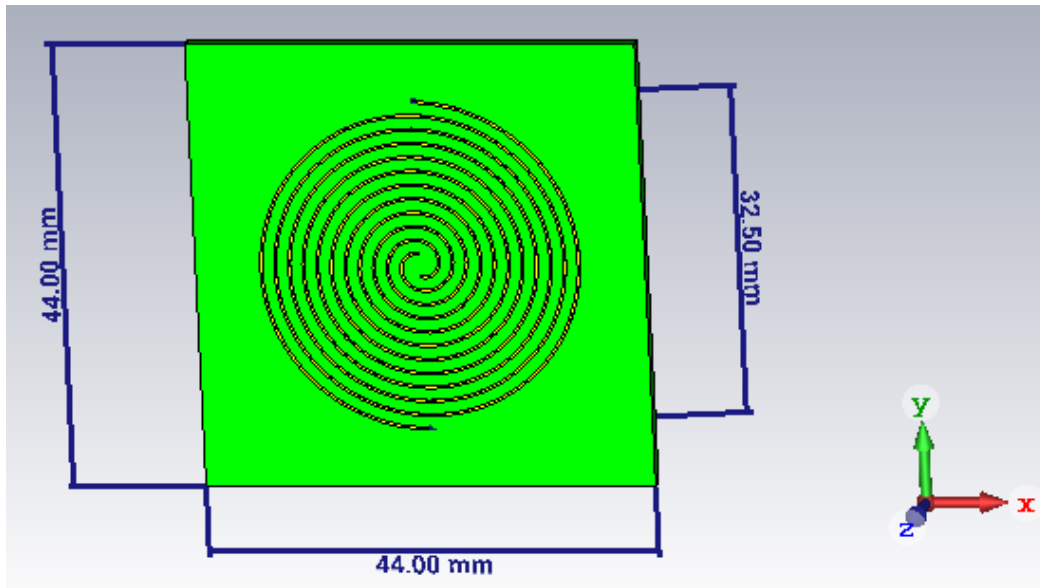
Figure 2.1: The Spiral Sensor
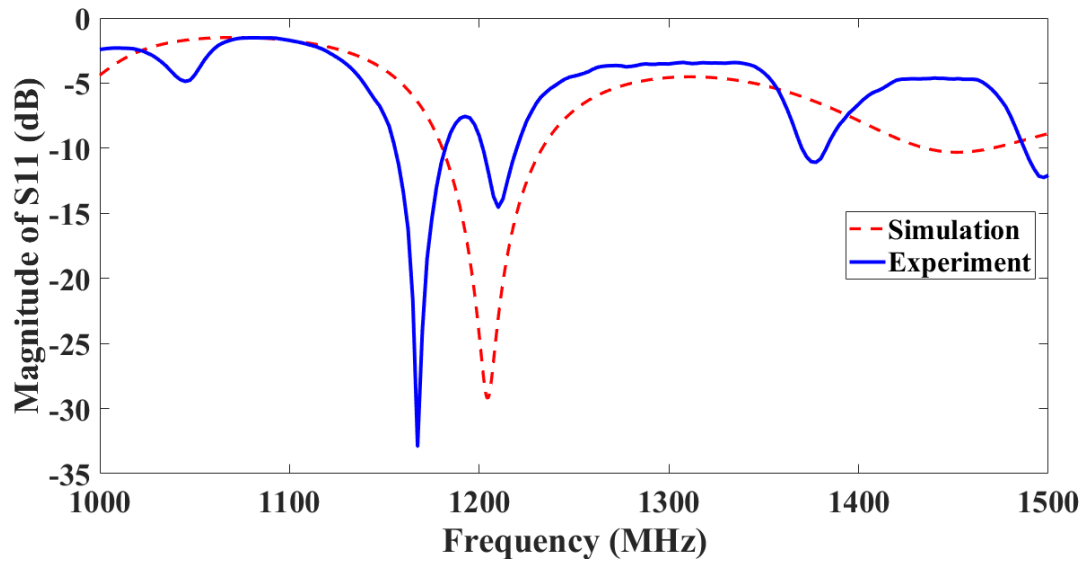
Figure 2.2: Response of the spiral sensor over the 1000-1500 MHz frequency band

2.2.2 Dipole sensor design and simulation

The sensor was designed as a printed dipole of length 92 mm and trace width of 2 mm hosted on a RO4003 Rogers material with a thickness of 1.52 mm and a dielectric substrate of a relative permittivity of ϵ_r=3.38 as shown in Figure 2.3. The electrical length of the dipole is $\lambda/12$, (where λ is the wavelength in free space) which makes its radiation efficiency very low implying a flat, or near unity reflection coefficient (S_{11}) response. To enable the sensor to be highly sensitive, the sensor must be made an efficient radiator which lead to a highly defined ultra-narrow band-stop filter response by using a loss-less matching network. The network was designed using the full-wave numerical simulation tool CST Microwave Studio [100]. The optimized matching network consisted of a series and parallel inductors having inductances of 0.36 μH and 0.49 μH, respectively with specific placement of the elements as shown in Figure 2.3(a) [3]. I emphasize that both the values of the inductors and their location and orientation were optimized using CST.
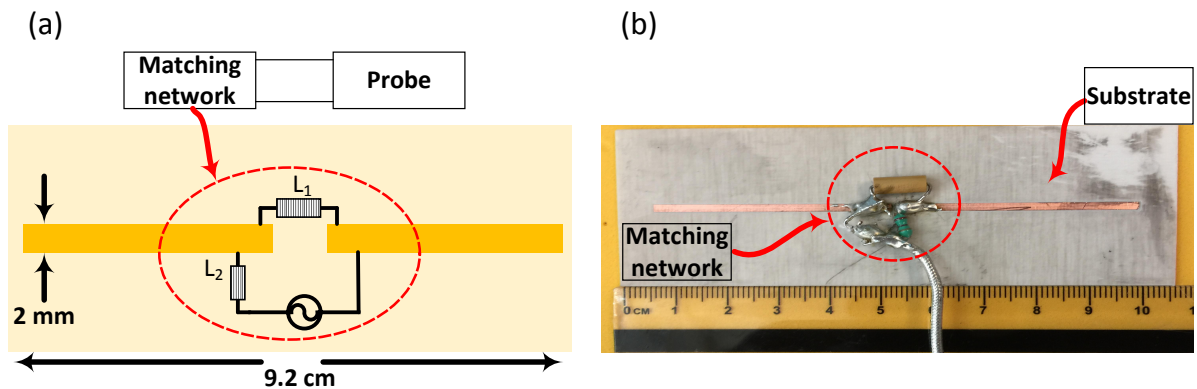
Figure 2.3: The printed electrically-small sensor hosted on a dielectric substrate. (a) Schematic showing sensor and the location of the matching network elements. (b) Photo of the fabricated sensor and the matching network.[3]

The dipole sensor is operated at frequencies 100-300 MHz which makes it an electrically-small sensor and have the ability to penetrate more in the human body. The dipole is resonated at 200 MHz. The sensor was fabricated and tested yielding strong agreement in the resonance frequency between the measurements and the simulations as can be observed from Figure 2.4.

Specifically, the agreement was very strong for the resonance frequency but I observe a deviation in the bandwidth which I attribute to non-ideal behavior of the elements and
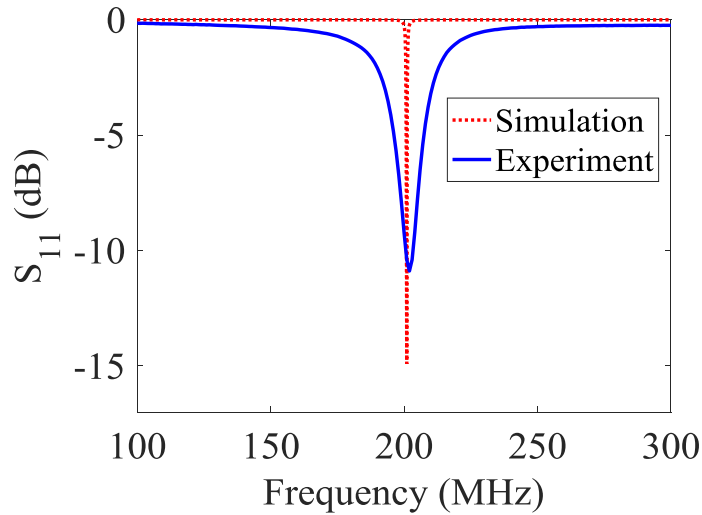
Figure 2.4: Response of the dipole sensor over the 100-300 MHz frequency band

particularly the dispersive nature of all material involved in the fabrication of the sensor (for instance, the inductance of real inductors is frequency dependent whereas the simulated ones have frequency independent inductance).

As both microwave sensors working in the near-filed, there is no need for radiation pattern as it is not relevant.

2.3 Permittivity and S-parameters

When the electric field passing through the dielectric medium its transmission is governed by a phenomenon called permittivity ϵ. The relative permittivity (ε_r) of a material is also called the dielectric constant and is an experimentally measurable parameter. The relative permittivity of a material is a ratio of its permittivity to the permittivity of free space (ε_0) [54].

Scattering parameters or S-parameters (the elements of a scattering matrix or S-matrix) describe the electrical behavior of linear electrical networks when undergoing various steady state stimuli by electrical signals. The parameters are useful for several branches of electrical engineering, including electronics, communication systems design, and especially for microwave engineering. The 2-port S-parameters have the following generic descriptions:

27

- $S11$, is the input port voltage reflection coefficient

- $S12$, is the reverse voltage gain

- $S21$, is the forward voltage gain

- $S22$, is the output port voltage reflection coefficient.

Figure 2.5 shows a schematic description of the s parameter of a two-ports network. The relationship between the reflected, incident power waves and the S-parameter matrix is given by:

$$\begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix} \cdot \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \tag{2.8}$$
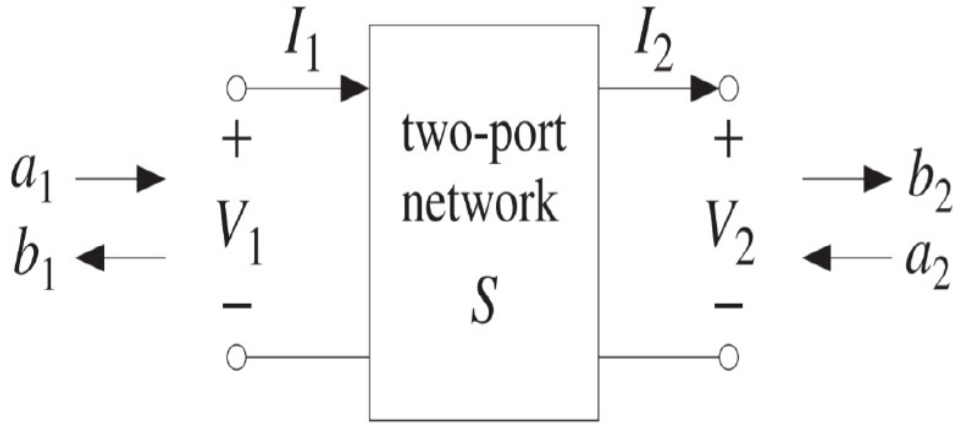
Figure 2.5: A schematic description of the S parameters of a two-ports network

2.4 Conclusion

In this chapter, I discussed the process of designing the spiral and the dipole microwave sensor, which is the hardware component of the non-invasive CGM system. The design process is governed by different factors such as the frequency of the sensor, the penetration depth and the size of the sensor. The frequency of the spiral sensor is 1000-1500 MHz and

it resonates at about 1200 MHz. Whereas, The frequency of the dipole sensor is 100-300 MHz and it resonates at about 200 MHz. I showed the response of both microwave sensor over their range of frequencies in both experiment and simulation, and I found a strong agreement between the experiment and simulation with some shifts due to fabrication. S-parameters are used to describe the interactions between the microwave sensors and the material under test (MUT).

Chapter 3

Machine Learning for Glucose Monitoring

3.1 Introduction

In this chapter, first I will give a general information about machine learning and its different types along with some of the very important algorithms and techniques. then, I will show how we can apply machine learning for CGM application and how we can calculate the performance metrics. More emphasis will be on the machine learning algorithms that will be used for CGM.

3.2 Machine Learning definition and Algorithms

Machine Learning (M.L.) is defined as a set of methods that can first automatically detect patterns in data, and then using those detected patterns in building a model that can classify or predict future data, or to perform other kinds of decision making under some uncertainty [101]. M.L is actually a sub field or type of Artificial Intelligence (A.I.).

There are four types of M.L.: supervised, semi-supervised, unsupervised and reinforcement learning. In **supervised learning**, the machine is taught by examples. Given a set of labeled input-output pairs $D = \{(x_i, y_i)\}_{i=1}^{N}$, the goal is to find a mapping from input (x) to output (y), where D is called the training set, and N is the number of training samples. Each training input x_i is a vector of d-dimension where vector elements are called features, attributes or variables. Input x_i can be an image, a transaction, an email message, a DNA

30

sequence, a graph, etc. Output y_i or response variables can be categorical or nominal variables from some finite set C, or it can be a real-valued scalar. When y_i is categorical, the problem is known as **classification or pattern recognition**, and when y_i is real-valued, the problem is known as **regression** [101].

The second type of M.L. is **semi-supervised**, in which we have some labeled example (generally a small portion) and the remaining are unlabeled. The goal is to try to learn a mapping as in supervised M.L. Constrained clustering is an example of semi-supervised M.L.

The third type of M.L is **unsupervised**, in which inputs are only given $D = \{(x_i)\}_{i=1}^{N}$, and the goal is to mine for patterns in the data. This type is sometimes called knowledge discovery. Clustering is an example of unsupervised learning such as clustering fishes into different species.

The fourth type of M.L is **reinforcement learning** where the machine learns from a series of rewards and punishments and adapts its behaviour based on these series, for example, playing chess.

Now, I will give some details about classification, regression and clustering.

Classification

Classification is one of the types of supervised learning used to assign an unknown data sample to one of the predefined classes based on some features. A classifier is modeled using data with known classes. This process consists of two phases: training phase and testing phase. In the training phase, the system is learning to find a mapping (classifier or model) between the features extracted from the samples in the training set and their classes. In the testing phase, the system uses the learned model and the features extracted from new samples (i.e. testing set) to assign them to classes. The purpose of a training phase is to produce a model, which is able to predict target values of data instances in the testing phase. In the dataset to be tested, only the attributes are known. It is required to produce a classifier that is not only able to separate the different classes by a function induced from available training data, but also generalizes well (works well on unseen instances) [102].

The goal is to find a mapping from inputs (x) to outputs (y), where $y \in \{1, ..., C\}$, where C is the number of classes and it defines the type of classification, either binary $(C = 2)$ or multiple $(C > 2)$. Function approximation is one way to formalize classification problems. We assume $y = f(x)$, where f is some unknown function, the goal is to estimate the function f using a labeled training set, and then to make predictions using $\hat{y} = f(\hat{x})$ (hat symbol denoting estimation). The accuracy of the classifier is measured by the number

of correct predictions on new input data and this is called generalization [101]. Classifiers are of different types, and one of the distinguished characteristic between them is the number of parameters. When the classifier has a fixed number of parameters, it is called **parametric**, whereas, classifiers that has non-fixed number of parameters where their parameters depends on the dimension of the data are called **non-parametric** classifiers.

Regression

Regression is another type of supervised M.L. and is just like classification except the output values are continuous. It allows us to estimate/predict the value of one or more quantities based on the values of other quantities, for example, predicting stock market prices, predicting temperature at specific location, etc. Regression analysis is an important tool for modelling and analyzing data. Here, a curve/line is fitted to the data points, in such a way where the differences between the distances of data points from the curve or line is minimized. There are different types of regression: linear, multiple, logistic, polynomial etc...

Clustering

Clustering is the common type of unsupervised M.L. and it is about grouping data elements into groups or clusters based on their attributes without any given labels or even number of groups (clusters). The first goal is to estimate the number of clusters K, and then estimate to which cluster each point belongs. Here are two common examples of clustering:

- Document clustering: in which we group similar documents. For example news reports can be clustered to politics, sports, fashion, arts, etc.

- Image compression: in which we assigns pixels with similar colors to the same color to reduce image size.

Next, I will highlight the most common and popular machine learning algorithms which will be relevant to my research.

Naïve Bayes Classifier Algorithm

The Naïve Bayes classifier is used for classification and is among supervised learning. It is based on Bayes' theorem so every value is classified independently of any other value. It uses probability to predict a class/category, based on given set of features.

Despite its simplicity, the Naïve Bayes classifier performs surprisingly well and is often used due to the fact it outperforms more sophisticated classification methods.

Support Vector Machine and Support Vector Regression Algorithms

Support Vector Machine (SVM) and SVR algorithms are among the best supervised learning models for solving classification and regression problems. The main idea for classification using SVM is to build a model to have a decision boundary between different classes with maximum margin. SVR for regression is following the same theory of SVM and the objective is to find the points (for the function that approximates mapping from an input domain to real numbers) that are within the decision boundary line [103].

A SVM classifier defines a hyperplane that separates the data into different classes. To find the separating hyperplane there are two possibilities: the classes are linearly separable or nonlinearly separable (as shown in Figure 3.1).
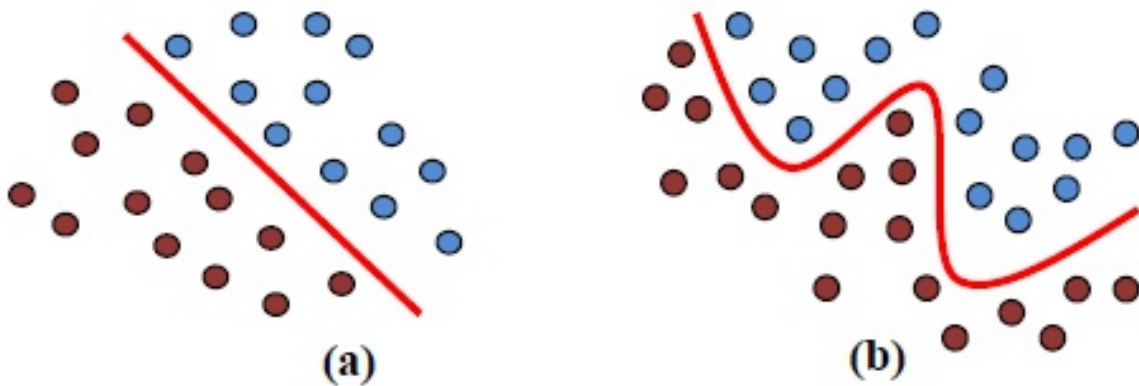
Figure 3.1: The Classes Separability Possibility: a) Linear. b) Non-Linear

Linearly separable case:

There are infinite numbers of hyperplanes that separate the samples of two different linearly separable classes. The optimal hyperplane that enhances the generalization of the classifier is the one with maximum margin, (i.e. maximum distance between the hyperplane and

the closest samples). In Figure 3.2, direction2 hyperplane is greater than direction1. The samples that are closest to the hyperplane are called **support vectors**. In general, the goal of SVM is to find the best orientation of the hyperplane that maximized the margin between support vectors.
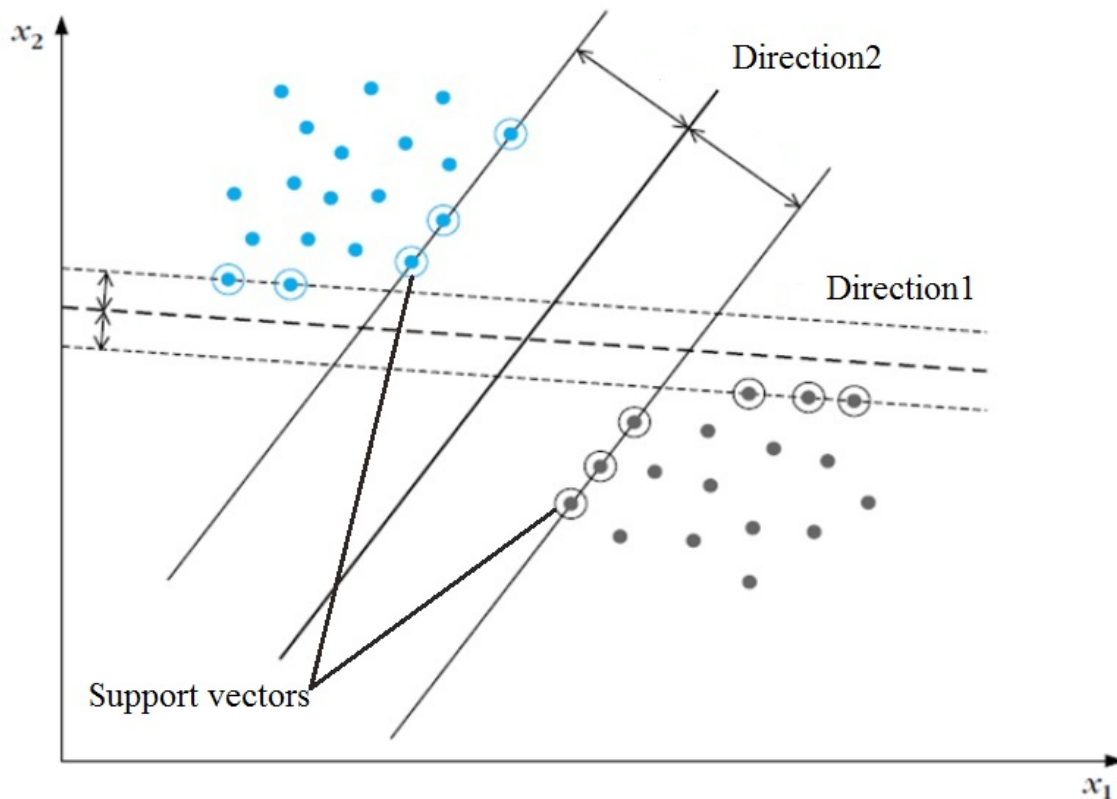
Figure 3.2: An example of a linearly separable two-class problem with two possible linear classifiers

Non-linearly separable case:

In addition to performing linear classification, SVM can efficiently perform a non-linear classification using what is called **the kernel trick**. The kernel functions map the samples to a higher dimension space where the classes become linearly separable (See Figure 3.3). Many kernel functions have been used, but the most popular ones are Radial Basis Function (RBF), Sigmoid and polynomial [104].
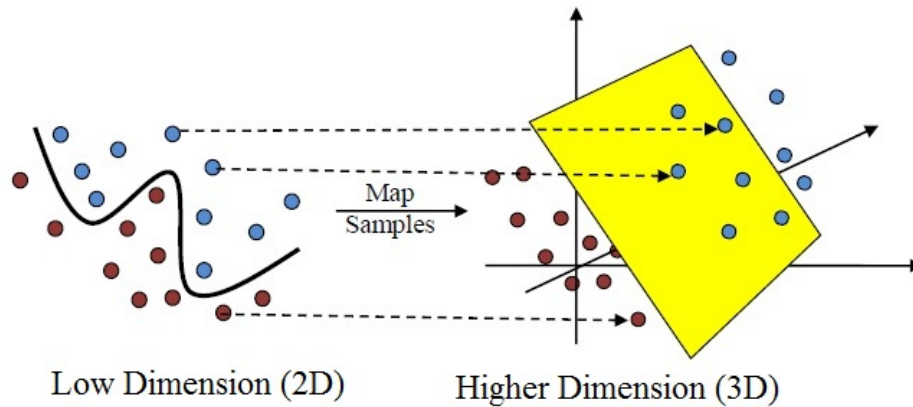
Figure 3.3: SVM kernel for Non-linearly separable classes

Now I will talk about SVR, it uses the same principle as SVM. The main idea behind SVR is to find a function that approximates mapping from an input domain to real numbers on the basis of a training samples. Let's consider the two black lines in Figure 3.4 as the decision boundaries and the red line as the hyperplane. SVR objective is to basically consider the points that are within the decision boundary line, and the best fit line is the hyperplane that has a maximum number of points. The SVR algorithm try to find the best fitted objective function, however, what to do if there are some points still fall outside the margin? (i.e. the error is greater than ε). We can do this with something called slack variables. for any value that falls outside of ε, slack variable is the deviation of that point from the margin. We want to minimize those deviation as much as possible, and we do that by adding this constraint to the objective function which add a new hyperparameter, C, that we can tune. When we increase the value for C, the tolerance for points outside of ε also increases. The best value for C can be obtained using grid search.

Nearest Neighbours Algorithm

The K-Nearest-Neighbour (KNN) algorithm is one of the supervised learning algorithms used for both classification and regression. The input consists of the k closest training instance in the data set (k is a positive integer, typically small). The output depends on whether KNN is used for classification or regression: In classification, it is the class membership which estimates how likely a data point is to be a member of one group or another. Whereas, in regression, it is the property value for the object, which is the average of the values of k nearest neighbors. Assign weights to the contribution of the neighbours is very helpful in this algorithm and the weights can be expressed as the inverse of the

35
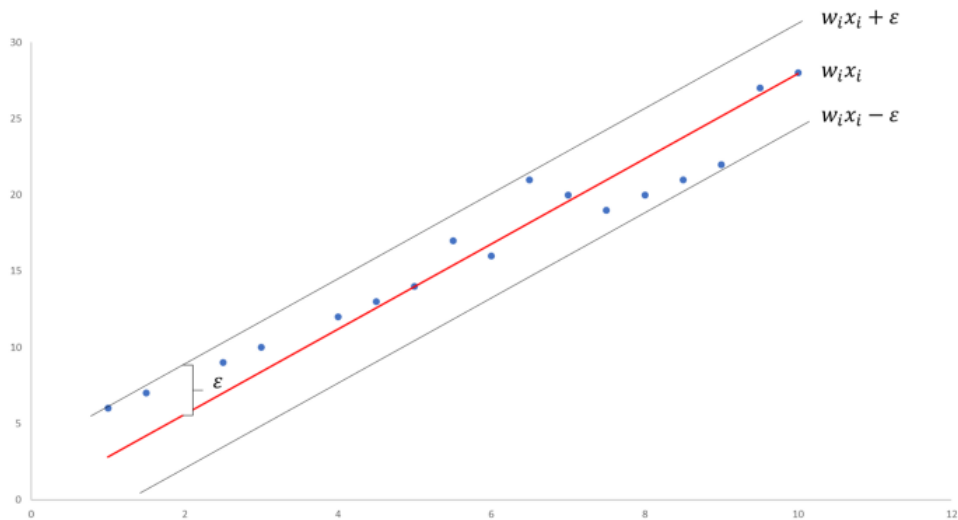
Figure 3.4: Illustrative Example of Simple SVR

distance. Different types of distance measures can be used such as Euclidean, Manhattan, or Minkowski ...etc. The algorithm is very simple and easy to implement, however, it gets significantly slower as the number of examples increase.

Decision Trees

Decision tree algorithm is a supervised learning algorithm used for both classification and regression. It is a hierarchical model in which the local region is identified in a sequence of recursive splits in smaller number of steps. A decision tree can be expressed as a flow-chart-like tree structure that uses a branching method to illustrate every possible outcome of a decision. Each node within the tree represents a test on a specific variable – and each branch is the outcome of that test. Each leaf node has an output, which is the class label (for classification) and is a numeric value (for regression). Decision trees are easy to interpret, in fact they can be expressed as IF-THEN statements. On the other hand, decision tress training is costly and complex in time and effort.

Random Forests

Random forests algorithm is a supervised learning algorithms used for both classification and regression. It is an ensemble learning method, combining multiple algorithms in order

to generate better results than individual algorithm. The algorithm starts with a decision tree and an input at the top. It then travels down the tree, with data being segmented into smaller and smaller sets, based on specific variables. It is a very good way to reduce the variance of decision trees, make them more stable to different inputs. On the other hand, Random forests are harder to interpret than decision trees.

K Means Clustering Algorithm

The K Means Clustering algorithm is a type of unsupervised learning, which is used to categorise unlabelled data, i.e. data without defined categories or groups. The algorithm works by finding groups within the data based on similarity metrics, with the number of groups represented by the variable K. It then works iteratively to assign each data point to one of K groups based on the features provided. It starts with a first group of randomly selected centroids, which are used as the beginning points for every cluster, and then performs iterative (repetitive) calculations to optimize the positions of the centroids. It halts creating and optimizing clusters when either: the centroids have stabilized (no more change), or it reached the defined number of iterations.

3.3   Applying Machine Learning for CGM

Predicting the glucose level can be represented as a regression task. In Machine Learning, the most common types of regressions are: Linear, Multiple, Logistic, Polynomial, Stepwise, Ridge, Lasso and ElasticNet.

Linear Regression

Linear regression is the most basic type of regression. Simple linear regression allows us to understand the relationships between two continuous variables. It establishes the relationship between two variables using a straight line. Linear regression learns a function to predict a continuous variable output of continuous or discrete input variables. It models the relationship between a nonrandom, one-dimensional X that is known, and a random, one-dimensional Y as:

$$Y = \beta_1 X + \beta_0 + \varepsilon \tag{3.1}$$

where β_1 and β_0 are unknown constants and ε is a random variable which may represent measurement error or some other source of [105]. Linear Regression is very sensitive to outliers. It can terribly affect the regression line and eventually the predicted values.

Multiple Regression

Multiple linear regression is when two or more variables have a linear relationship with the dependent variable, the regression is called a multiple linear regression. Many data relationships do not follow a straight line, so statisticians use nonlinear regression instead. The two are similar in that both track a particular response from a set of variables graphically. But nonlinear models are more complicated than linear models because the function is created through a series of assumptions that may stem from trial and error.

Logistic Regression

Logistic regression is used to find the probability of event=Success and event=Failure. It is used when the dependent variable is binary (0/1, True/False, Yes/No) in nature. Logistic Regression can be shown as logistic curve which is a common S-shaped curve (sigmoid curve) as shown in Figure 3.5. Not like normal regression, the parameters are chosen to maximize the likelihood of observing the sample values rather than minimizing the sum of squared errors. It is widely used for classification problems and does not require linear relationship between dependent and independent variables. It applies a non-linear log transformation to the predicted odds ratio, so it can handle various types of relationships. Logistic Regression requires large sample sizes because maximum likelihood estimates are less powerful at low sample sizes than ordinary least square.The independent variables should not be correlated with each other.

Polynomial Regression

A polynomial regression is the one where the power of independent variable in its equation equals more than 1. On the other hand, we should be careful from overfitting which will occur if we try to fit a higher degree polynomial to get lower error. We should always focus on making sure that the curve fits the nature of the problem not every single detail of it.

Stepwise Regression

Stepwise Regression is used when dealing with multiple independent variables. In this regression, the selection of independent variables does not involve any human intervention, it is done with the help of an automatic process. Stepwise regression basically fits the regression model by adding/dropping variables one at a time based on a specified criterion such
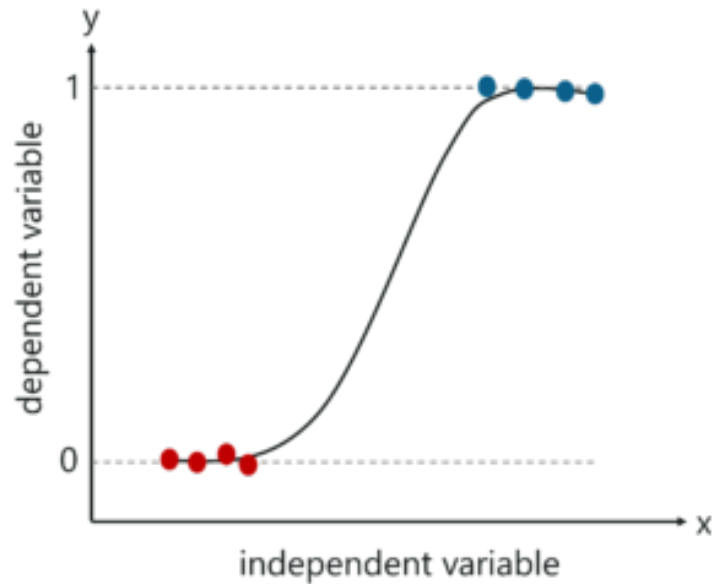
Figure 3.5: Logistic Regression

as statistical values like R-square, t-stats. The following are some of the most commonly used Stepwise regression methods:

- **Standard stepwise regression**: it adds and removes predictors as needed for each step.

- **Forward selection**: it starts with most significant predictor in the model and adds variable for each step.

- **Backward elimination**: it starts with all predictors in the model and removes the least significant variable for each step.

The goal of this regression technique is to maximize the prediction power with minimum number of predictor variables. It is one of the useful methods in case of higher dimensionality of data set.

Ridge Regression

Ridge Regression is used in case of multicollinearity (a situation in which independent variables are highly correlated). In multicollinearity, even though the least squares esti-

mates are unbiased, their variances are large which deviates the observed value far from the true value. Ridge regression reduces the standard errors by adding a degree of bias to the regression estimates. this is done by adding shrinkage parameter λ (lambda). Ridge regression can be expressed as a regularization method and uses $L2$ regularization.

LASSO Regression

LASSO (Least Absolute Shrinkage and Selection Operator) regression is similar to ridge regression except that it uses absolute values in the penalty function, instead of squares. LASSO penalizes the absolute size of the regression coefficients. LASSO is a feature selection technique as it shrinks coefficients to zero (exactly zero), and it is a regularization method and uses $L1$ regularization. If group of predictors are highly correlated, LASSO selects only one of them and shrinks the others to zero.

ElasticNet Regression

ElasticNet Regression is hybrid of Lasso and Ridge Regression techniques. It is trained with $L1$ and $L2$ prior as regularizer. Elastic-net is useful when there are multiple features which are correlated. ElasticNet encourages group effect in case of highly correlated variables, and there are no limitations on the number of selected variables, however, it can suffer with double shrinkage.

3.4   Performance metrics

Performance of the regression is obtained by finding the best fit line/curve of the estimated function. This task can be easily accomplished by Least Square Method. It is the most common method used for fitting a regression line. It calculates the best-fit line for the observed data by minimizing the sum of the squares of the vertical deviations from each data point to the line. Because the deviations are first squared, when added, there is no cancelling out between positive and negative values. We can evaluate the model performance using the metric R-square. However, for CGM, there is a special performance analysis called the Clarke Error Grid Analysis (EGA). Mean Absolute Relative Difference (MARD) is also another metric used for perfornace in CGM

The Clarke Error Grid Analysis (EGA)

The Clarke Error Grid Analysis (EGA) [106, 107] was developed in 1987 to quantify clinical accuracy of estimation of patient blood glucose as compared to the blood glucose value obtained in their glucometer. It was then used to quantify the clinical accuracy of blood glucose estimates generated by glucometers as compared to a reference value. Eventually, the EGA became accepted as one of the "gold standards" for determining the accuracy of blood glucose meters.

The grid breaks down a scatter plot of a reference glucose meter and an evaluated glucose meter into five regions (as shown in Figure 3.6):

- Region A are those values within 20% of the reference sensor,

- Region B contains points that are outside of 20% but would not lead inappropriate treatment,

- Region C are those points leading to unnecessary treatment,

- Region D are those points indicating a potentially dangerous failure to detect hypoglycemia or hyperglycemia, and

- Region E are those points that would confuse treatment of hypoglycemia for hyperglycemia and vice versa.

MARD

The MARD is based on the comparison between paired measurements of a given CGM system (predicted) and a reference method. MARD is computed as mean value of the absolute relative differences (ARD) as follows:

$$ARD_k = 100 \cdot \frac{|Y_{pred}(t_k) - Y_{ref}(t_k)|}{Y_{ref}(t_k)} \tag{3.2}$$

$$MARD = \frac{1}{N_{ref}} \sum_{k=1}^{N_{ref}} ARD_k \tag{3.3}$$

,where Y_{pred} is the value predicted by the CGM device, Y_{ref} is the value measured by the reference measurement device at t_k different times when reference measurements are available, $k = 1, 2, 3, ..., N_{ref}$ [108].

41
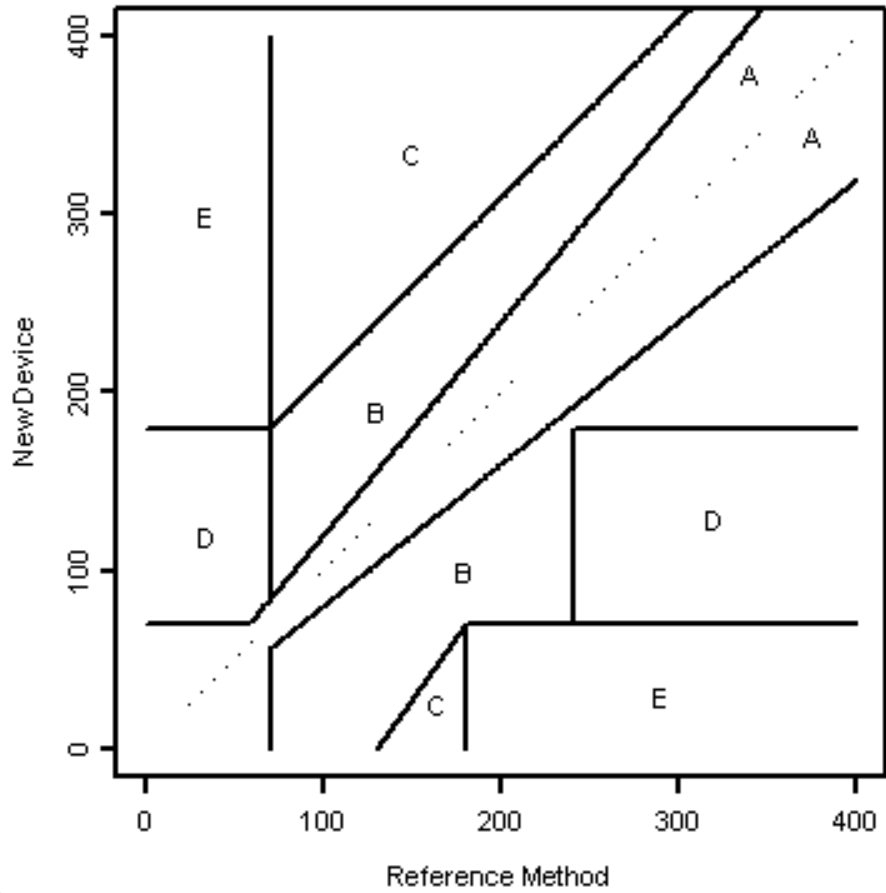
Figure 3.6: The Clarke Error Grid

3.5 Conclusion

In this chapter, I discussed different types of machine learning along with some important algorithms and techniques. I showed how we can apply machine learning for CGM application and how we can calculate the performance metrics. The main part of applying machine learning algorithms for CGM is the regression and I gave more emphasis on the different regression types and algorithms as I am going to use them in the next result's chapter.

Chapter 4

Results and Discussions

4.1   Introduction

In this chapter, I will show the experimental and simulation results using the microwave sensors with both aqueous solutions in experiment and the 3-D hand model in simulation.

4.2   Experimental Results:

In this section, I will show the experimental results using both sensors (the dipole and the spiral) with aqueous solutions with 9 different glucose concentrations prepared in the lab as a proof of concept and in order to check the sensitivity of the sensor to those concentrations. For the purpose of accuracy and validation, the experiments were repeated many times in many different days and the average of the results were conducted. I followed the same methodology (shown previously in chapter 1, Figure 1.4) in all the models of the results.

4.2.1   Preparing the aqueous glucose solutions

Water is suitable for glucose experiments based on the fact that 50% of the volume of the human blood is water, and 92% of the blood plasma is made up of water too [109].

In addition, tissues exposed to radiation in the microwave frequency range exhibit properties strongly dependant on the behavior of water. It is commonly held that percentage of water content is related to dielectric constant [110]. For example, water is the main constituent of blood, whereas fat has very little water.

In this subsection, I will show how to prepare the aqueous glucose solutions. There are two different ways of preparing or measuring glucose concentrations: It can either be in terms of a molar concentration, measured in mmol/L or a mass concentration, measured in mg/dL. A conversion factor of 1 mg/dL equals to 0.0555 mmol/L can be used interchangeably to convert from one unit to another. To prepare the glucose samples, a concentrated sample (i.e., 30 mg/dL) is prepared first by dissolving pure particles of glucose (molecular weight (MW)=180.16 g/mol) in distilled water and then other samples can be formed using a dilution method. For instance, in order to prepare 30 ml of the glucose sample with a concentration of 30 mg/dL, the amount of glucose particle that will be dissolved in distilled water should be determined first using the following equation:

$$M_G = \frac{C_G * V_{DI}}{100} \tag{4.1}$$

, where M_G: mass of pure glucose in mg, C_G: desired concentration in mg/dL (30 mg/dL), V_{DI}: volume of the distilled water in ml (30 ml). Using the above equation 4.1, the amount of glucose will be 9 mg.

After that, the lower concentrations samples (¡30 mg/dL) can be produced by diluting the concentrated sample. For example, 10 ml sample that has a mass concentration of 20 mg/dL can be formed after finding the required volume of the concentrated sample (i.e., 30 mg/dL) that should be mixing with distilled water. The volume of the concentrated sample can be calculated as follows:

$$V_{Concentrated} = \frac{C_{desired(diluted)} * V_{desired(diluted)}}{C_{Concentrated}} \tag{4.2}$$

, where $V_{Concentrated}$: volume of the concentrated sample in ml, $C_{Concentrated}$: mass concentration of the highly concentrated sample in mg/dL, $C_{desired(diluted)}$: the desired concentration aimed to prepare in mg/dL, $V_{desired(diluted)}$: the desired volume of the diluted sample in ml [111]. Using equation 4.2, the volume of the concentrated sample (30 mg/dL) will be 6.67 ml which will be mixed with 3.33 ml (10 ml – 6.67 ml) of distilled water (DI) in order to form 10 ml of glucose sample with a concentration of 20 mg/dL. In the following table (Tabel 4.1), I will show some of the values of glucose concentrations in both mg/dL and mmol/L.

Table 4.1: Some glucose concentrations values in both molar and mass

mg/dL	10	15	20	40	65	80	100	150	200	300
mmol/L	0.6	0.8	1.1	2.2	3.6	4.4	5.6	8.3	11.1	16.7

4.2.2 The Dipole Sensor

The experimental setup consisted of the electrically-small dipole sensor (the sensor used is adopted from [3]), a keysight 8.5 GHz VNA (E5071C), glucose-water solutions with nine different concentrations as shown in Figure 4.1

The Dipole Sensor **Glucose Solutions** **VNA**
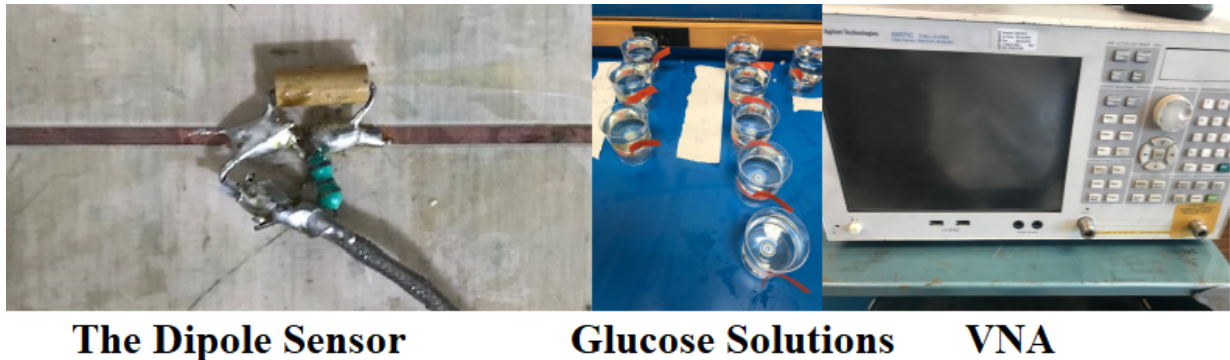
Figure 4.1: The Experimental Setup

First, I prepared the aqueous glucose solutions and placed the microwave sensor on top of those solutions to start reading the signals. Then, machine learning is applied by starting with prepossessing to clean and prepare the data for feature extraction. In the feature extraction step, I reduce the dimensionality of the feature space and include the most relevant features only. Next, I train the system to build the regression model which will be used to predict actual glucose concentrations. In the next subsections, I will give a brief explanation of the different components and techniques used by the system. The sensor was placed above the solution with a standoff distance of 5 mm (see Figure 4.2).

The Dipole Sensor

5 mm

Glucose Solution

Figure 4.2: The Dipole Sensor and its proximity to the solution

This standoff distance value was obtained experimentally as the optimal value and has no relation with the operating frequency.
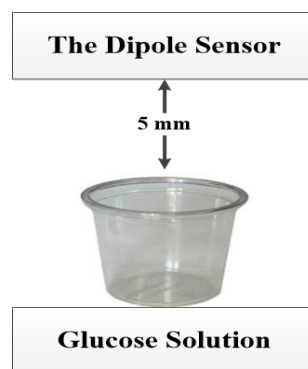
The aqueous solutions are made up of 9 different concentrations: (0 (i.e. water only), 2, 4, 6, 8, 10, 14, 28, 42 mg\dl). The magnitude and phase of the reflection coefficient (S_{11}) of the sensor were then recorded via the VNA at 201 uniformly spaced frequencies spanning the operating frequency range of 100 to 300 MHz.

It is very important to understand the nature of the data and how it looks like before starting any data analysis. In order to have a clear picture about the nature of the data, in Figures 4.3 and 4.4, I show the responses of the dipole sensor with the water-glucose solutions for the nine different concentrations using the (S_{11}) magnitude and phase. I observe that the range of frequencies that have the most notable discrimination between the responses due to different glucose levels using S_{11} magnitude and phase, are 140-240 MHz and 170-200 MHz, respectively, as shown in Figures 4.3(b) and 4.4(b). I observe that the most notable difference between the nine different glucose concentrations occurs around 180 MHz. Next, I imported and prepared the data using MATLAB [112] to be in a suitable form for feature extraction.
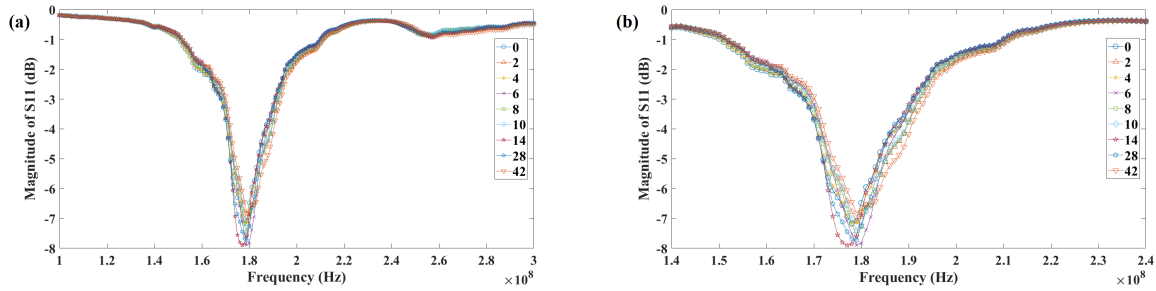

Figure 4.3: Magnitude of S_{11} for different glucose levels (a) Entire frequency range. (b) Frequency range of interest

Now, I highlighted the step of **feature extraction** in order to reduce the dimensionality of the feature space by excluding any redundant or irrelevant features and include the most discriminative features only. The data consisted of three feature vectors: magnitude, phase and frequency. Each feature vector contains 201 values, which corresponds to the 201 frequencies and the S_{11} magnitude and phase values of each frequency. I adopted two different approaches to extract the feature: **a data-driven approach and a domain-knowledge approach**. In the data-driven approach, I used Principle Component Analysis (PCA) [113] and selected the highest two Principle Components (PCs) as they preserve 95% of the variance of the whole data. In the domain knowledge-driven approach, I selected the
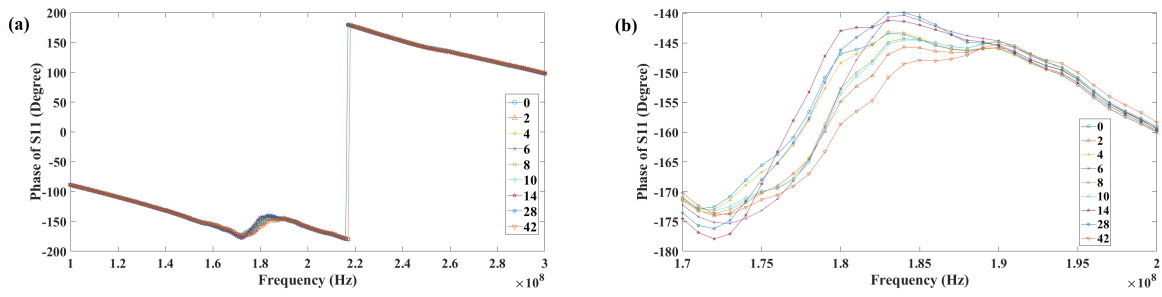
Figure 4.4: Phase of S_{11} for different glucose levels (a) Entire frequency range. (b) Frequency range of interest

frequency with the minimum S_{11} magnitude and phase (The resonance area). I extracted the minimum two values of S_{11} magnitude and phase. It is worth mentioning here that in both approaches, only two features have been selected from the entire 201 features. This selection of reduced features will help to easily train the regression model in the next step. Therefore, I have four different features: highest 2 PCs using S_{11} magnitude, highest 2 PCs using S_{11} phase, minimum two S_{11} magnitude, and minimum two S_{11} phase. Figures 4.5 and 4.6 show the values of the nine glucose levels representing by the highest two PCs using magnitude and phase of S_{11}, respectively.
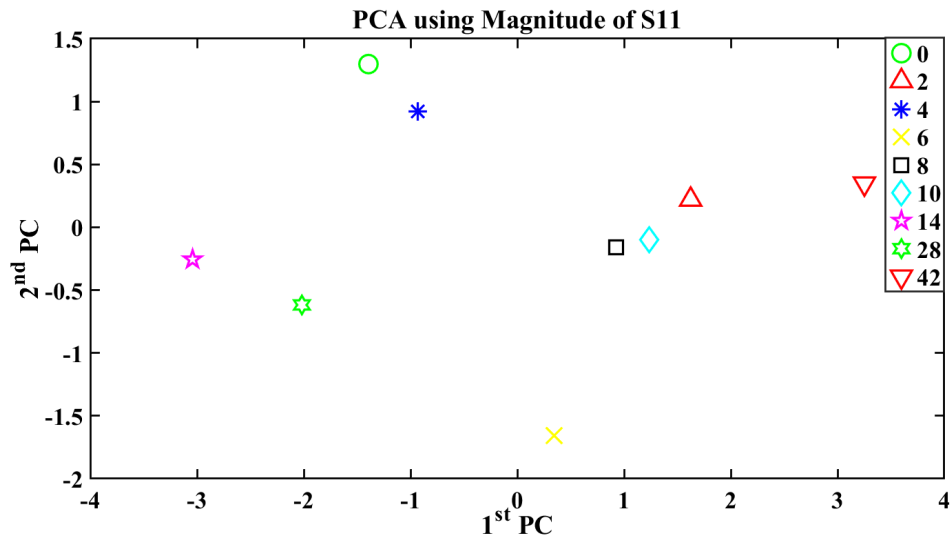

Figure 4.5: PCA using Magnitude of S_{11} with different glucose levels

Figures 4.7 and 4.8 show the values of the nine glucose levels representing by the
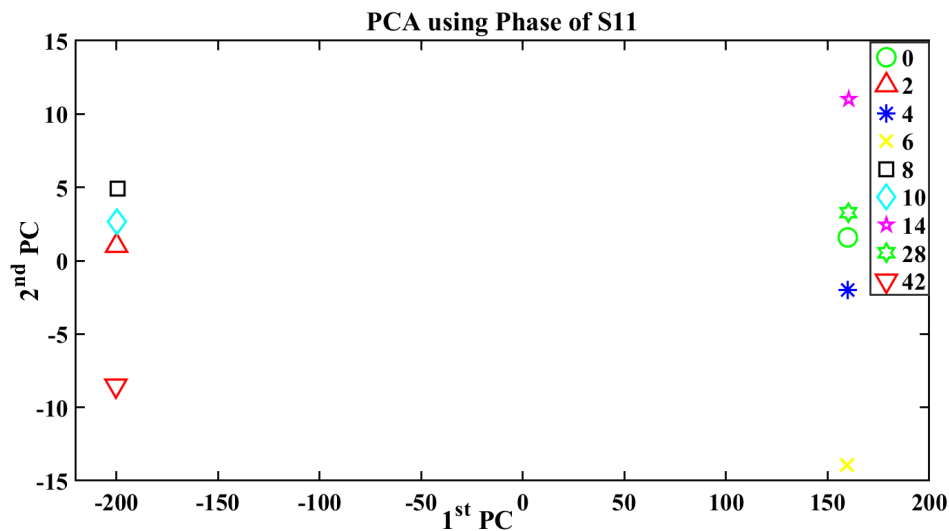
47

Figure 4.6: PCA using Phase of S_{11} with different glucose levels

minimum two S_{11} magnitude and phase, respectively. As we can see from all the figures of representing the glucose levels with the different extracted features, there is a noticeable difference between the glucose levels, however, this difference is not the main goal of the model, I need to have a good regression function that will easily map each value to its corresponding glucose level with a high accuracy.

Now I will show the results of the best regression models using all the different features approaches (Minimum-two S_{11} Magnitude and Phase , and Highest-two PCs using both magnitude and phase). I used 3-folds cross validation as it is more suitable in this case based on the size of the data. As shown in Table 4.2, results from the Gaussian SVR algorithm [114] showed the least RMSE with 11.9 using the two-minimum magnitude of S_{11}. Figure 4.9 shows the response plot of the prediction model by the Gaussian SVR algorithm using the the minimum magnitude of S_{11} to predict the actual aqueous glucose levels. Whereas, the least RMSE using the two-minimum phase of S_{11} was 12.3 given by same Gaussian SVR algorithm.

Results of the regressions using highest two PCs of magnitude and phase of S_{11} gave RMSE = 13.4 and 14, respectively. Those results obtained using linear SVR algorithm and Rational Quadratic GPR, respectively [115].
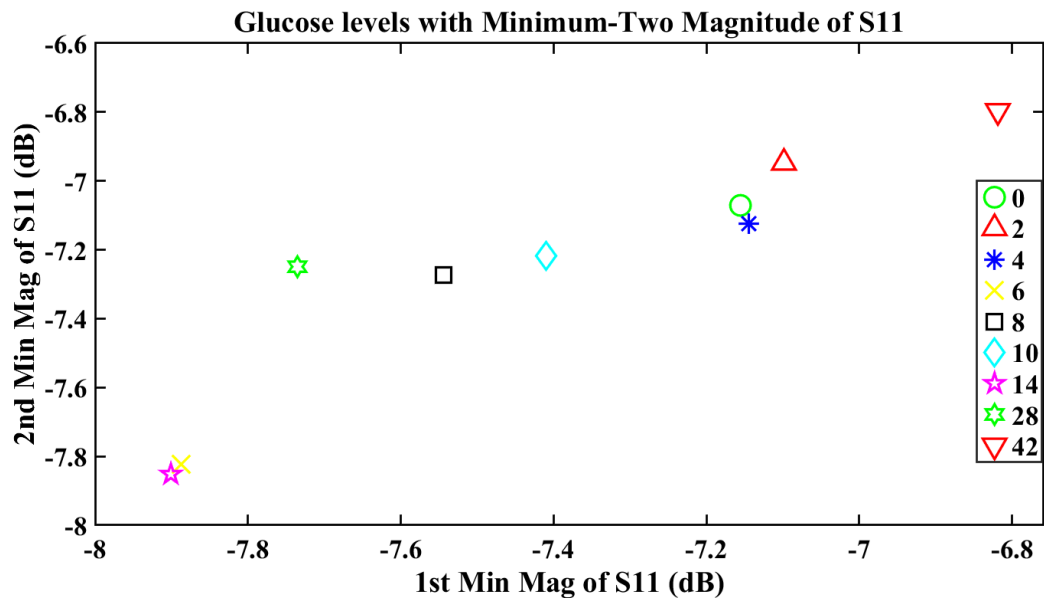
48

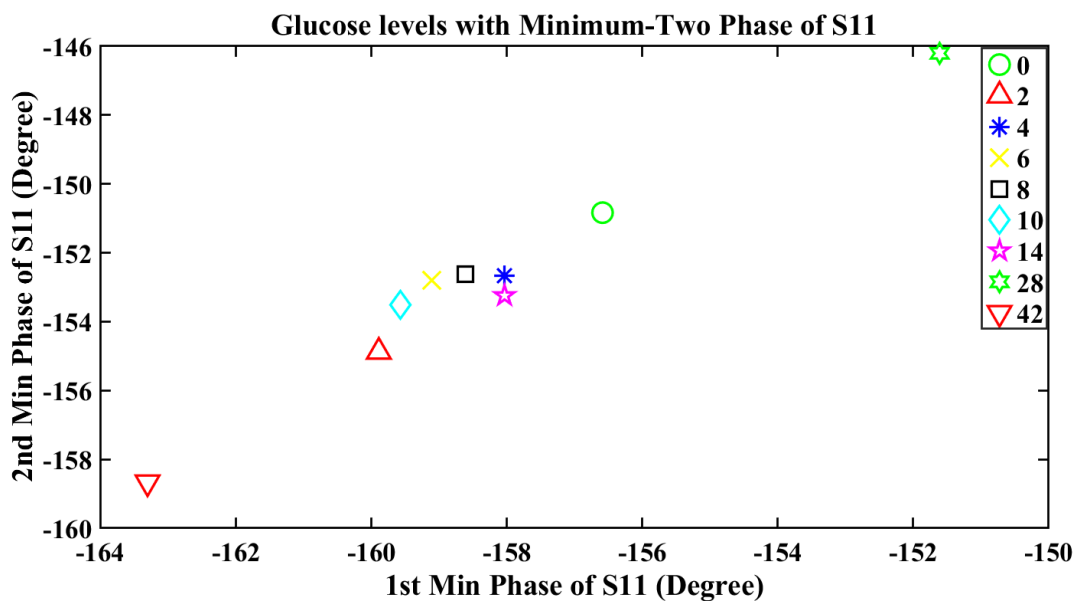Figure 4.7: Minimum two Magnitude of S_{11} with different glucose levels

Figure 4.8: Minimum two Phase of S_{11} with different glucose levels

49

Table 4.2: RMSE of some of the regression algorithms using the dipole sensor in experiments

Regression Algorithm	Min Mag.	Min Phase	PCA Mag.	PCA Phase
Linear Regression	15.2	22.7	16	20.6
SVR	11.9	12.3	13.4	16.4
Matern5/2 GPR	15.5	13.3	13.5	14.8
Exponential GPR	14.7	13.4	13.5	14.2
Rational Quadratic GPR	15.6	13.4	13.5	14
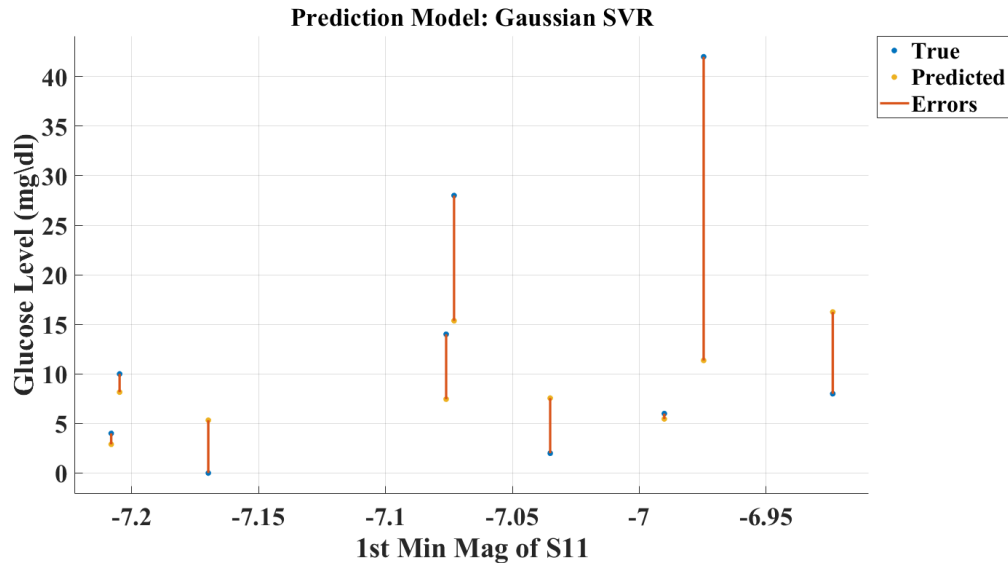
Figure 4.9: Response plot of the prediction model using minimum magnitude of S_{11}

4.2.3   The Spiral Sensor

In this subsection, I will show the experimental results using the spiral sensor. **The experimental setup** is the same as shown in the previous subsection about dipole sensor except the sensor changed only. Figures 4.10 and 4.11 show the responses of the spiral sensor with the water-glucose solutions for the nine different concentrations using the (S_{11}) magnitude and phase. I observed that the range of frequencies that have the most notable discrimination between the responses due to different glucose levels using S_{11} magnitude and phase, are 1150-1200 MHz, as shown in Figures 4.10(b) and 4.11(b). I observed that the most notable difference between the nine different glucose concentrations occurs around

1170 MHz.

Figure 4.10: Magnitude of S$_{11}$ for different glucose levels (a) Entire frequency range. (b) Frequency range of interest
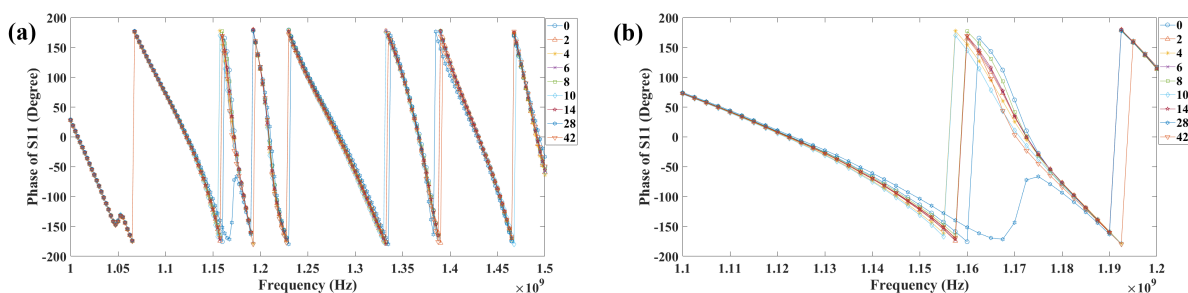
Figure 4.11: Phase of S$_{11}$ for different glucose levels (a) Entire frequency range. (b) Frequency range of interest
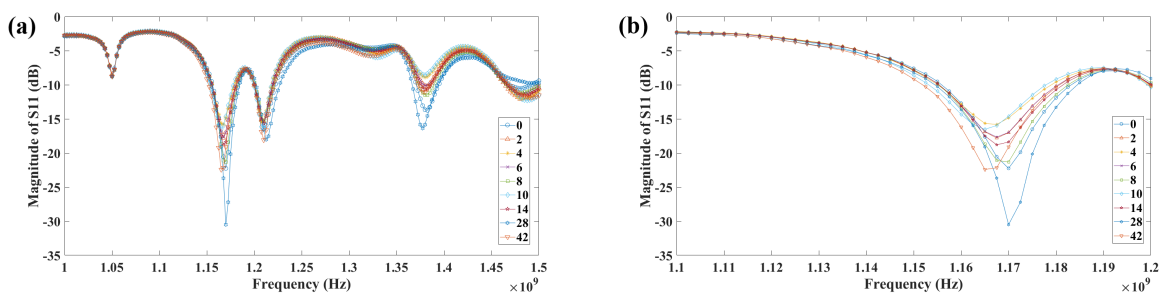
In **feature extraction**, Figures 4.12 and 4.13 show the values of the nine glucose levels representing by the highest two PCs using magnitude and phase of S$_{11}$, respectively.

In the same way, Figures 4.14 and 4.15 show the values of the nine glucose levels representing by the minimum two S$_{11}$ magnitude and phase, respectively.

Now I will show the results of the best regression models using all the different features approaches (Minimum-two S$_{11}$ Magnitude and Phase , and Highest-two PCs using both magnitude and phase). I used 3-folds cross validation. As shown in Table 4.3, results from the linear regression algorithm showed the least RMSE with 7.26 using PCA of magnitude of S$_{11}$. Figure 4.16 shows the response plot of the prediction model by the linear regression algorithm using the highest PCs of magnitude of S$_{11}$ to predict the actual aqueous glucose levels. On the other hand, the least RMSE using PCA of phase of S$_{11}$ was 13.5 given by Matern 5/2 GPR algorithm [115].

51
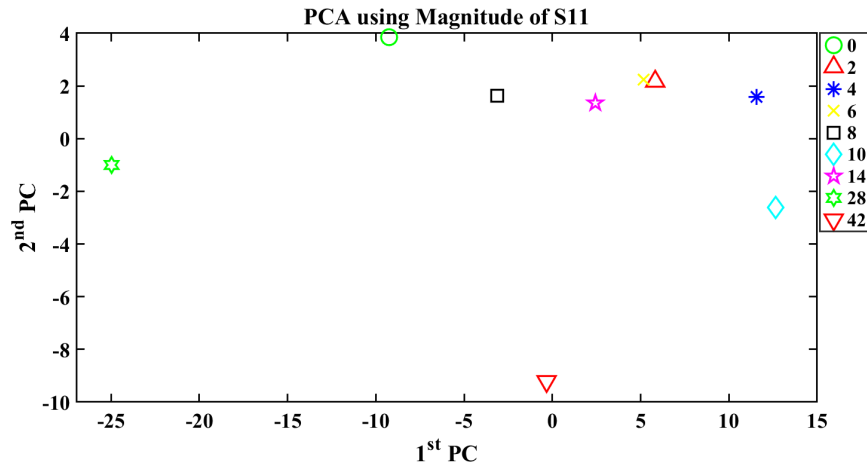
Figure 4.12: PCA using Magnitude of S_{11} with different glucose levels
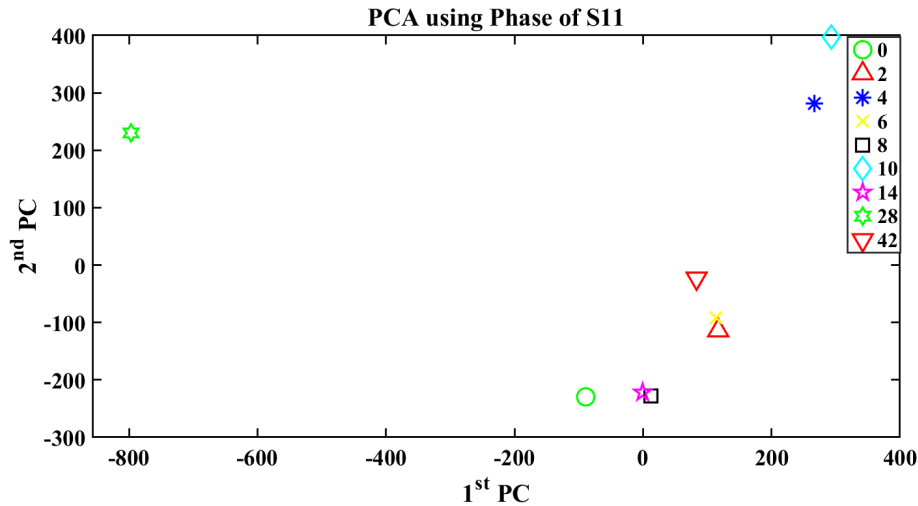
Figure 4.13: PCA using Phase of S_{11} with different glucose levels

Results of the regressions using minimum-two magnitude and phase of S_{11} gave RMSE = 8.8 and 13, respectively. Those results obtained using linear SVR algorithm and Exponential GPR, respectively [114].

After that, I train the system using the data from one sensor and test the system using the data of the other sensor. When I test the system using dipole on the trained spiral regression model, I got RMSE = 13 despite the different frequencies of the sensors. In the same way, when I test the system using spiral on the trained dipole regression model, I got
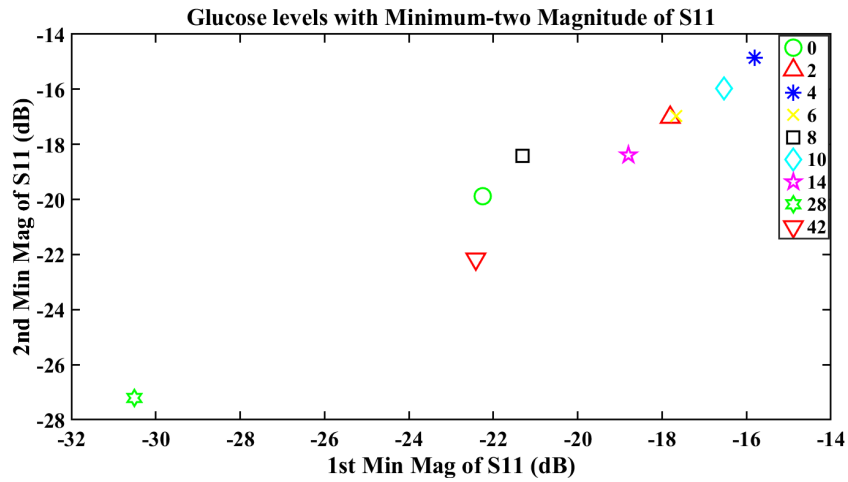
Figure 4.14: Minimum two Magnitude of S$_{11}$ with different glucose levels
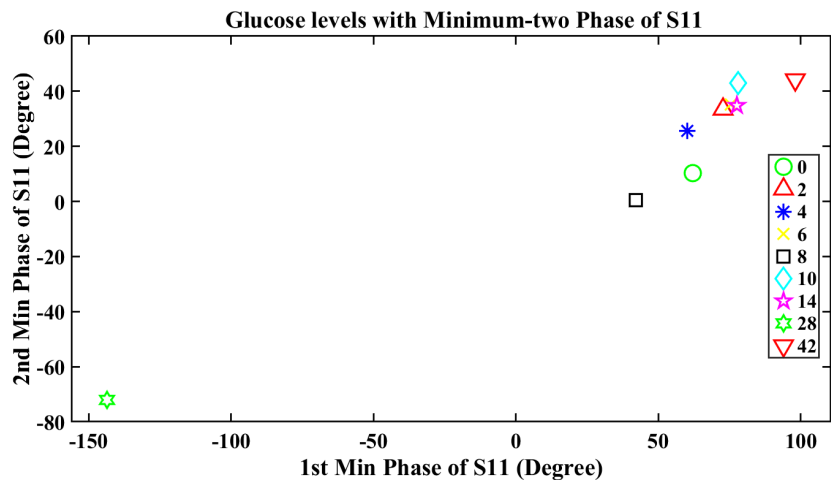
Figure 4.15: Minimum two Phase of S$_{11}$ with different glucose levels

RMSE = 13.4

4.3   Simulation Results:

In this section, I will show the results of simulation of the microwave sensor for constructing a non-invasive CGM system using the full-wave numerical simulation tool CST Microwave

Table 4.3: RMSE of some of the regression algorithms using the spiral sensor in experiments

| Regression Algorithm	Min Mag.	Min Phase	PCA Mag.	PCA Phase
Linear Regression	8.8	15.8	7.26	14
SVR	11.1	13.6	11.7	13.6
Matern5/2 GPR	14.8	13	14.5	13.5
Exponential GPR	14.4	13	14.9	13.6
Rational Quadratic GPR	14.4	13.1	14.7	13.7
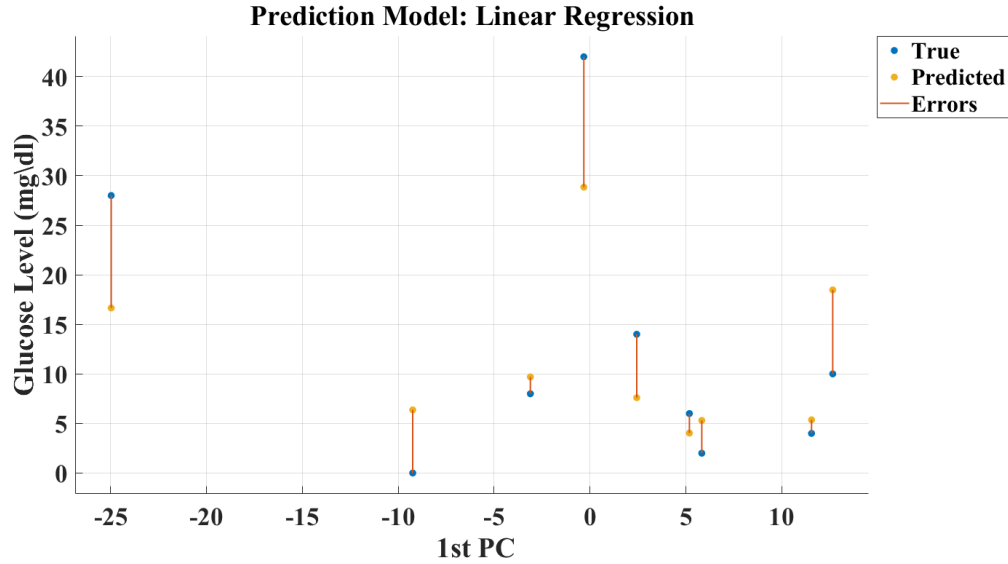
Figure 4.16: Response plot of the prediction model using highest PC of magnitude of S_{11}

Studio [100]. The main building blocks of the **simulation setup** are the microwave sensor and the hand phantom on which I test the sensor. I already showed how the sensor was designed in 2. In the next subsection, I will highlight the design of the hand phantom with different layers.

4.3.1 Hand model design

In order to simulate the effect of the sensors, a hand phantom model is designed as four layers namely, skin (1.5 mm), fat (2 mm), blood (2.5 mm) and muscle (15 mm) layers as shown in Figure. 4.17. Those thickness values are adopted from the literate [116, 59, 117].

I chose the hand part of the human body because of the following reasons:

- One of the easiest place to put the sensor on it.

- Has minimal fat and muscles close to the skin.

- No dangerous because of microwave radiation.

- The goal of the final product is to build a system and wear it like a watch.
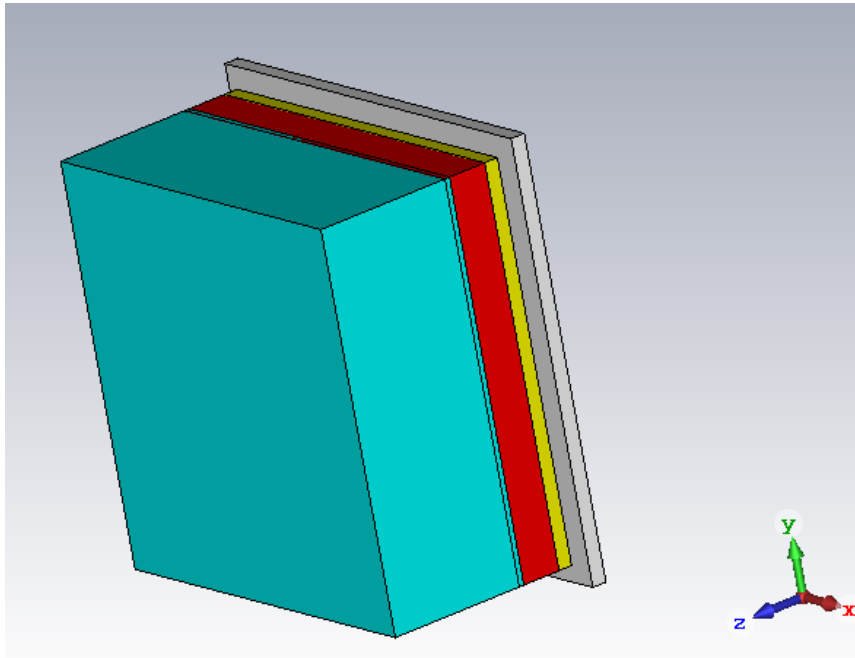
Figure 4.17: The Hand Model layers

4.3.2 Dielectric properties of the hand model layers/tissues

The main idea of the simulation results is coming from changing the dielectric properties (permittivity ($\hat{\varepsilon}$) and conductivity σ) of each layer and saving the reflection coefficient (S11) of the sensor based on the interaction of the microwave sensor with the tissues having those dielectric properties. I will have more emphasize on the blood layer because of the reasons discussed earlier in chapter 1, section 1.10 about the body fluids considerations.

In the next subsection, I will investigate the change in microwave resonator response as a result of the variation in dielectric properties (permittivity and conductivity) of one tissue/layer at a time. At each layer, dielectric properties are varied while other layers remain constant. By this way, I can make sure that the sensor can see the corresponding layer/tissue, i.e. the penetration depth of the sensor is enough. Note that I focused first on the spiral sensor for the simulation.

The dimension of the simulation data is 1001 features corresponding to the 1001 uniformly spaced frequencies spanning the operating frequency range of 1000-1500 MHz. Because I used the magnitude and phase of S11, simulation data consisted of three feature vectors: magnitude, phase and frequency, each feature vector contains 1001 values.

4.3.3 Blood Layer

In this subsection, I will show the results of applying the detailed steps of the methodology of the system as shown earlier in chapter 1, Figure 1.4. The steps will be applied for the results obtained by both changing the permittivity and the conductivity of the blood tissue/layer of the hand model.

Permittivity(ε)

Here I will show the results of the system by changing the permittivity of the blood layer. Permittivity values for the blood layer are in the range [40,105], where the base value (60) are adopted from [118], for the corresponding frequency range of the spiral sensor (1000-1500 MHz), with some increments and decrements in order to reflect the expected change caused by glucose molecules in the blood.

In order to see how the simulation data looks like and to have a clear picture about the nature of these data, Figures 4.18 and 4.19 show the responses of the spiral sensor to the different permittivity values using magnitude and phase of (S_{11}), respectively. I observed that the range of frequencies that have the most notable discrimination between the responses due to different permittivity values using magnitude and phase of S_{11}, are 1410-1470 MHz.

Next, I imported and prepared the data using MATLAB [112] to be in a suitable form for the next steps. In **data-driven feature extraction**, Figures 4.20 and 4.21 show the different values of the permittivity representing by the highest two PCs using magnitude
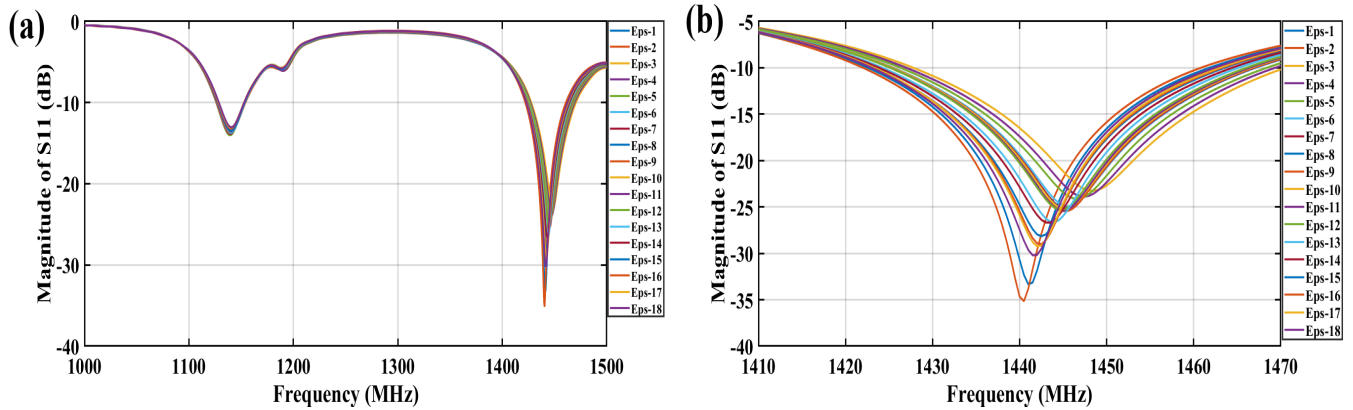
Figure 4.18: Magnitude of S_{11} for different values for Permittivity of Blood (a) Entire frequency range. (b) Frequency range of interest
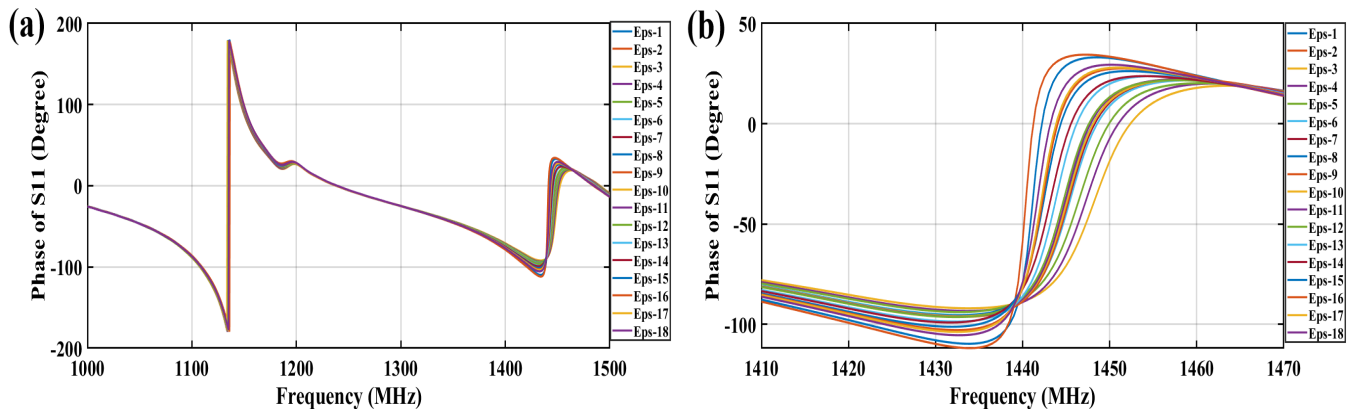
Figure 4.19: Phase of S_{11} for different values for Permittivity of Blood (a) Entire frequency range. (b) Frequency range of interest

and phase of S_{11}, respectively. As we can see from the figures, PCA using magnitude of S11 gave us some good patterns that may help in the next step of regression.

After that, I will show the results of the 2nd category of feature engineering, which is based on domain-knowledge. Figures 4.22 and 4.23 show the different values of the permittivity representing by the minimum-two magnitude and phase of S_{11}, respectively. It is very clear that this representation of features can be represented by a linear line as shown in the figures.

Based on the fact that by increasing the frequency, permittivity decreases [116], I tried

57
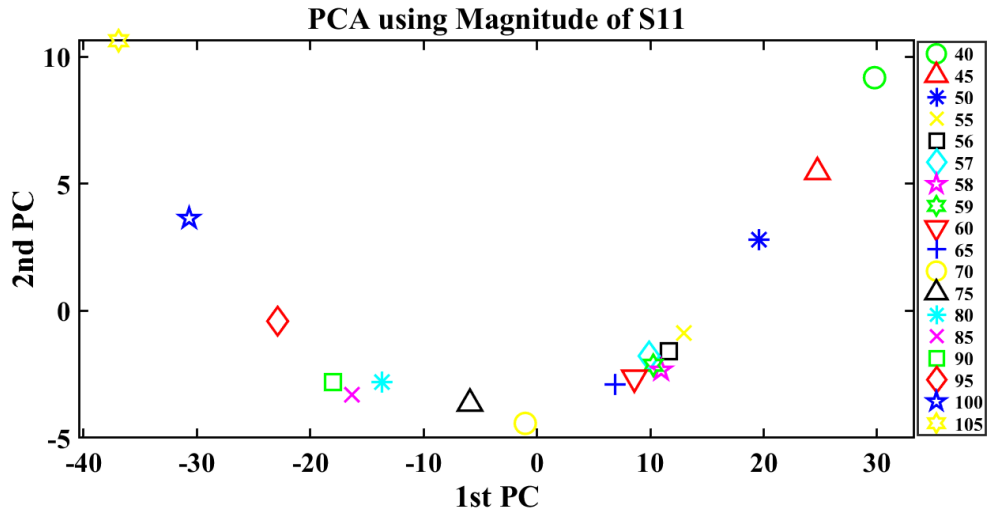
Figure 4.20: PCA using Magnitude of S_{11} with different values for Permittivity of Blood
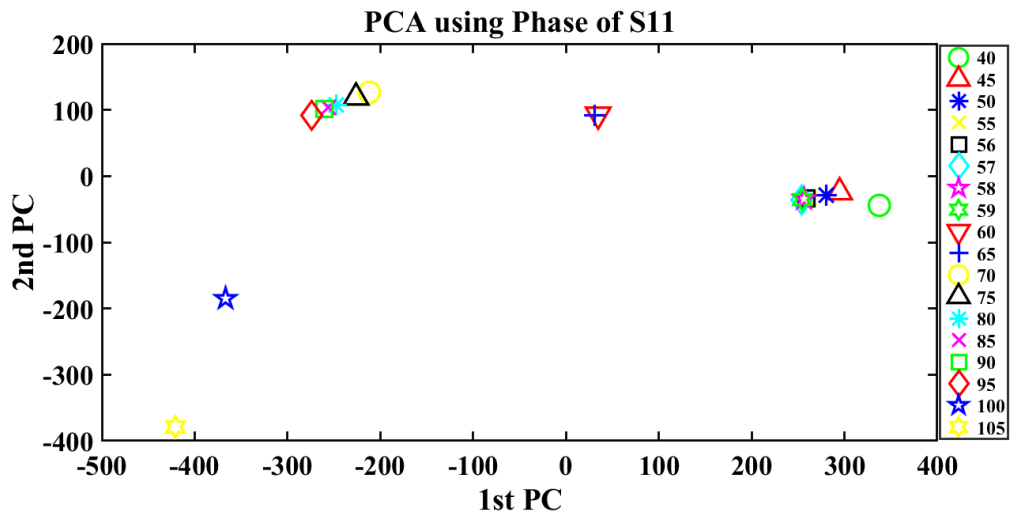
Figure 4.21: PCA using Phase of S_{11} with different values for Permittivity of Blood

to benefit from the resonance frequency as a feature. Figures 4.24 and 4.25 show the different values of the permittivity representing by the minimum frequency (resonance) and minimum S_{11} magnitude and phase, respectively. In case of the phase of S_{11}, the resonance frequency is at about 1133.5-1136 MHz which is not in the frequency range of interest and therefore I can tell that this feature is not a discriminating one.
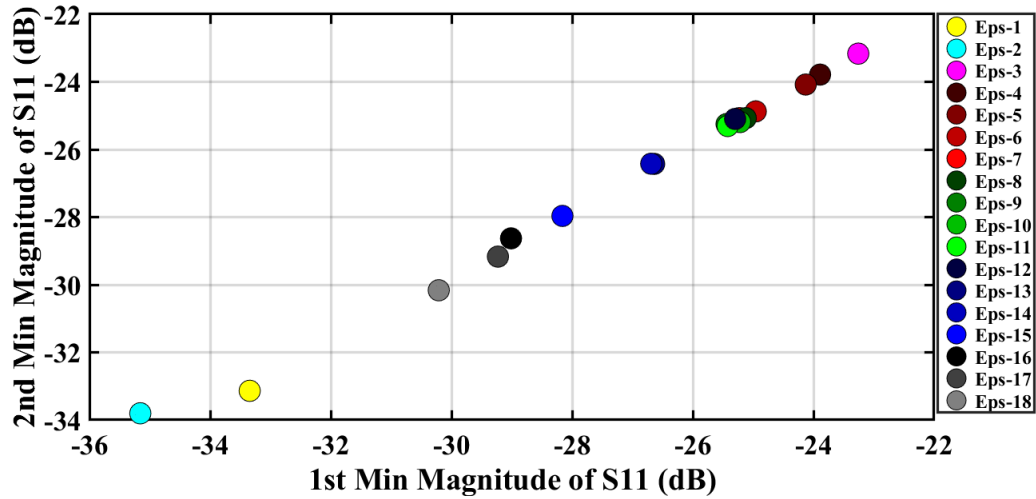
Figure 4.22: Minimum-two Magnitude of S_{11} with different values for Permittivity of Blood
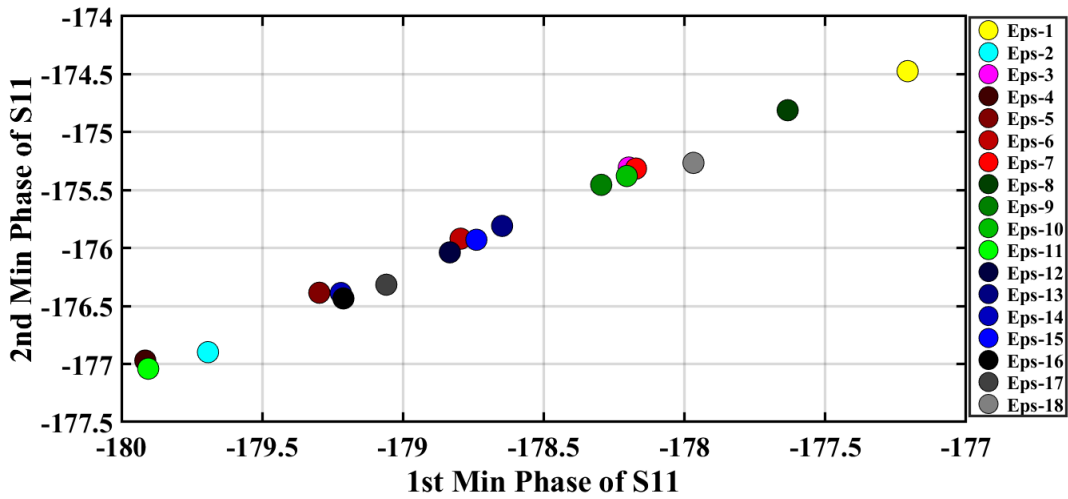
Figure 4.23: Minimum-two Phase of S_{11} with different values for Permittivity of Blood

After preparing the features, I reached to the final step of the system which is predicting the values of the permittivity of blood tissue using those feature extracted from the S11 reflected coefficient of the spiral sensor. The goal is to find a regression function to map between the S11 features and the permittivity of the tissue. Therefore, after selecting the most relevant features, they were inputted to train the regression model which will be used
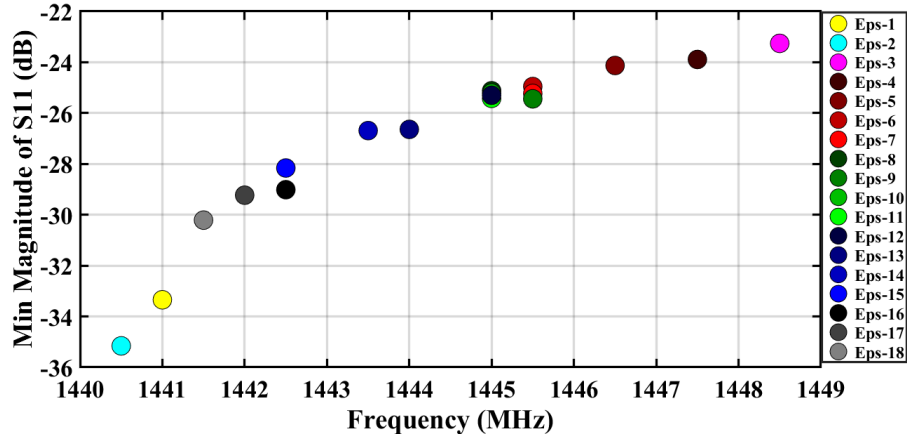
Figure 4.24: Minimum Frequency and Magnitude of S_{11} with different values for Permittivity of Blood
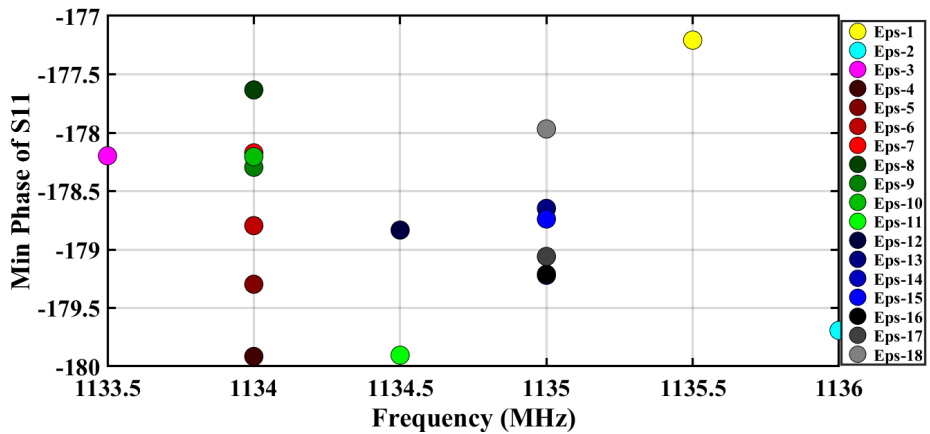
Figure 4.25: Minimum Frequency and Phase of S_{11} with different values for Permittivity of Blood

to predict the blood permittivity. A total of 19 different regression models were trained using Matlab regression learning App. RMSE is used as the criteria to select the most accurate regression model (least RMSE) Table 4.4 shows the results of the best regression models using all the different features approaches (Minimum-two S_{11} Magnitude and Phase , and Highest-two PCs using both magnitude and phase). I used 3-folds cross validation.

Results from the Quadratic SVR (QSVR) algorithm showed the least RMSE with 1.98 using PCA of S_{11} magnitude. Whereas, the least RMSE using Minimum-two magnitude

Table 4.4: RMSE of some of the regression algorithms using the spiral sensor with Permittivity of Blood

Regression Algorithm	Min Mag.	Min Phase	PCA Mag.	PCA Phase
Linear Regression	2.8	9.3	2.1	5.8
SVR	3.7	20.1	1.98	6.7
Matern5/2 GPR	3.9	19.6	3.5	10.6
Exponential GPR	2.7	19.6	2.2	11.2
Rational Quadratic GPR	4.1	20.6	2.2	10.2
Decision Tree	11.5	20.4	10.8	10.5

of S_{11} was 2.7 given by Matern 5/2 GPR algorithm. In this case, I found that the phase of S_{11} did not give us good results using both minimum-two phase and highest-two PCs.

Conductivity (σ)

Now I will show the results of the system by changing the conductivity of the blood layer. Conductivity values for the blood layer are in the range [0.1,1.9], values are adopted from [118] for the corresponding frequency range of the spiral sensor (1000-1500 MHz) in order to reflect the expected change in conductivity caused by glucose molecules in the blood.

In order to see how the simulation data looks like and to have a clear picture about the nature of these data, Figures 4.26 and 4.27 show the responses of the spiral sensor to the different conductivity values using magnitude and phase of (S_{11}), respectively. I observed that the range of frequencies that have the most notable discrimination between the responses due to different conductivity values using magnitude and phase of S_{11} , are 1410-1470 MHz, which is the same as in case of permittivity.

Next step is **feature extraction** following data-driven approach, Figures 4.28 and 4.29 show the different values of the conductivity representing by the highest two PCs using magnitude and phase of S_{11}, respectively.

After that, the results of domain-knowledge feature extraction category will be shown. Figures 4.30 and 4.31 show the different values of the conductivity representing by the minimum-two magnitude and phase of S_{11}, respectively. This representation of features shows some kinds of linearity as shown in the figures.

Figures 4.32 and 4.33 show the different values of the conductivity representing by the
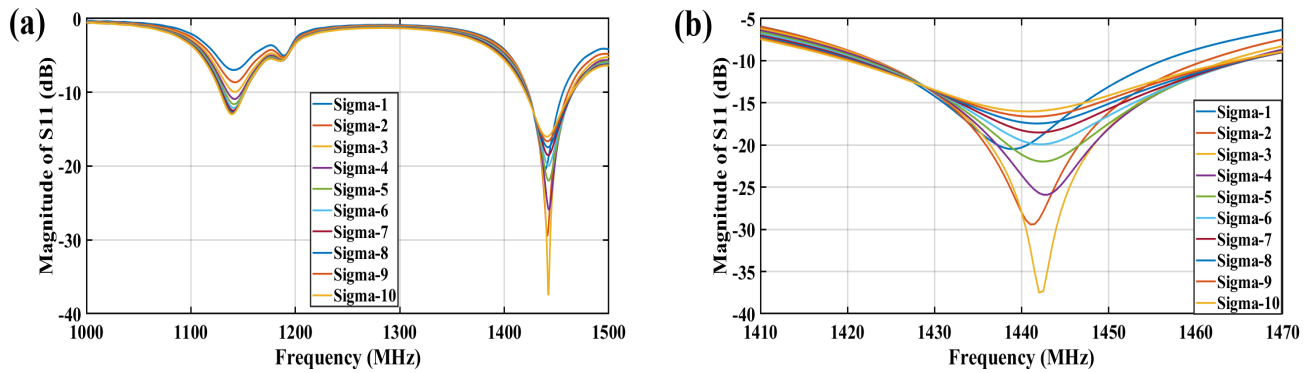
Figure 4.26: Magnitude of S_{11} for different values for Conductivity of Blood (a) Entire frequency range. (b) Frequency range of interest
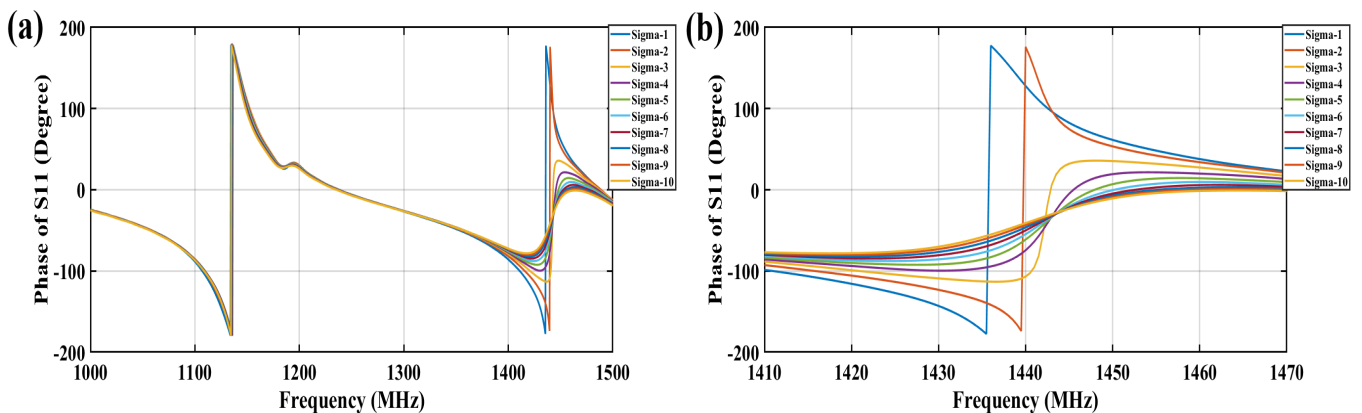
Figure 4.27: Phase of S_{11} for different values for Conductivity of Blood (a) Entire frequency range. (b) Frequency range of interest

minimum frequency (resonance) and minimum S_{11} magnitude and phase, respectively. In case of the phase of S_{11}, the resonance frequency is at about 1134-1136 MHz which is not in the frequency range of interest and therefore I can tell that this feature is not a discriminating one.

Next, I will show the results of predicting the values of the conductivity of blood tissue using those feature extracted from the S11 reflected coefficient of the spiral sensor. The goal is to find a regression function to map between the S11 features and the conductivity of the tissue. Therefore, after selecting the most relevant features, they were inputted to train the regression model which will be used to predict the blood conductivity. A total of
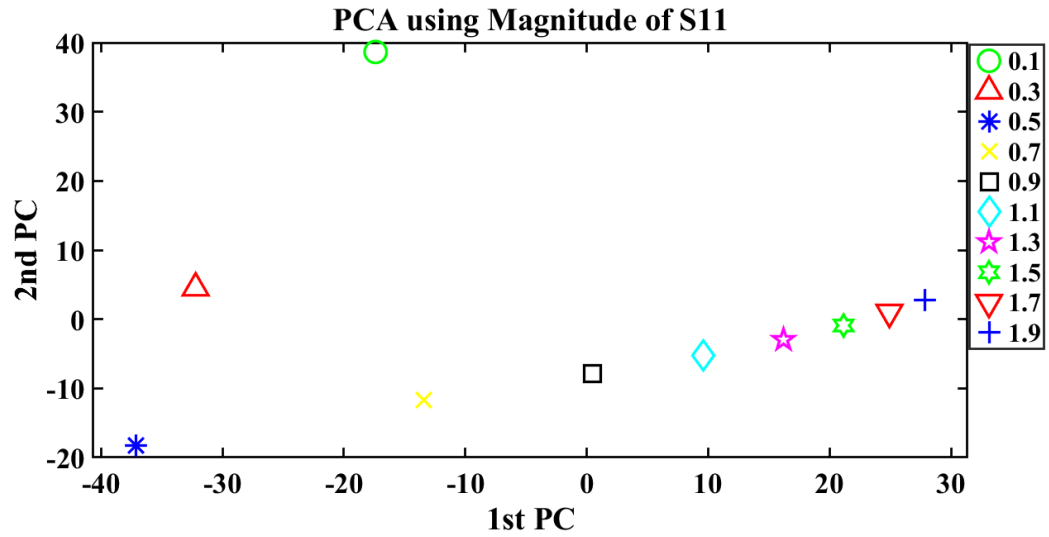
Figure 4.28: PCA using Magnitude of S_{11} with different values for Conductivity of Blood
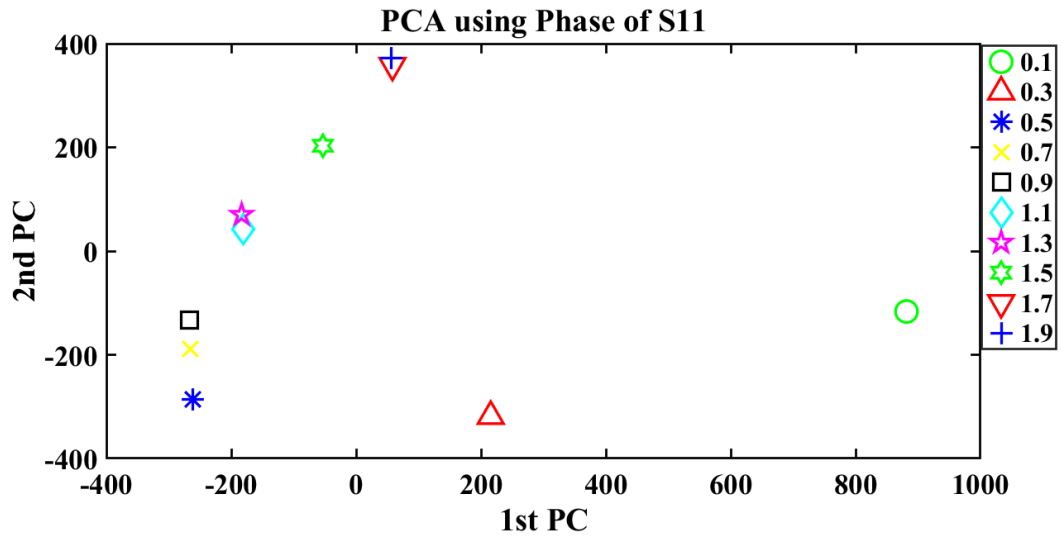
Figure 4.29: PCA using Phase of S_{11} with different values for Conductivity of Blood

19 different regression models were trained using Matlab regression learning App. RMSE and R^2 are used as the criteria to select the most accurate regression model (least RMSE and highest R^2).

Table 4.5 shows the results of only the best regression models using all the different
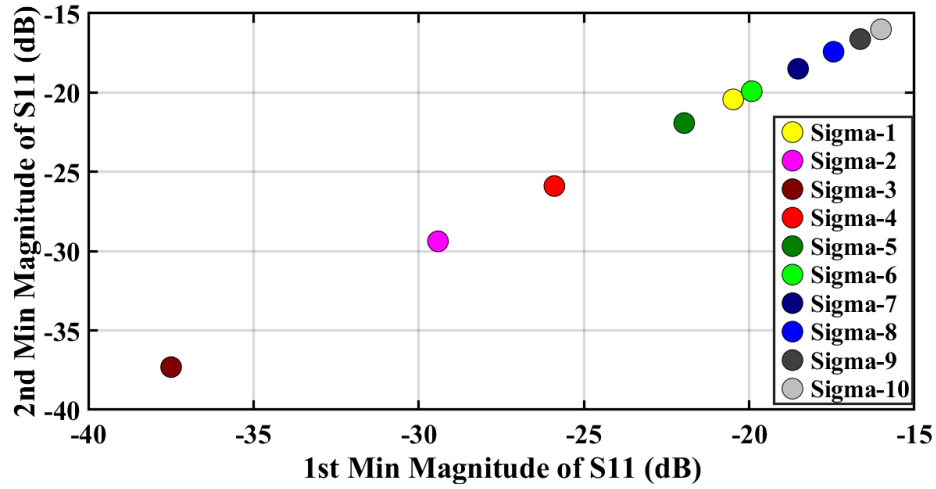
Figure 4.30: Minimum-two Magnitude of S_{11} with different values for Conductivity of Blood
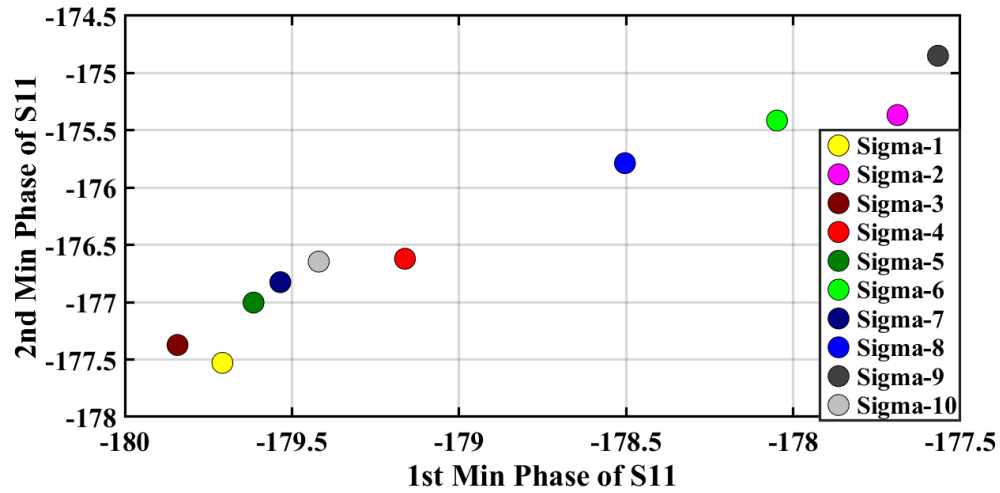
Figure 4.31: Minimum-two Phase of S_{11} with different values for Conductivity of Blood

features approaches (Minimum-two S_{11} Magnitude and Phase, and Highest-two PCs using both magnitude and phase). 3-folds cross validation was used to train and test the system.

Results from the Linear Regression algorithm showed the least RMSE with 0.1 and

64

Figure 4.32: Minimum Frequency and Magnitude of S$_{11}$ with different values for Conductivity of Blood
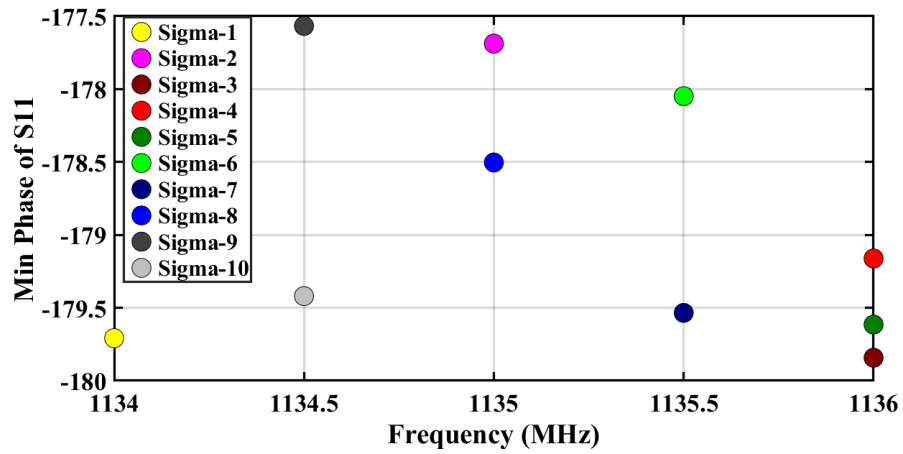
Figure 4.33: Minimum Frequency and Phase of S$_{11}$ with different values for Conductivity of Blood

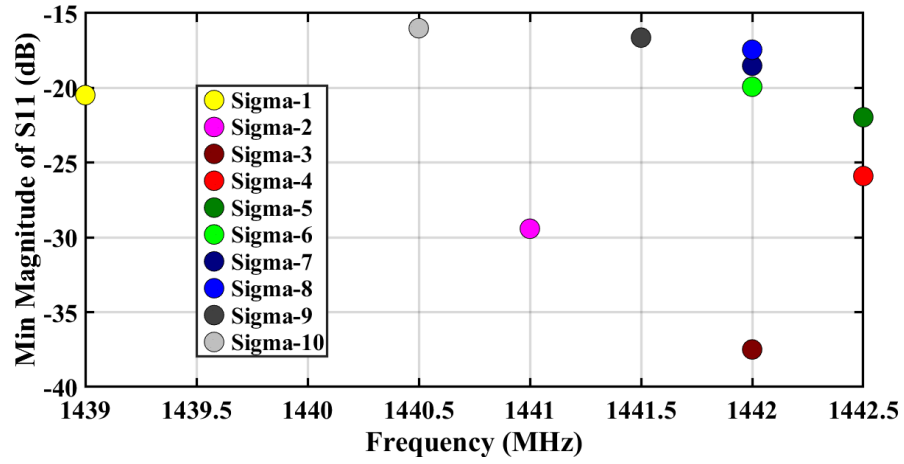highest $R^2 = 0.95$ using PCA of Phase of S$_{11}$.

Table 4.5: RMSE of some of the regression algorithms using the spiral sensor with Conductivity of Blood

Features	Best Regression Algorithm	Metrics	
		RMSE	R^2
Min Mag.	Quadratic SVR	0.4	0.67
Min Phase	Linear Regression	0.3	0.66
PCA Mag.	Linear Regression	0.2	0.84
PCA Phase	Linear Regression	0.1	0.95

4.3.4 Skin Layer

In this subsection, I will show the results of changing the permittivity and the conductivity of the skin tissue/layer of the hand model.

Permittivity(ε)

I will show the results of the system by changing the permittivity of the skin layer. Permittivity values for the skin layer are in the range [40,300], where the values for the dry skin and wet skin from [118] are between 40-45 for the corresponding frequency range of the spiral sensor (1000-1500 MHz).

In order to see how the simulation data looks like and to have a clear picture about the nature of these data, Figures 4.34 and 4.35 show the responses of the spiral sensor to the different permittivity values using magnitude and phase of (S_{11}), respectively. I observed that the range of frequencies that have the most notable discrimination between the responses due to different permittivity values using magnitude and phase of S_{11}, are 1410-1470 MHz.

Next, is **the feature extraction** step. Figures 4.36 and 4.37 show the different values of the permittivity representing by the highest two PCs using magnitude and phase of S_{11}, respectively. This is the feature extraction following data-driven approach.

Next, I will show the results of the 2nd category of feature engineering, which is based on domain-knowledge. Figures 4.38 and 4.39 show the different values of the permittivity representing by the minimum-two magnitude and phase of S_{11}, respectively. It is very clear that this representation of features produced some interesting patterns as shown in the figures.

66
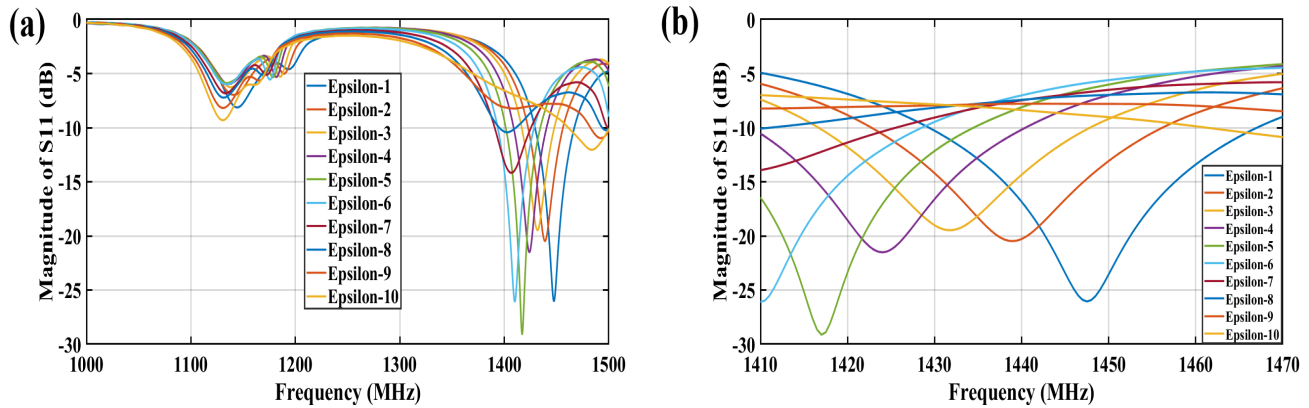
Figure 4.34: Magnitude of S_{11} for different values for Permittivity of Skin (a) Entire frequency range. (b) Frequency range of interest
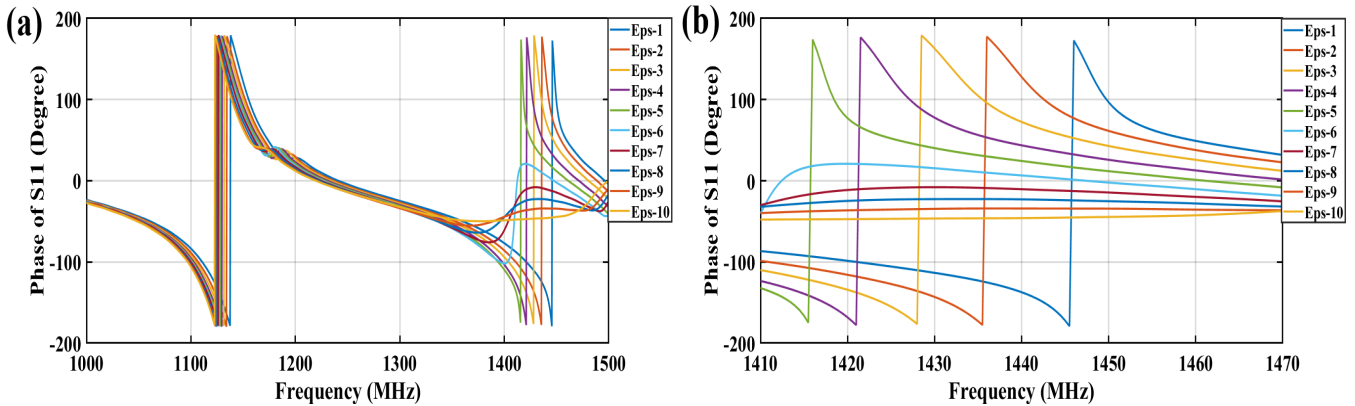
Figure 4.35: Phase of S_{11} for different values for Permittivity of Skin (a) Entire frequency range. (b) Frequency range of interest

Figures 4.40 and 4.41 show the different values of the permittivity representing by the minimum frequency (resonance) and minimum S_{11} magnitude and phase, respectively. In case of the phase of S_{11}, majority of the resonance frequency is at about 1100-1150 MHz which is not in the frequency range of interest and therefore I can tell that this feature is not a discriminating one.

Now, I reached to the final step of the system, using permittivity of skin, which is predicting the values of the permittivity of skin tissue using those feature extracted from the S11 reflected coefficient of the spiral sensor. The goal is to find a regression function to

67
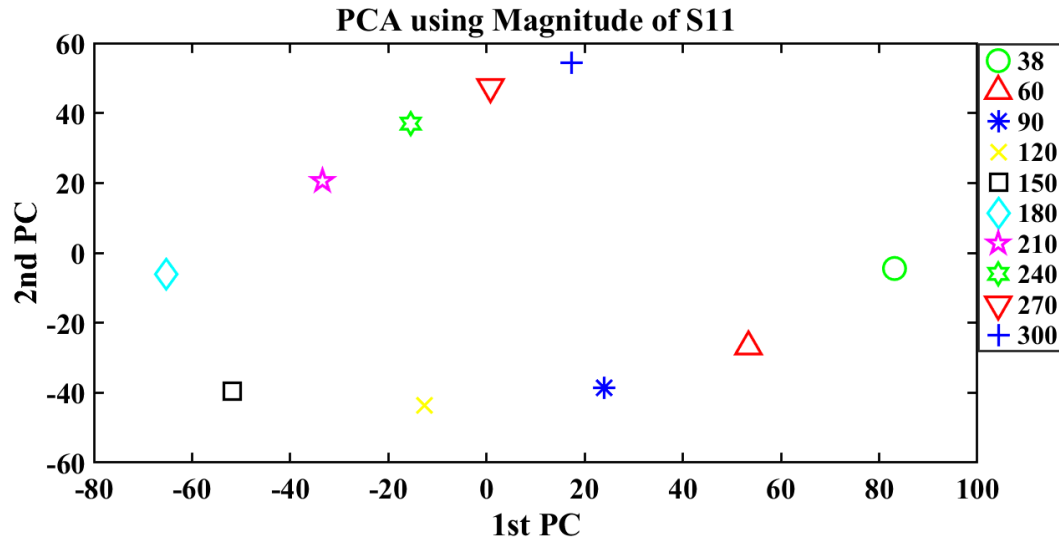
Figure 4.36: PCA using Magnitude of S_{11} with different values for Permittivity of Skin
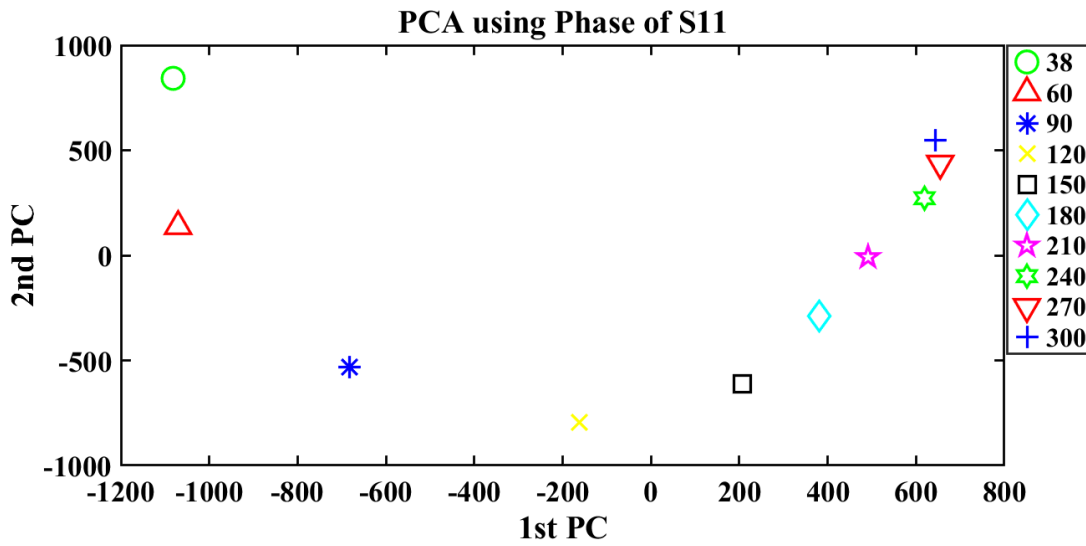
Figure 4.37: PCA using Phase of S_{11} with different values for Permittivity of Skin

map between the S11 features and the permittivity of the tissue. Therefore, after selecting the most relevant features, they were inputted to train the regression model which will be used to predict the skin permittivity. A total of 19 different regression models were trained using Matlab regression learning App. RMSE and R^2 are used as the criteria to select the

68
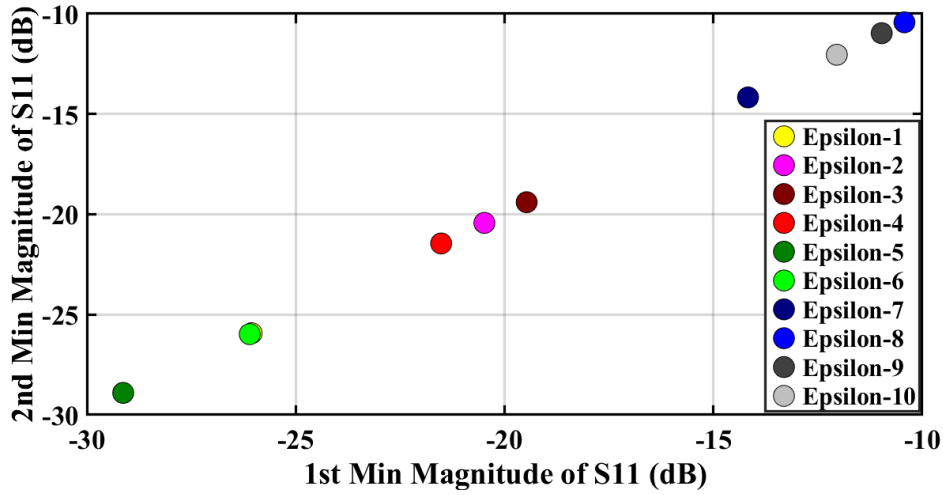
Figure 4.38: Minimum-two Magnitude of S_{11} with different values for Permittivity of Skin
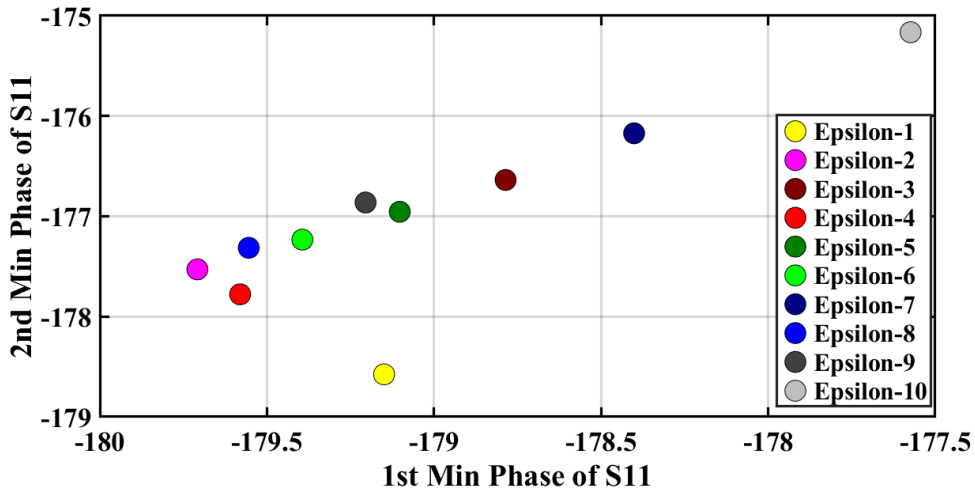
Figure 4.39: Minimum-two Phase of S_{11} with different values for Permittivity of Skin

most accurate regression model (least RMSE and highest R^2). Table 4.6 shows the results of the best regression models using all the different features approaches (Minimum-two S_{11} Magnitude and Phase , and Highest-two PCs using both magnitude and phase). I used 3-folds cross validation to validate the system.
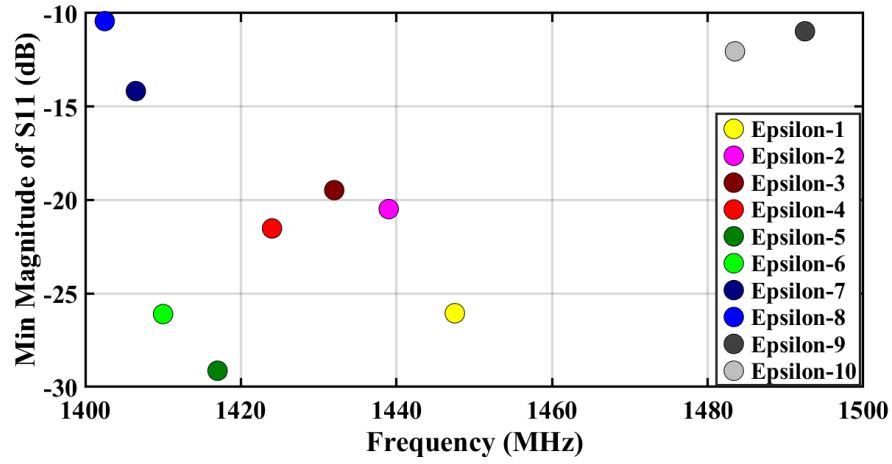
Figure 4.40: Minimum Frequency and Magnitude of S_{11} with different values for Permittivity of Skin
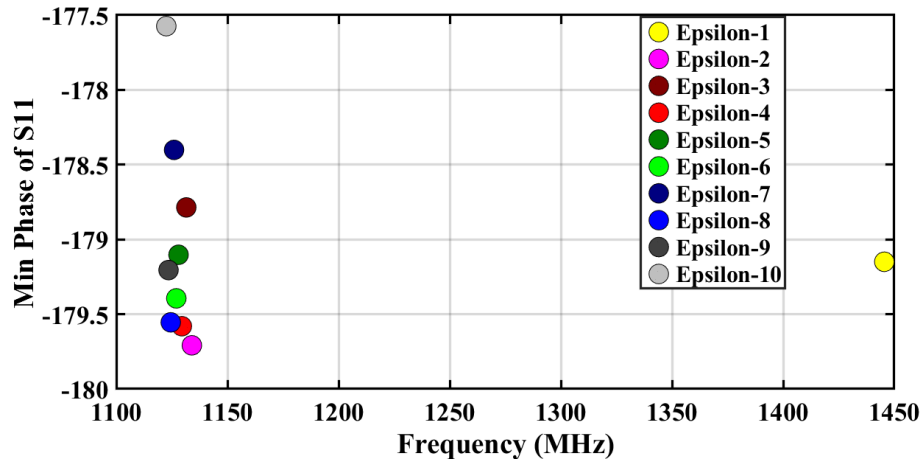
Figure 4.41: Minimum Frequency and Phase of S_{11} with different values for Permittivity of Skin

Results from the Rational Quadratic GPR algorithm showed the least RMSE with 11.2 and highest $R^2 = 0.98$ using PCA of Magnitude of S_{11}.

Table 4.6: RMSE of some of the regression algorithms using the spiral sensor with Permittivity of Skin

Features	Best Regression Algorithm	Metrics	
		RMSE	R^2
Min Mag.	Medium Gaussian SVR	60.9	0.68
Min Phase	Linear SVR	68	0.55
PCA Mag.	Rational Quadratic GPR	11.2	0.98
PCA Phase	Linear Regression	14.1	0.98

Conductivity (σ)

Now I will show the results of the system by changing the conductivity of the skin layer. Conductivity values for the skin layer are in the range [1.5,3.3], values are adopted from [118] for the corresponding frequency range of the spiral sensor (1000-1500 MHz) in order to reflect the expected change in conductivity caused by glucose molecules in the skin.

In order to see how the simulation data using the skin conductivity and to have a clear picture about the nature of these data, Figures 4.42 and 4.43 show the responses of the spiral sensor to the different conductivity values using magnitude and phase of (S_{11}), respectively. I observed that the range of frequencies that have the most notable discrimination between the responses due to different conductivity values using magnitude and phase of S_{11} , are 1410-1470 MHz, which is the same as in case of permittivity.

Next step is **data-driven feature extraction**, Figures 4.44 and 4.45 show the different values of the skin conductivity representing by the highest two PCs using magnitude and phase of S_{11}, respectively.

After that, the results of domain-knowledge feature extraction category will be shown. Figures 4.46 and 4.47 show the different values of the skin conductivity representing by the minimum-two magnitude and phase of S_{11}, respectively. This representation of features shows some kinds of linearity as shown in the figures.

Figures 4.48 and 4.49 show the different values of the skin conductivity representing by the minimum frequency (resonance) and minimum magnitude and phase of S_{11}, respectively. In case of the phase of S_{11}, majority of the resonance frequency is at about

Figure 4.42: Magnitude of S_{11} for different values for Conductivity of Skin (a) Entire frequency range. (b) Frequency range of interest

Figure 4.43: Phase of S_{11} for different values for Conductivity of Skin (a) Entire frequency range. (b) Frequency range of interest

1100-1150 MHz which is not in the frequency range of interest and therefore this feature may not be a discriminating one.

Finally, I will show the results of predicting the values of the conductivity of skin tissue using those feature extracted from the S11 reflected coefficient of the spiral sensor. The goal is to find a regression function to map between the S11 features and the conductivity of the tissue. Therefore, after selecting the most relevant features, they were inputted to train the regression model which will be used to predict the skin conductivity. A total of 19 different regression models were trained using Matlab regression learning App. RMSE

72
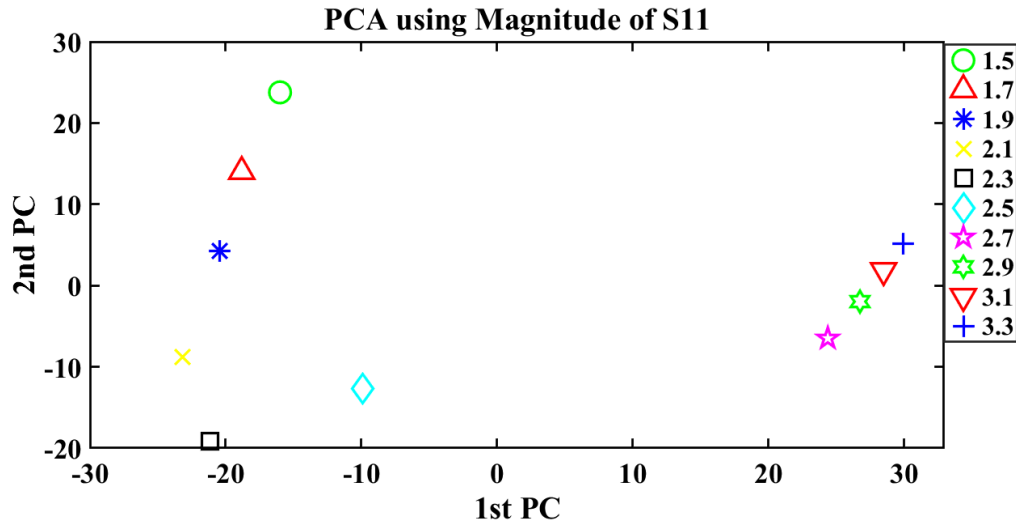
Figure 4.44: PCA using Magnitude of S_{11} with different values for Conductivity of Skin
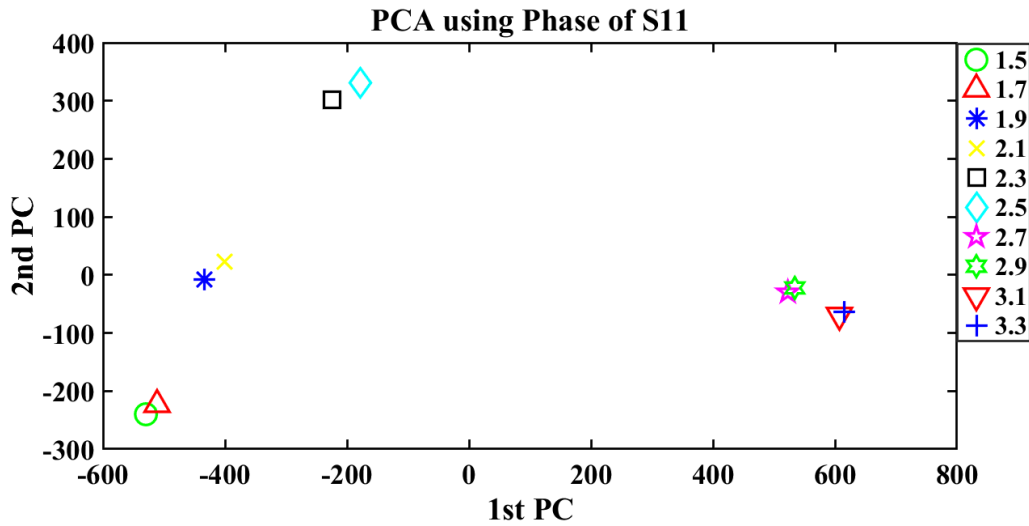
Figure 4.45: PCA using Phase of S_{11} with different values for Conductivity of Skin

and R^2 are used as the criteria to select the most accurate regression model.

Table 4.7 shows the results of only the best regression models using all the different features approaches (Minimum-two S_{11} Magnitude and Phase, and Highest-two PCs using both magnitude and phase). 3-folds cross validation was used to train and test the system.
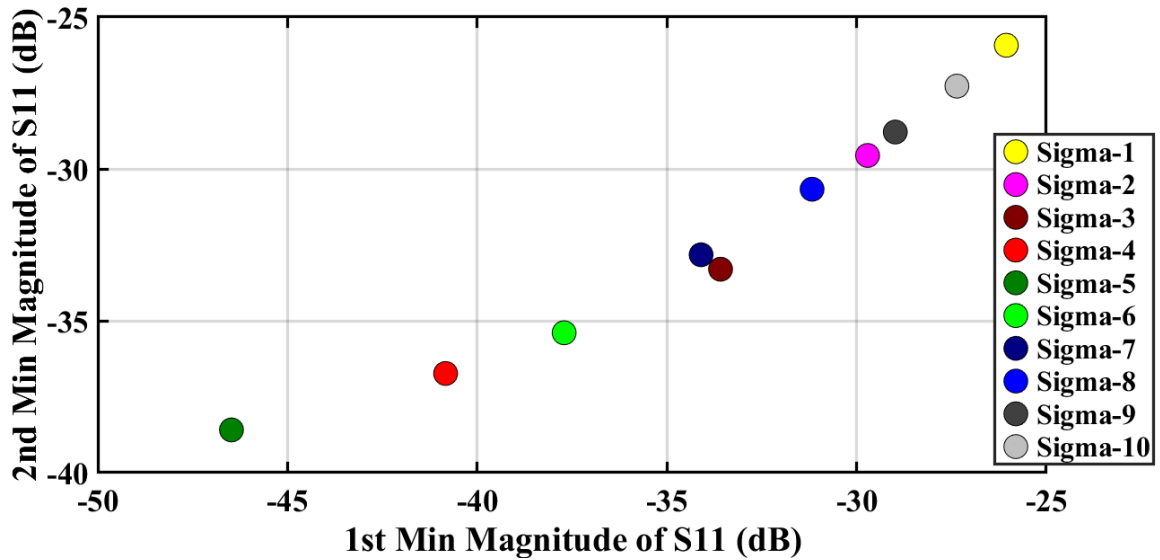
Figure 4.46: Minimum-two Magnitude of S_{11} with different values for Conductivity of Skin
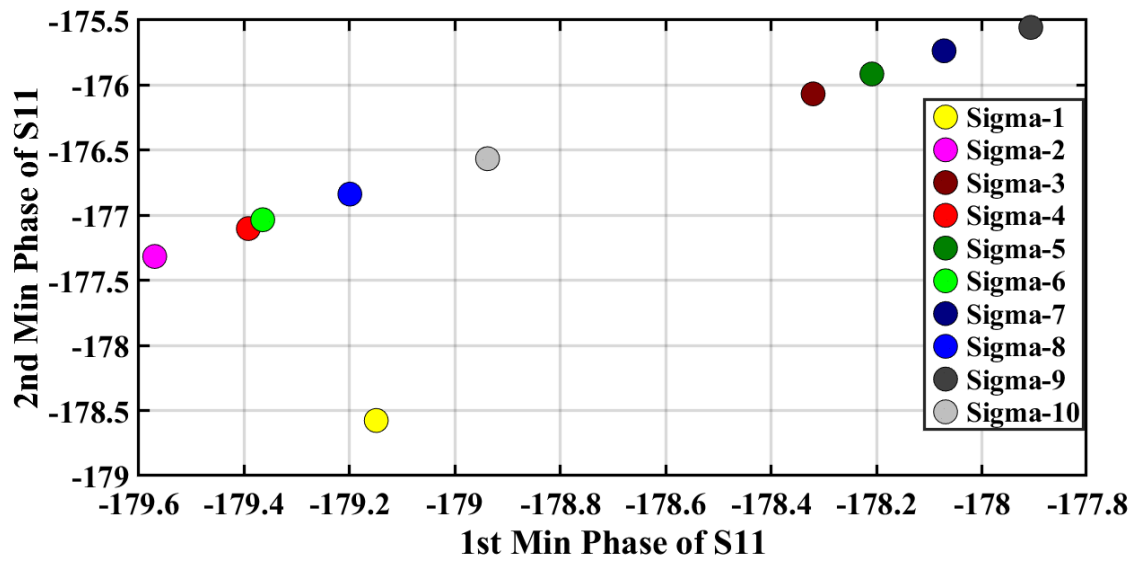
Figure 4.47: Minimum-two Phase of S_{11} with different values for Conductivity of Skin

Results from the Quadratic SVR algorithm showed the least RMSE with 0.15 and highest $R^2 = 0.94$ using PCA of Magnitude of S_{11}.
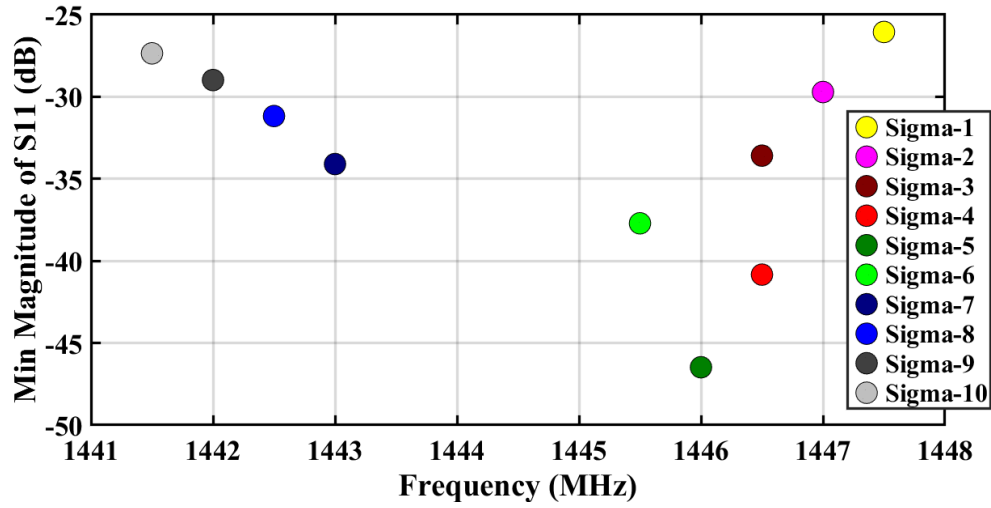
Figure 4.48: Minimum Frequency and Magnitude of S_{11} with different values for Conductivity of Skin
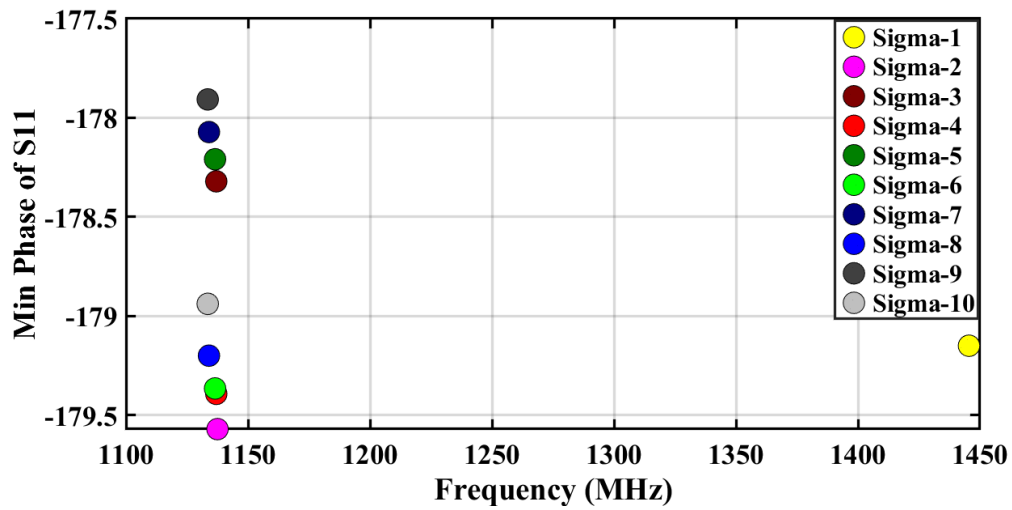
Figure 4.49: Minimum Frequency and Phase of S_{11} with different values for Conductivity of Skin

75

Table 4.7: RMSE of some of the regression algorithms using the spiral sensor with Conductivity of Skin

Features	Best Regression Algorithm	Metrics	
		RMSE	R^2
Min Mag.	Ensemble Boosted trees	0.77	0.02
Min Phase	Squared Exponential GPR	0.64	0.37
PCA Mag.	Quadratic SVR	0.15	0.94
PCA Phase	Linear Regression	0.28	0.86

4.3.5   FAT Layer

In this subsection, I will show the results of changing the permittivity and the conductivity of the FAT tissue/layer of the hand model.

Permittivity(ε)

Here, I will show the results of the system by changing the permittivity of the FAT layer. Permittivity values for the FAT layer are in the range [5.3,95.3].

In order to see how the simulation data using FAT permittivity and to have a clear picture about the nature of these data, Figures 4.50 and 4.51 show the responses of the spiral sensor to the different FAT permittivity values using magnitude and phase of (S_{11}), respectively. I observed that the range of frequencies that have the most notable discrimination between the responses due to different permittivity values using magnitude and phase of S_{11}, are 1410-1470 MHz.

In **feature extraction** following data-driven approach, Figures 4.52 and 4.53 show the different values of the permittivity representing by the highest two PCs using magnitude and phase of S_{11}, respectively. As we can see from the figures, PCA using magnitude of S11 gave us some good patterns that may help in the next step of regression.

Next, I will show the results of the 2nd category of feature engineering, which is based on domain-knowledge. Figures 4.54 and 4.55 show the different values of the permittivity representing by the minimum-two magnitude and phase of S_{11}, respectively. It is very clear that this representation of features produced some interesting patterns as shown in the figures.
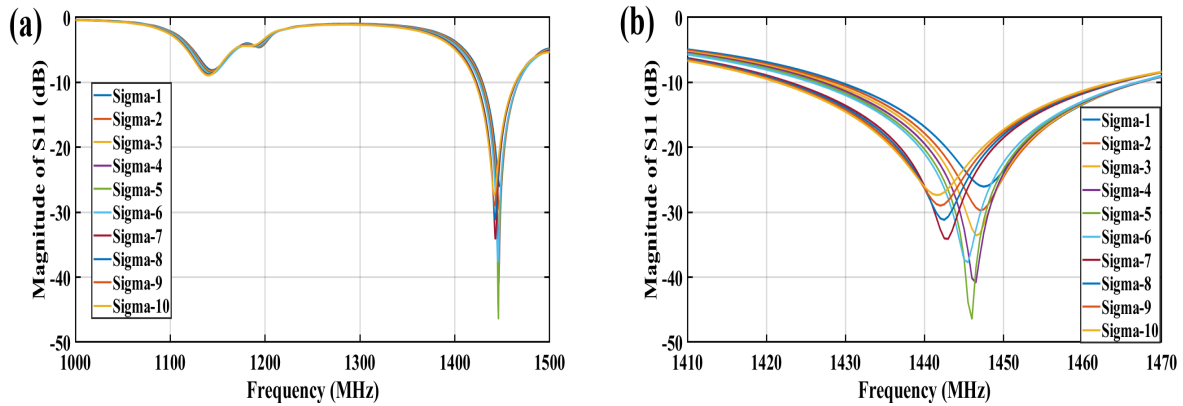
Figure 4.50: Magnitude of S_{11} for different values for Permittivity of FAT (a) Entire frequency range. (b) Frequency range of interest
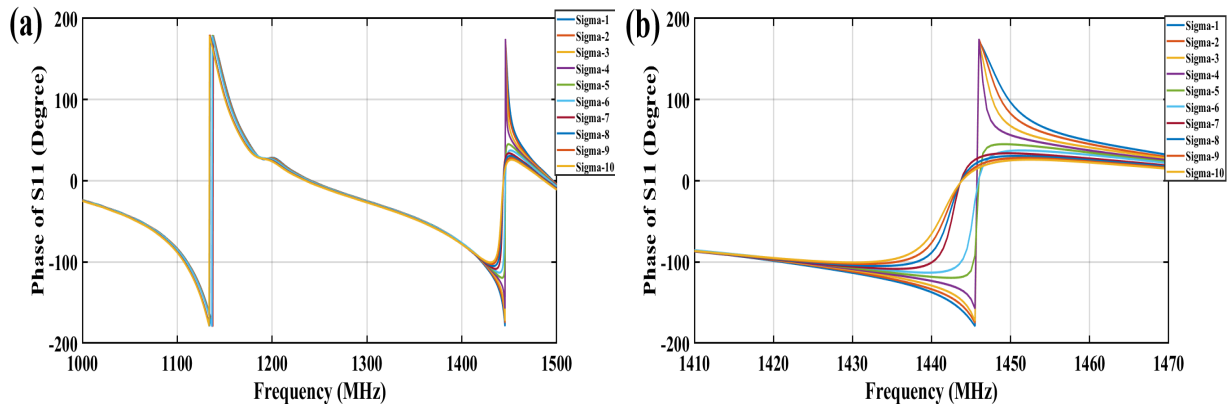

Figure 4.51: Phase of S_{11} for different values for Permittivity of FAT (a) Entire frequency range. (b) Frequency range of interest

Figures 4.56 and 4.57 show the different values of the permittivity representing by the minimum frequency (resonance) and minimum S_{11} magnitude and phase, respectively. In case of the phase of S_{11}, the resonance frequency is at about 1132-1135 MHz which is not in the frequency range of interest and therefore this feature may not be a discriminating one.

Now, I reached to the final step of the system which is predicting the values of the permittivity of FAT tissue using those feature extracted from the S11 reflected coefficient of

77

Figure 4.52: PCA using Magnitude of S_{11} with different values for Permittivity of FAT
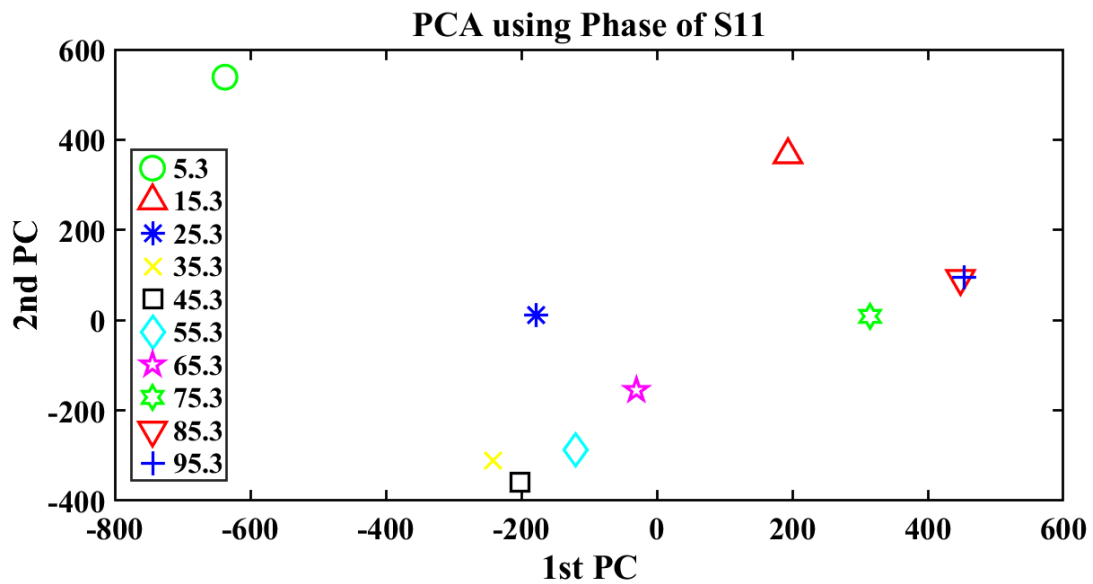
Figure 4.53: PCA using Phase of S_{11} with different values for Permittivity of FAT

the spiral sensor. The goal is to find a regression function to map between the S11 features and the permittivity of the tissue. Therefore, after selecting the most relevant features,

Figure 4.54: Minimum-two Magnitude of S_{11} with different values for Permittivity of FAT
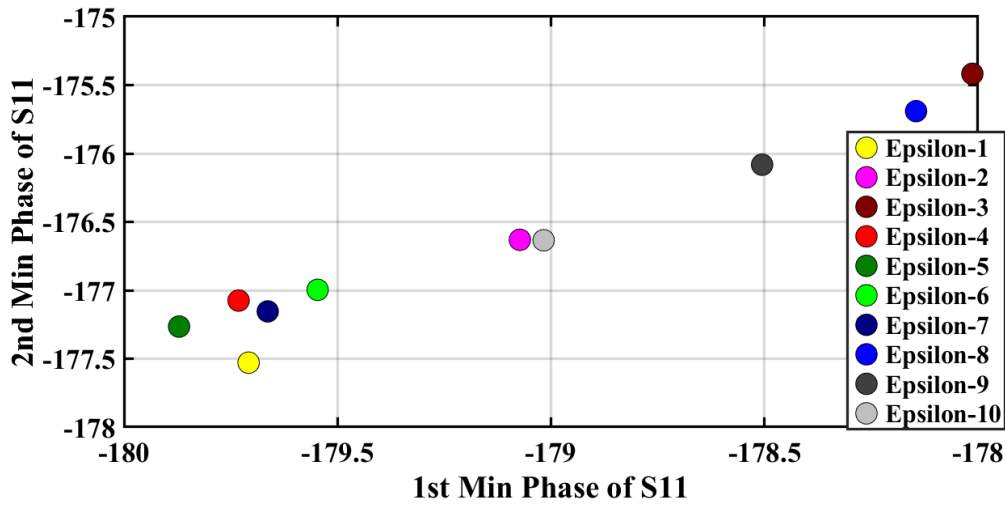
Figure 4.55: Minimum-two Phase of S_{11} with different values for Permittivity of FAT

they were inputted to train the regression model which will be used to predict the FAT permittivity. A total of 19 different regression models were trained using Matlab regression learning App. RMSE is used as the criteria to select the most accurate regression model. Table 4.8 shows the results of the best regression models using all the different features approaches (Minimum-two S_{11} Magnitude and Phase , and Highest-two PCs using both

79

Figure 4.56: Minimum Frequency and Magnitude of S_{11} with different values for Permittivity of FAT
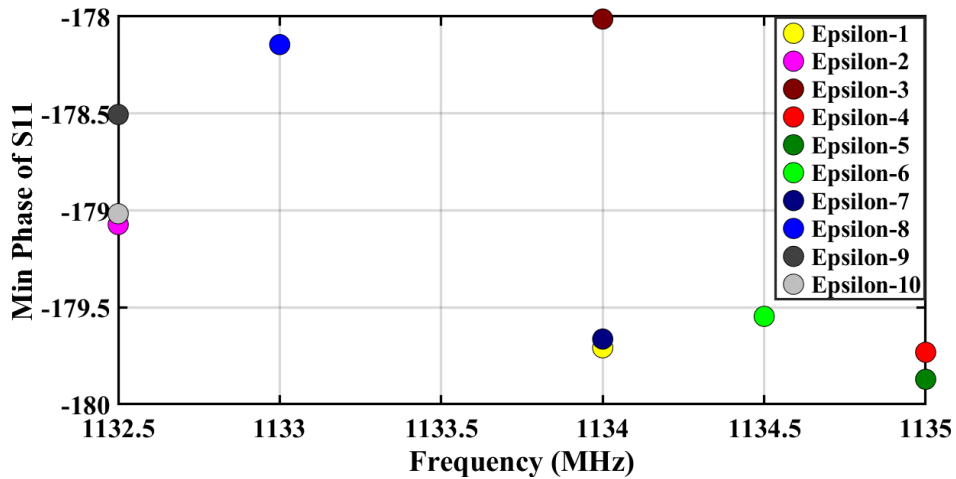
Figure 4.57: Minimum Frequency and Phase of S_{11} with different values for Permittivity of FAT

magnitude and phase). I used 3-folds cross validation to validate the system.

Results from Linear Regression algorithm showed the least RMSE with 7.1 using PCA of Magnitude of S_{11}.

Table 4.8: RMSE of some of the regression algorithms using the spiral sensor with Permittivity of FAT

Features	Best Regression Algorithm	RMSE
Min Mag.	Quadratic SVR	12.3
Min Phase	Decision Tree	28.8
PCA Mag.	Linear Regression	7.1
PCA Phase	Linear Regression	9.7

Conductivity (σ)

In this subsection, I will show the results of the system by changing the conductivity of the FAT layer. Conductivity values for the FAT layer are in the range $[0.02, 1.4]$, values are adopted from [118] for the corresponding frequency range of the spiral sensor (1000-1500 MHz) in order to reflect the expected change in conductivity caused by glucose molecules in FAT tissue.

In order to see how the simulation data looks like and to have a clear picture about the nature of these data, Figures 4.58 and 4.59 show the responses of the spiral sensor to the different conductivity values using magnitude and phase of (S_{11}), respectively. I observed that the range of frequencies that have the most notable discrimination between the responses due to different conductivity values using magnitude and phase of S_{11} , are 1410-1470 MHz, which is the same as in case of permittivity.
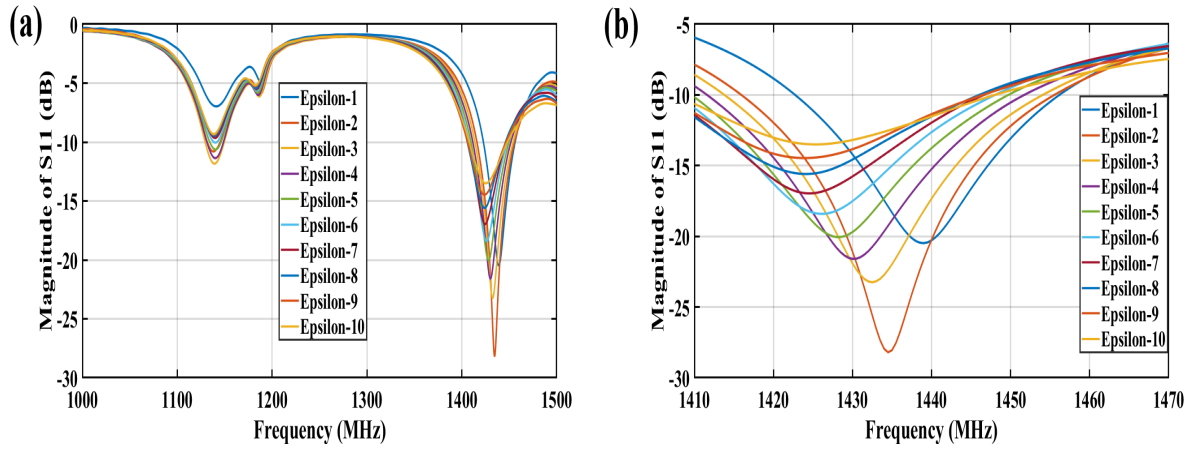

Figure 4.58: Magnitude of S_{11} for different values for Conductivity of FAT (a) Entire frequency range. (b) Frequency range of interest
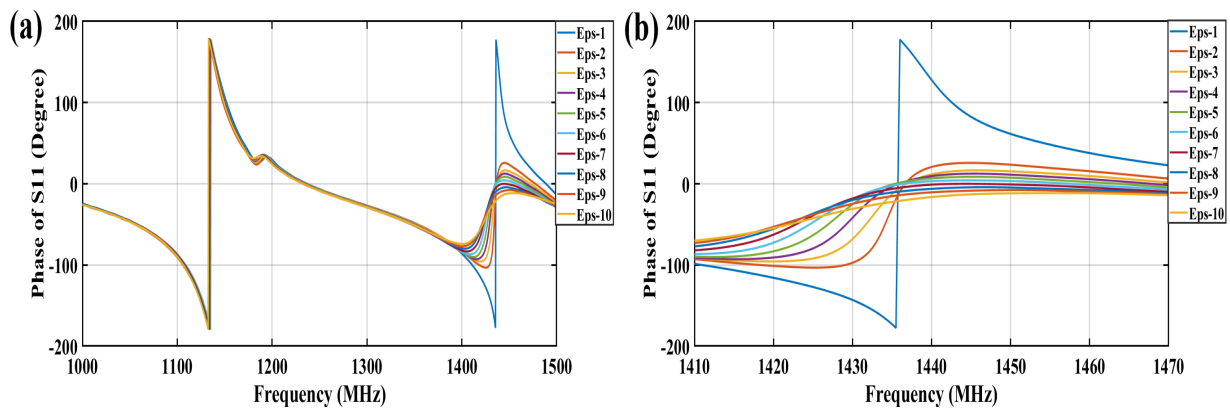
Figure 4.59: Phase of S_{11} for different values for Conductivity of FAT (a) Entire frequency range. (b) Frequency range of interest

Next step is **feature extraction** following data-driven approach, Figures 4.60 and 4.61 show the different values of the conductivity representing by the highest two PCs using magnitude and phase of S_{11}, respectively.

After that, the results of domain-knowledge feature extraction category will be shown. Figures 4.62 and 4.63 show the different values of the conductivity representing by the minimum-two magnitude and phase of S_{11}, respectively. This representation of features shows some kinds of linearity as shown in the figures.

Figures 4.64 and 4.65 show the different values of the conductivity representing by the minimum frequency (resonance) and minimum S_{11} magnitude and phase, respectively. In case of the phase of S_{11}, the resonance frequency is at about 1134-1136 MHz which is not in the frequency range of interest.

Next, I will show the results of predicting the values of the conductivity of FAT tissue using those feature extracted from the S11 reflected coefficient of the spiral sensor. The goal is to find a regression function to map between the S11 features and the conductivity of the tissue. Therefore, after selecting the most relevant features, they were inputted to train the regression model which will be used to predict the FAT conductivity. A total of 19 different regression models were trained using Matlab regression learning App. RMSE and R^2 are used as the criteria to select the most accurate regression model.
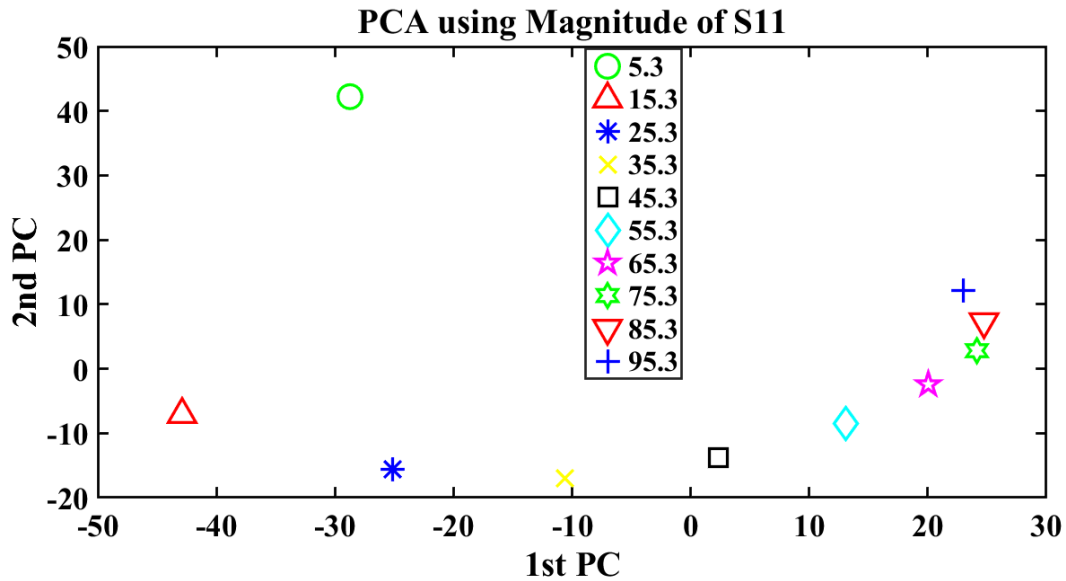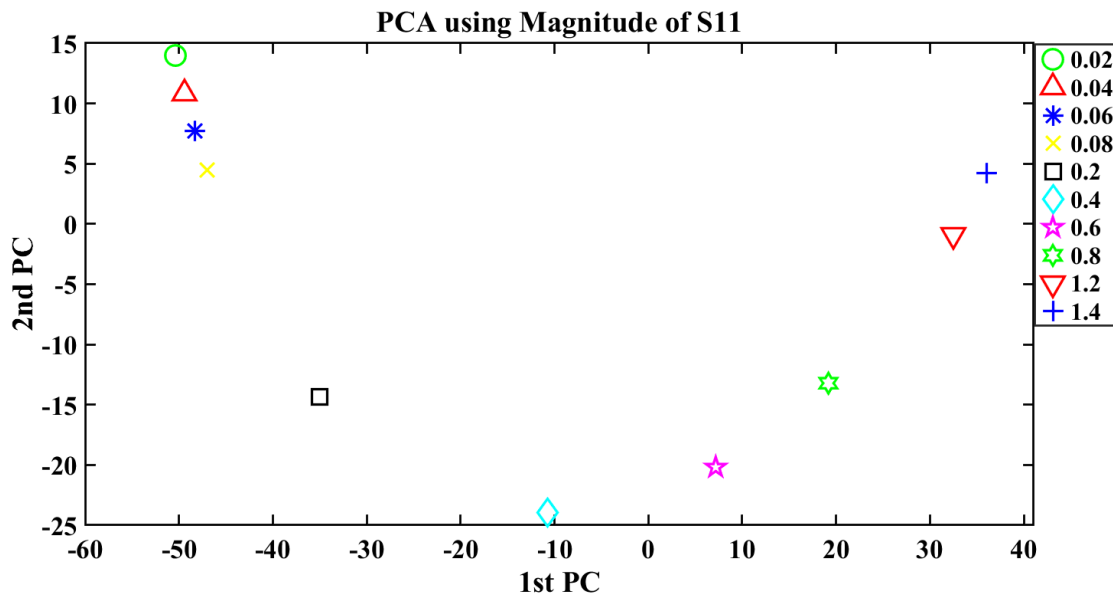
Figure 4.60: PCA using Magnitude of S_{11} with different values for Conductivity of FAT
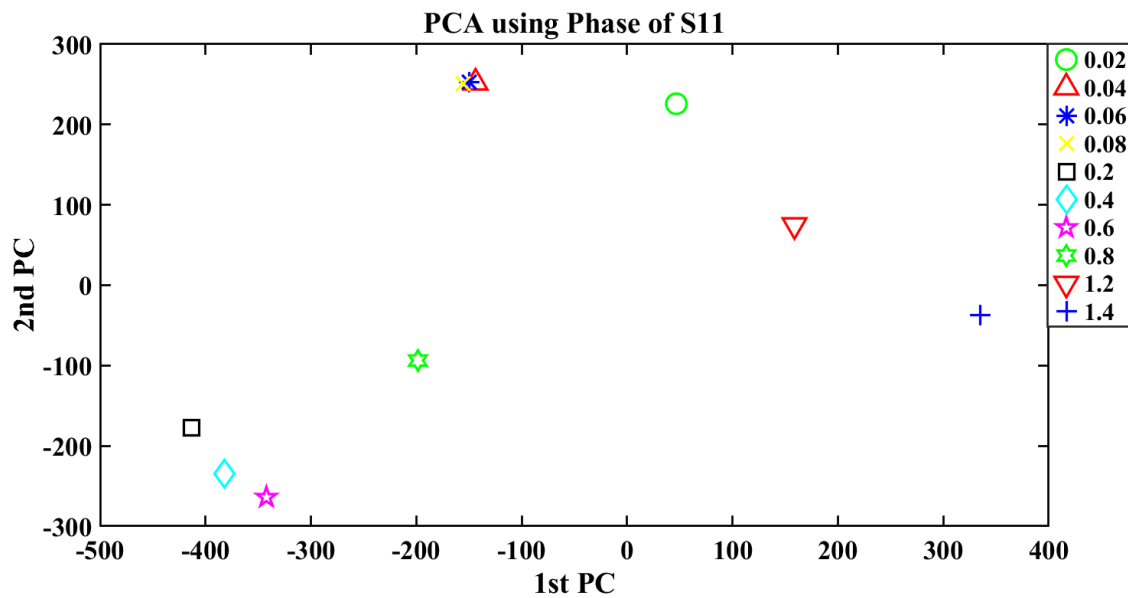
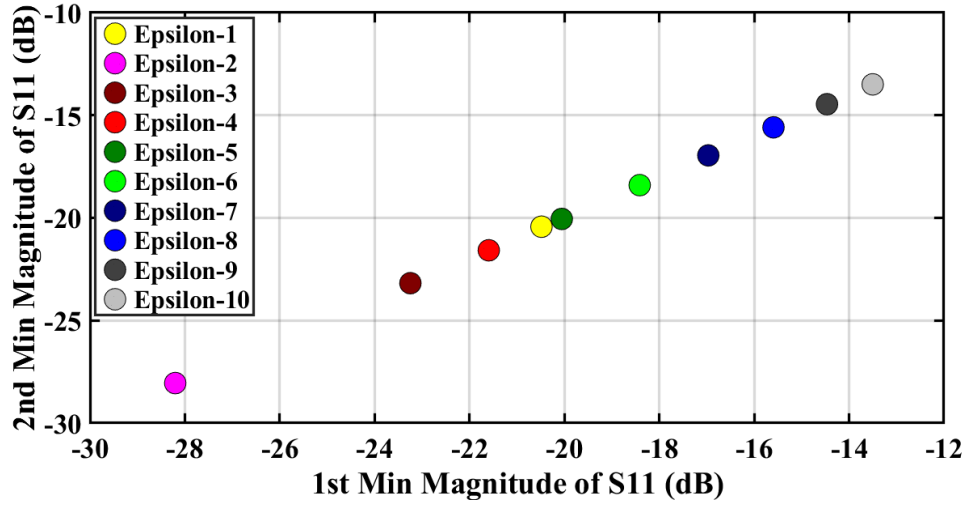Figure 4.61: PCA using Phase of S_{11} with different values for Conductivity of FAt

Figure 4.62: Minimum-two Magnitude of S_{11} with different values for Conductivity of FAT
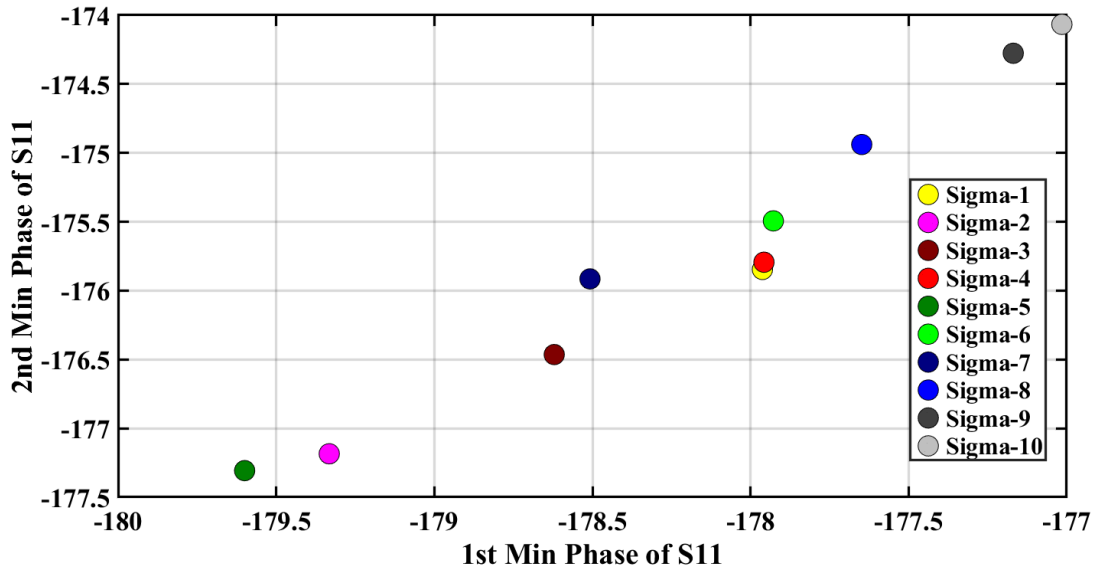
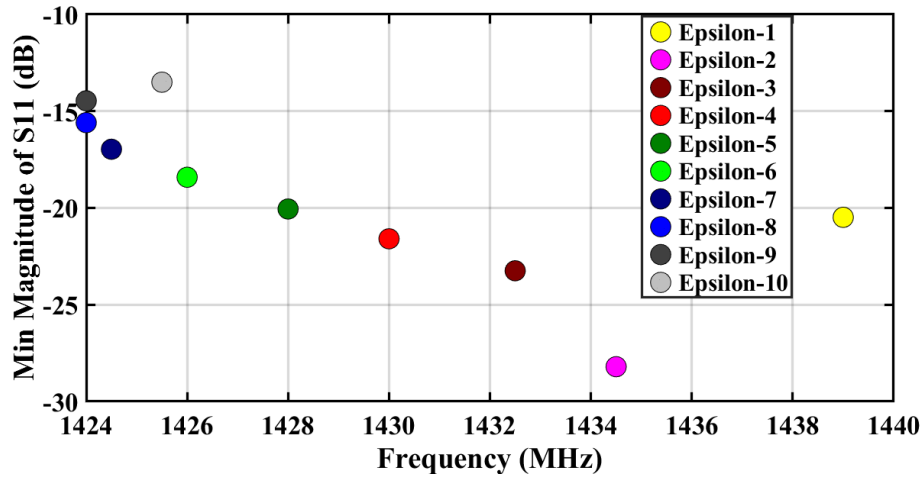Figure 4.63: Minimum-two Phase of S_{11} with different values for Conductivity of FAT

Figure 4.64: Minimum Frequency and Magnitude of S_{11} with different values for Conductivity of FAT
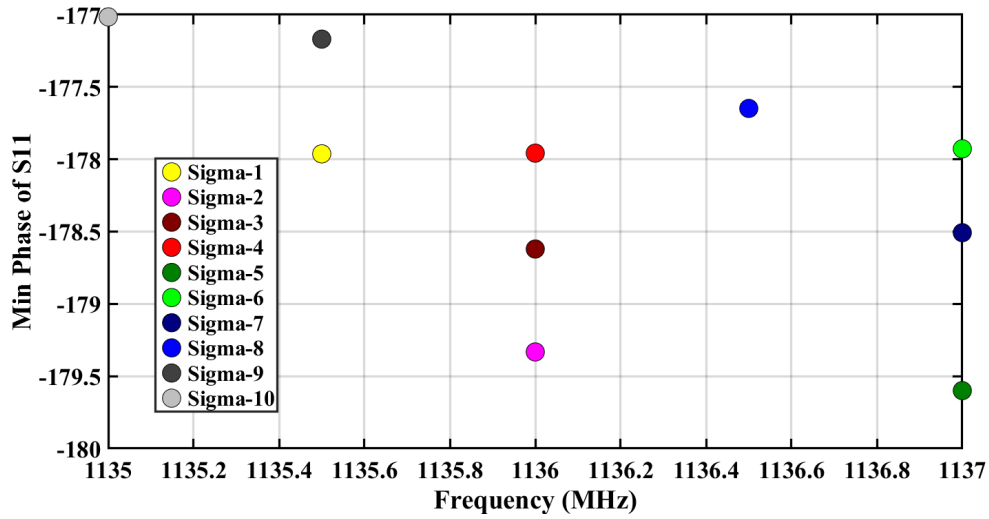
Figure 4.65: Minimum Frequency and Phase of S_{11} with different values for Conductivity of FAT

Table 4.9 shows the results of only the best regression models using all the different features approaches (Minimum-two S_{11} Magnitude and Phase, and Highest-two PCs using

85

both magnitude and phase). 3-folds cross validation was used to train and test the system.

Table 4.9: RMSE of some of the regression algorithms using the spiral sensor with Conductivity of FAT

Features	Best Regression Algorithm	Metrics	
		RMSE	R^2
Min Mag.	Quadratic SVR	0.2	0.83
Min Phase	Linear Regression	0.1	0.96
PCA Mag.	Rational Quadratic GPR	0.03	1
PCA Phase	Linear SVR	0.2	0.81

Results from the Rational Quadratic GPR algorithm showed the least RMSE ever with 0.03 and highest $R^2 = 1$ using PCA of magnitude of S_{11}.

4.3.6 Muscle Layer

In this subsection, I will show the results of changing the permittivity and the conductivity of the Muscle tissue/layer of the hand model.

Permittivity(ε)

I will show the results of the system by changing the permittivity of the Muscle layer. Permittivity values for the Muscle layer are in the range [52.7,142.7], where the permittivity values for the muscle from [118] are between 53-54 for the corresponding frequency range of the spiral sensor (1000-1500 MHz).

In order to see how the simulation data looks like and to have a clear picture about the nature of these data, Figures 4.66 and 4.67 show the responses of the spiral sensor to the different permittivity values using magnitude and phase of (S_{11}), respectively. I observed that the range of frequencies that have the most notable discrimination between the responses due to different permittivity values using magnitude and phase of S_{11}, are 1410-1470 MHz.
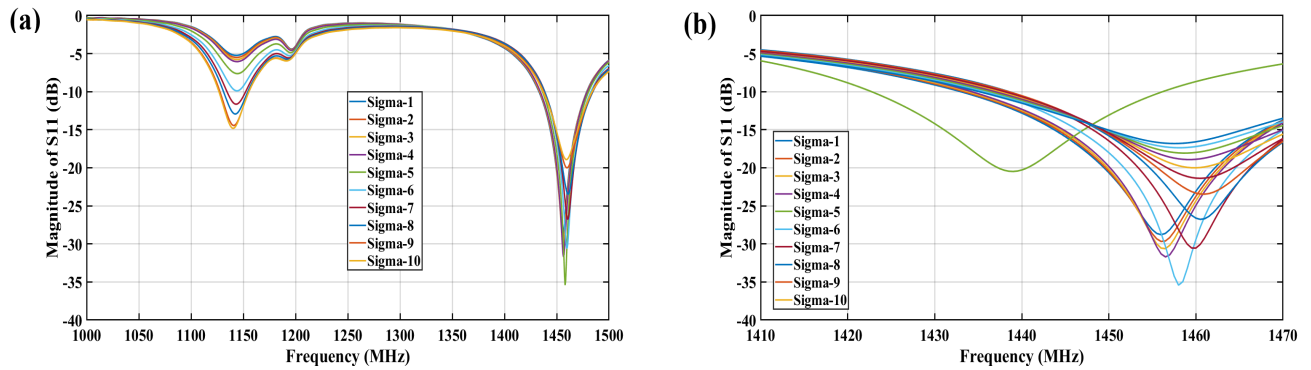
Figure 4.66: Magnitude of S_{11} for different values for Permittivity of Muscle (a) Entire frequency range. (b) Frequency range of interest
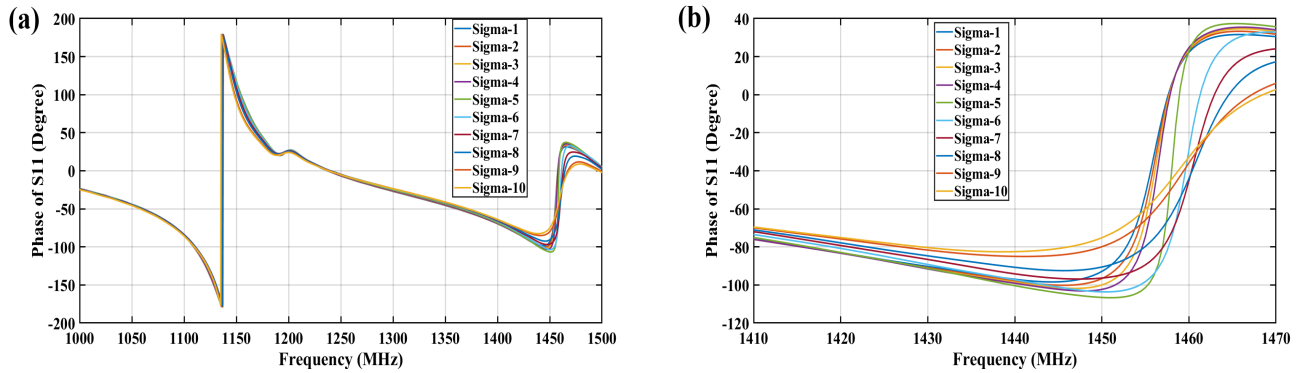

Figure 4.67: Phase of S_{11} for different values for Permittivity of Muscle (a) Entire frequency range. (b) Frequency range of interest

In **Data-Driven feature extraction**, Figures 4.68 and 4.69 show the different values of the permittivity representing by the highest two PCs using magnitude and phase of S_{11}, respectively.

Next, I will show the results of the 2nd category of feature engineering, which is based on domain-knowledge. Figures 4.70 and 4.71 show the different values of muscle permittivity representing by the minimum-two magnitude and phase of S_{11}, respectively. It is very clear that this representation of features produced some interesting patterns as shown in the figures.

Figures 4.72 and 4.73 show the different values of the muscle permittivity representing by the minimum frequency (resonance) and minimum magnitude and phase of S_{11}, respec-

Figure 4.68: PCA using Magnitude of S_{11} with different values for Permittivity of Muscle
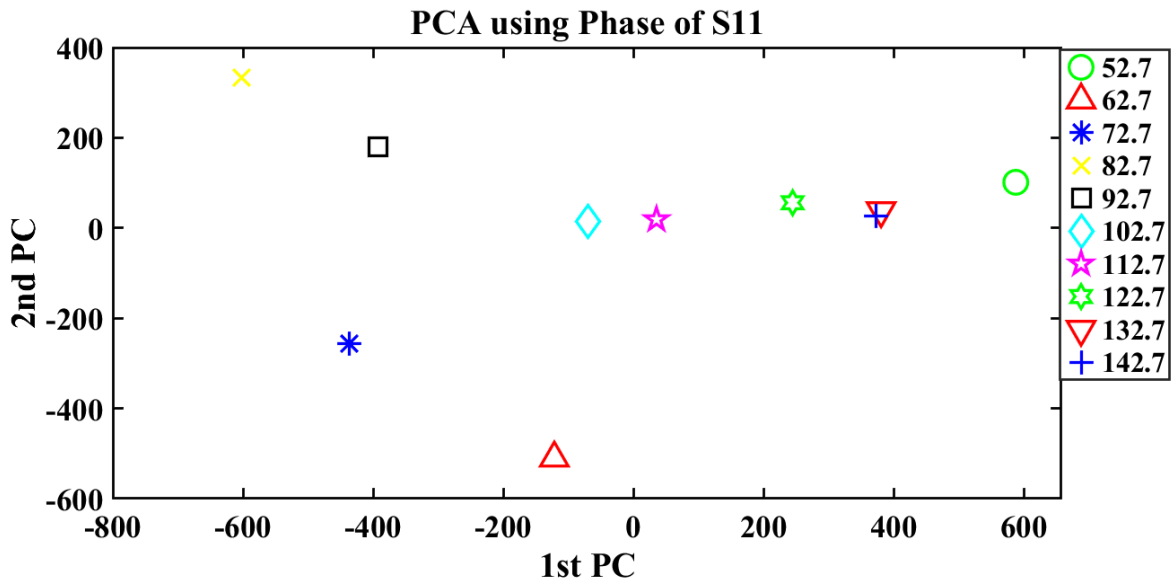
Figure 4.69: PCA using Phase of S_{11} with different values for Permittivity of Muscle

tively. In case of the phase of S_{11}, the resonance frequency is at about 1130-1145 MHz which is not in the frequency range of interest.
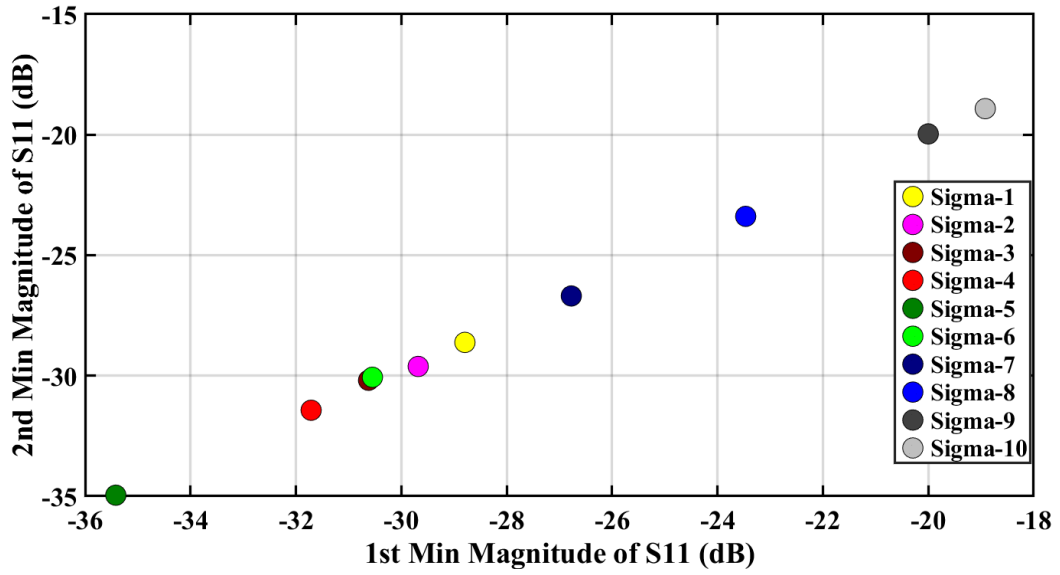
Figure 4.70: Minimum-two Magnitude of S_{11} with different values for Permittivity of Muscle
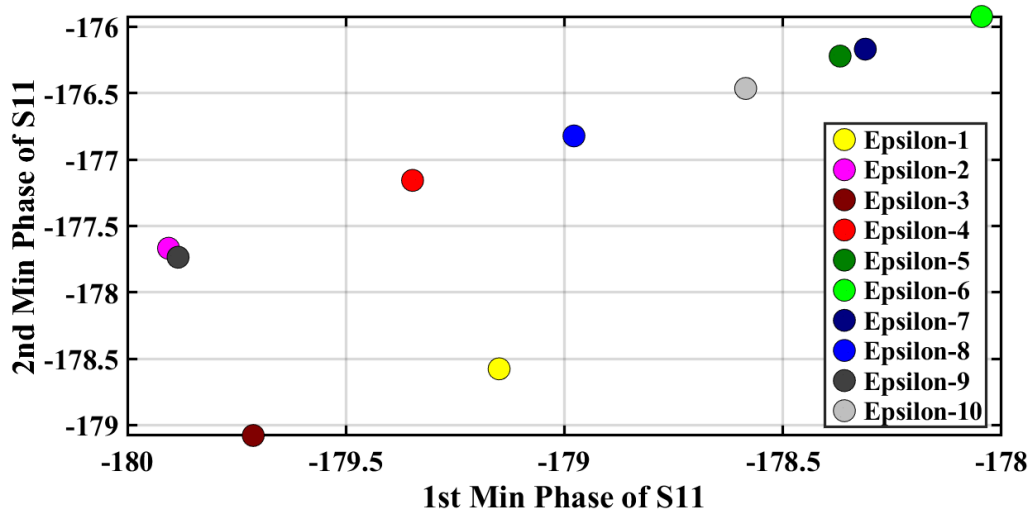
Figure 4.71: Minimum-two Phase of S_{11} with different values for Permittivity of Muscle

Now, I reached to the final step of the system which is predicting the values of the permittivity of muscle tissue using those feature extracted from the S11 reflected coefficient of the spiral sensor. The goal is to find a regression function to map between the S11
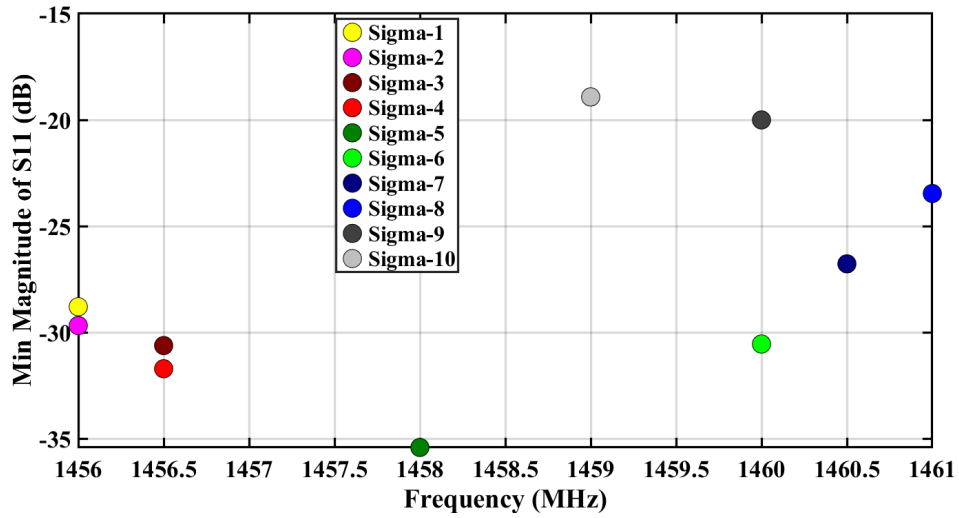
Figure 4.72: Minimum Frequency and Magnitude of S_{11} with different values for Permittivity of Muscle
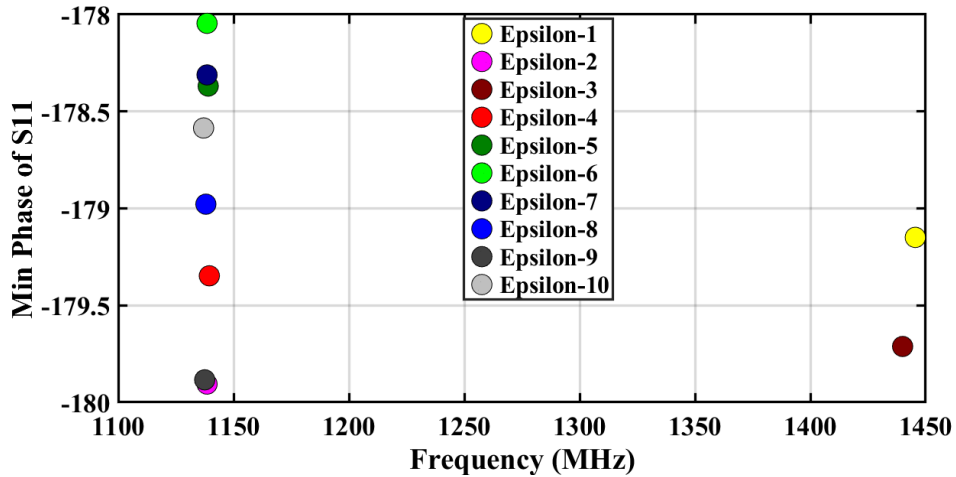
Figure 4.73: Minimum Frequency and Phase of S_{11} with different values for Permittivity of Muscle

features and the permittivity of the tissue. Therefore, after selecting the most relevant features, they were inputted to train the regression model which will be used to predict the muscle permittivity. A total of 19 different regression models were trained using Matlab regression learning App. RMSE and R^2 are used as the criteria to select the most accurate regression model. Table 4.10 shows the results of the best regression models using all the

different features approaches (Minimum-two S_{11} Magnitude and Phase , and Highest-two PCs using both magnitude and phase). I used 3-folds cross validation to validate the system.

Table 4.10: RMSE of some of the regression algorithms using the spiral sensor with Permittivity of Muscle

Features	Best Regression Algorithm	Metrics	
		RMSE	R^2
Min Mag.	Linear Regression	27	0.42
Min Phase	Decision Tree	30.1	0.0
PCA Mag.	Quadratic SVR	6.8	0.95
PCA Phase	Fine Gaussian SVR	27.6	0.45

Results from the Rational Quadratic SVR algorithm showed the least RMSE with 6.8 and highest $R^2 = 0.95$ using PCA of Magnitude of S_{11}.

Conductivity (σ)

Now I will show the results of the system by changing the conductivity of the muscle layer. Conductivity values for the muscle layer are in the range [1.74,3.54], values are adopted from [118] for the corresponding frequency range of the spiral sensor (1000-1500 MHz) in order to reflect the expected change in conductivity caused by glucose molecules in the skin.

In order to see how the simulation data using the muscle conductivity and to have a clear picture about the nature of these data, Figures 4.74 and 4.75 show the responses of the spiral sensor to the different conductivity values using magnitude and phase of (S_{11}), respectively. I observed that the range of frequencies that have the most notable discrimination between the responses due to different conductivity values using magnitude and phase of S_{11} , are 1410-1470 MHz, which is the same as in case of permittivity.
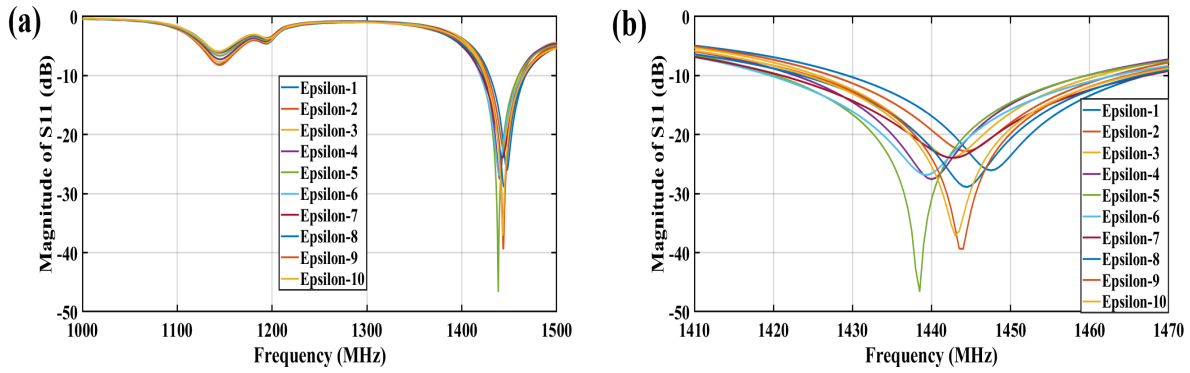
Next step is **data-driven feature extraction**, Figures 4.76 and 4.77 show the different values of the muscle conductivity representing by the highest two PCs using magnitude and phase of S_{11}, respectively.
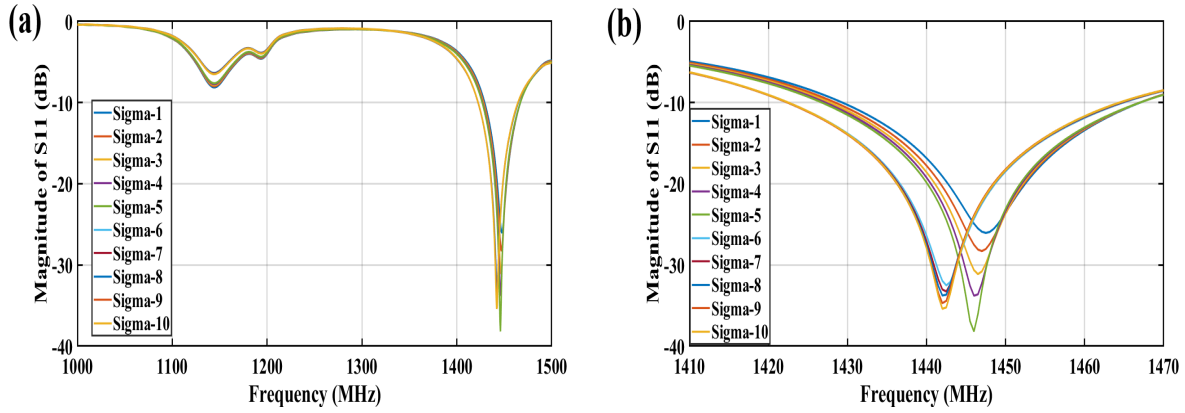
Figure 4.74: Magnitude of S_{11} for different values for Conductivity of Muscle (a) Entire frequency range. (b) Frequency range of interest
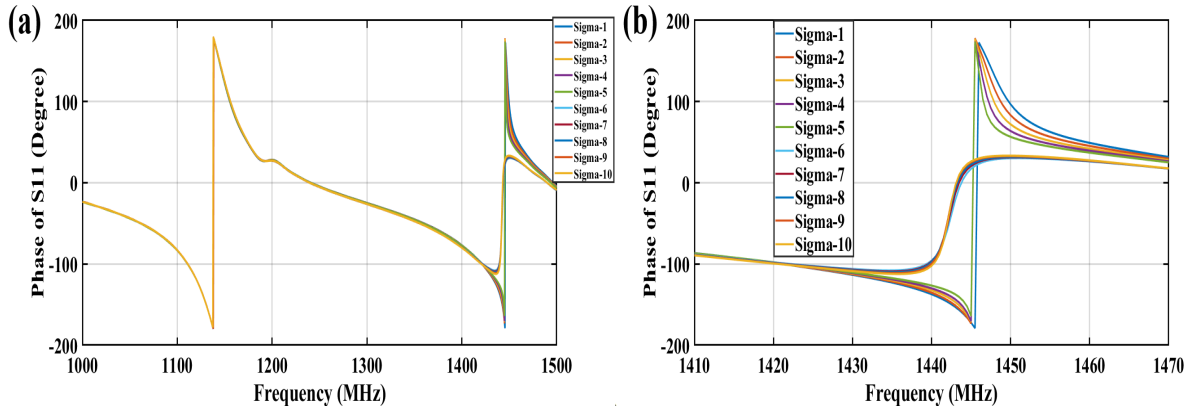
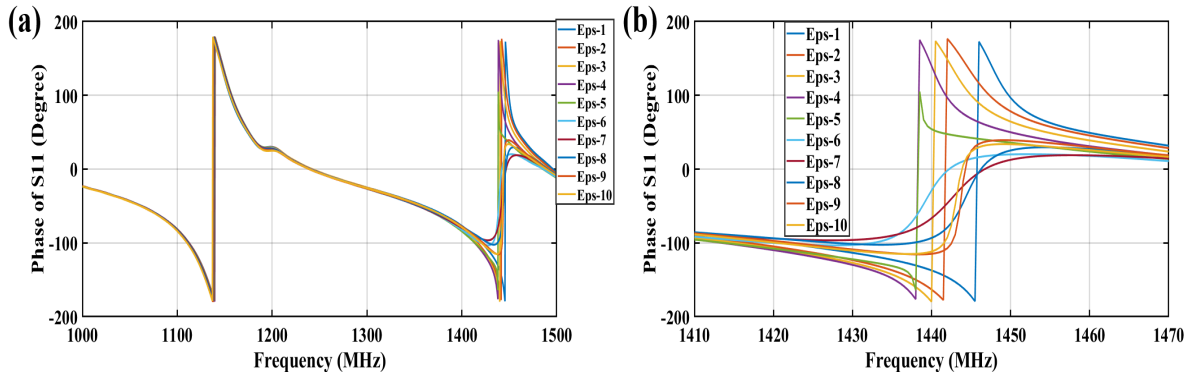Figure 4.75: Phase of S_{11} for different values for Conductivity of Muscle (a) Entire frequency range. (b) Frequency range of interest

After that, the results of domain-knowledge feature extraction category will be shown. Figures 4.78 and 4.79 show the different values of the skin conductivity representing by the minimum-two magnitude and phase of S_{11}, respectively. This representation of features shows some kinds of linearity as shown in the figures.

Figures 4.80 and 4.81 show the different values of the muscle conductivity representing by the minimum frequency (resonance) an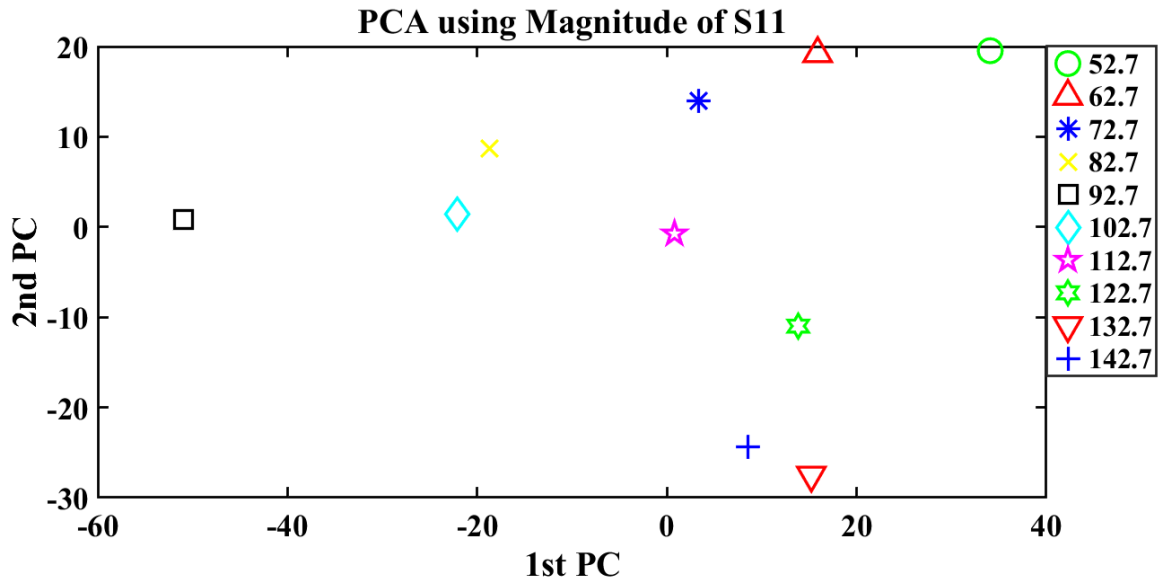d minimum magnitude and phase of S_{11}, respectively. In case of the phase of S_{11}, majority of the resonance frequency is at about 1100-1150 MHz which is not in the frequency range of interest and therefore this feature
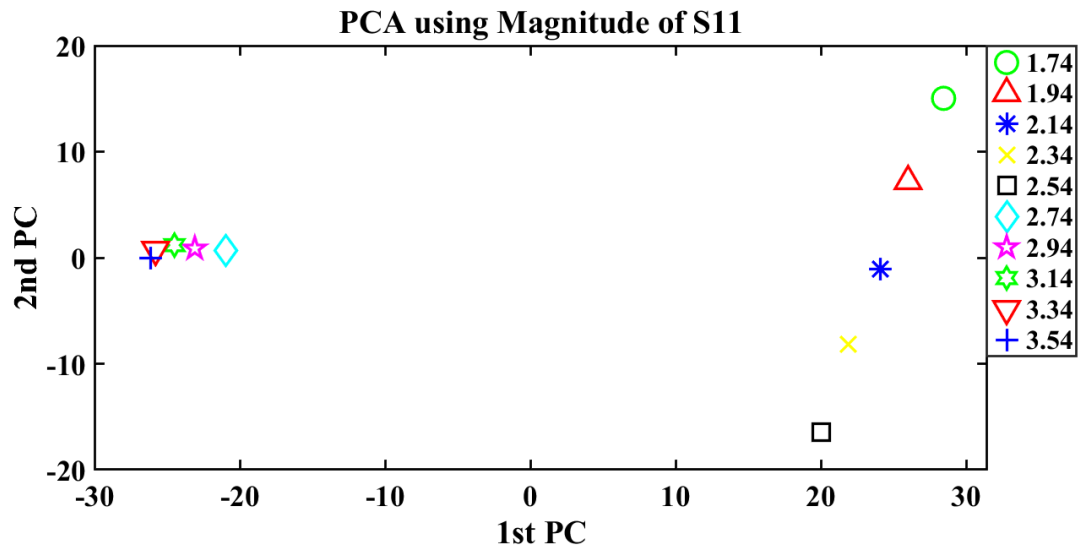
Figure 4.76: PCA using Magnitude of S_{11} with different values for Conductivity of Muscle
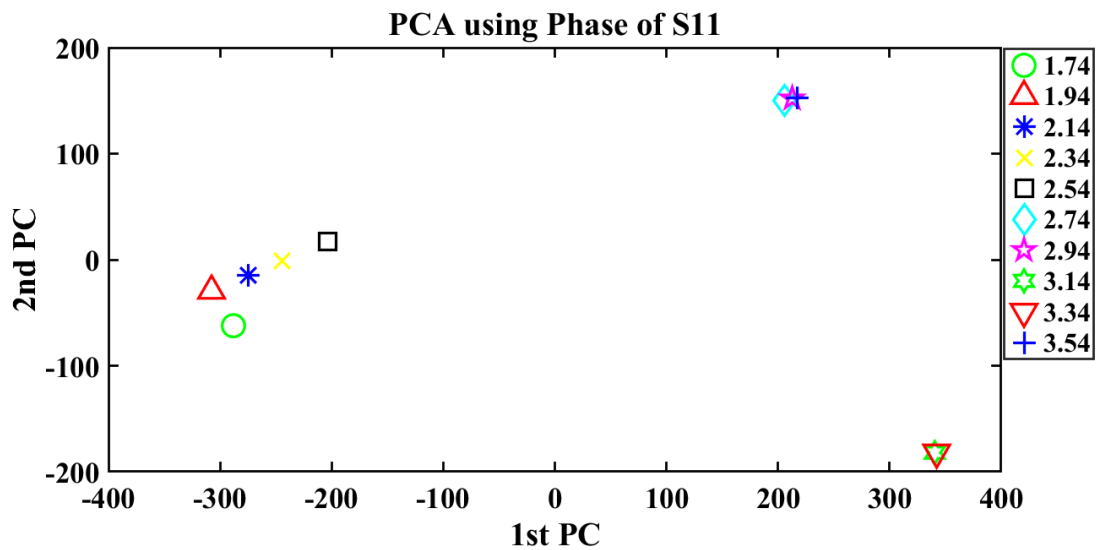
Figure 4.77: PCA using Phase of S_{11} with different values for Conductivity of Muscle

may not be a discriminating one.

Finally, I will show the results of predicting the values of the conductivity of muscle tissue using those feature extracted from the S11 reflected coefficient of the spiral sensor. The
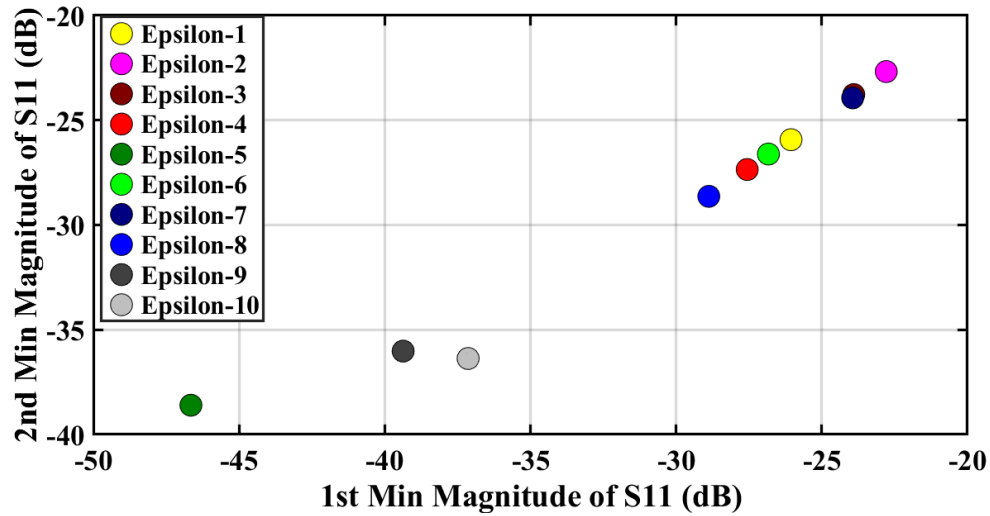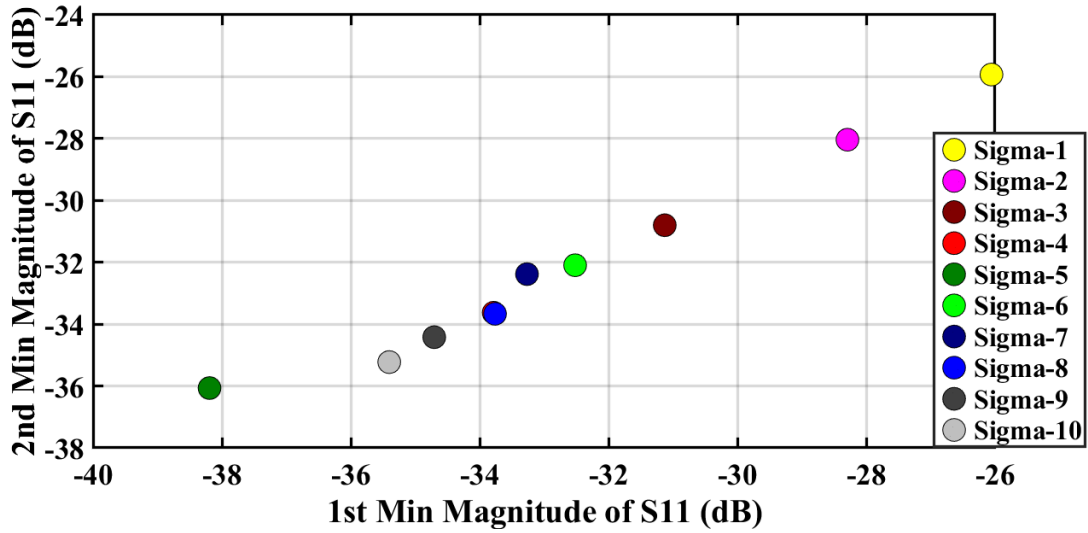
Figure 4.78: Minimum-two Magnitude of S_{11} with different values for Conductivity of Muscle
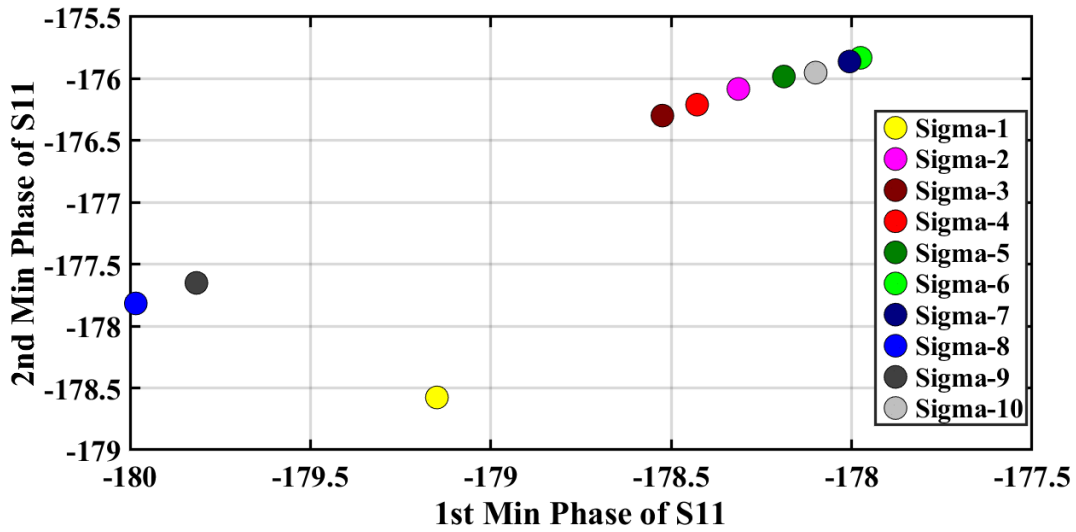
Figure 4.79: Minimum-two Phase of S_{11} with different values for Conductivity of Muscle

goal is to find a regression function to map between the S11 features and the conductivity of the tissue. Therefore, after selecting the most relevant features, they were inputted to
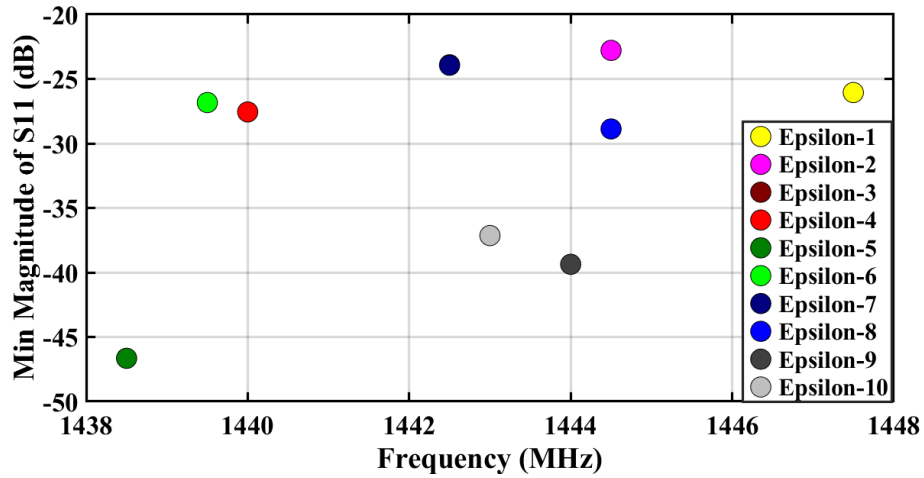
94
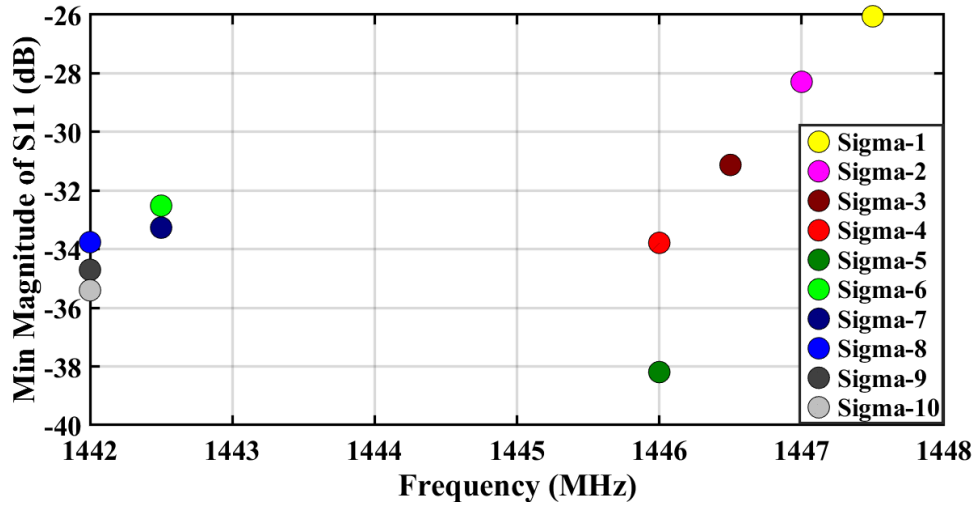
Figure 4.80: Minimum Frequency and Magnitude of S_{11} with different values for Conductivity of Muscle
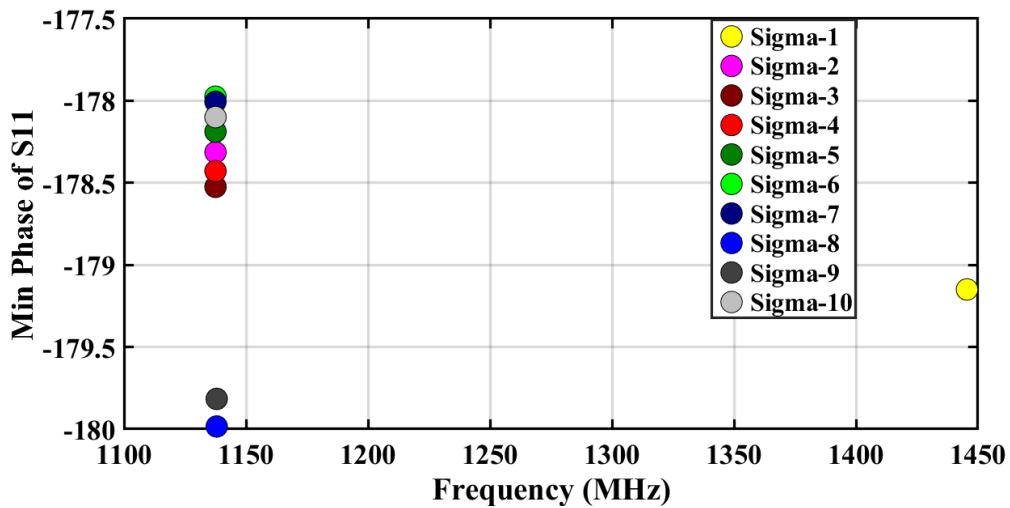
Figure 4.81: Minimum Frequency and Phase of S_{11} with different values for Conductivity of Muscle

train the regression model which will be used to predict the blood conductivity. A total of 19 different regression models were trained using Matlab regression learning App. RMSE and R^2 are used as the criteria to select the most accurate regression model.

Table 4.11 shows the results of only the best regression models using all the different features approaches (Minimum-two S_{11} Magnitude and Phase, and Highest-two PCs using both magnitude and phase). 3-folds cross validation was used to train and test the system.

Table 4.11: RMSE of some of the regression algorithms using the spiral sensor with Conductivity of Muscle

Features	Best Regression Algorithm	Metrics	
		RMSE	R^2
Min Mag.	Median Gaussian SVR	0.6	0.2
Min Phase	Fine Gaussian SVR	0.5	0.43
PCA Mag.	Linear Regression	0.24	0.94
PCA Phase	Linear SVR	0.3	0.75

Results from the Linear Regression algorithm showed the least RMSE with 0.24 and highest $R^2 = 0.94$ using PCA of magnitude of S_{11}.

4.3.7 Mapping glucose with dielectric properties and testing of Regression Systems

In this subsection, I will show the result of testing the regression system, trained using specific layer/tissue, with the values of other layers/tissues. The goal is to use those trained regression systems to predict the actual glucose levels after finding a model for dielectric properties as a function of glucose concentration in the human body.

After training the system using values of permittivity for blood (in the range [40,105]), FAT permittivity values are used to test the prediction accuracy of the trained system. The test values was 21 values in the range [4,6.5]. I tested those 21 values using the trained QSVR as in 4.4. Test RMSE was about 60, however, the ranges of the permittivity values used for training and testing have big difference.

After using the following equation 4.3 to map the output of the test values into their corresponding interval, I got RMSE $= 0.7$

$$f(t) = c + ((d - c)/(b - a)) * (t - a) \tag{4.3}$$

, where t is the value to be mapped from interval [a,b] to [c,d].

In the same way, I used the trained regression system for conductivity of FAT layer to test new values of conductivity for all the layers. Table 4.12 shows the RMSE results.

Table 4.12: Test RMSE of all the layers using the trained regression system for conductivity of FAT layer

Features	FAT	Blood	Skin	Muscle
Min Mag.	0.05	0.4	2.1	2.5
Min Phase	0.6	0.6	2.4	2.8
PCA Mag.	0.08	0.5	1.7	2.1
PCA Phase	0.3	1.3	1.8	1.9

When I build the relation between human body glucose concentrations and dielectric properties (permittivity or conductivity), I just need to map those glucose concentrations to the trained system in order for the system to predict those glucose concentrations. Finally, by the help of glucose ground truth created using the best glucometers and the eight trained regression systems using the dielectric properties (permittivity or conductivity) of the layers (Skin, Fat, Blood and Muscle), I will find a way to predict glucose concentrations in the human body by means of the microwave sensor reflected coefficient S11.

4.4 Conclusion

In this chapter, I showed the experimental and simulation results of my research. In the beginning, experimental results were shown starting from preparing the water-glucose solutions which were used in the experiments using both microwave sensors, the dipole and the spiral. Next, the results of the detailed steps of the research methodology were investigates for every sensor separately. In dipole, results from the Gaussian SVR algorithm showed the least RMSE with 11.9 using the two-minimum magnitude of S_{11}. whereas, in spiral, results from the linear regression algorithm showed the least RMSE with 7.26 using PCA of magnitude of S_{11}. After that, simulation results were shown starting from designing the hand model with four different layers in order to change the dielectric properties of those layers and then calculate the sensor's reflected signals due to the interaction of the microwave sensor with the tissues having those dielectric properties. Results of each layers were presented individually by changing of both permittivity and conductivity of each layer. Following the same research methodology each time, starting from reading sensor's signals and extracting the discriminating features from those raw signals to be inputted to the step of building the regression system, I have got reasonable RMSE. At the end, I have

eight different system, two per each layer of the four layers of the hand, one for permittivity and one for conductivity. Those systems will help in predicting the actual glucose levels after building a model mapping dielectric properties of the human body with the glucose levels.

Chapter 5

Conclusion and Future Work

In this chapter, I will summarize the accomplished work and the list of publications. In addition, some ideas on future work will be shown in order to complete the solution to the non-invasive CGM problem.

5.1 Accomplished Work

I presented an overview of glucose molecules and their relation with different types of diabetes in **chapter 1**. I showed how countries at all income levels suffer from high human, social and economic costs for diabetes in all forms. CGM is a very important process for people with diabetes to regularly check their blood glucose levels to keep track of any increase or decrease in these levels, and adjust the amount of medication accordingly. I showed the three main categories of CGM and the desperate need for a simple, non-invasive and pain-free CGM method. Before starting any work, I studied the literature to check some of the past and current non-invasive CGM techniques and the major challenges facing these techniques, to build a solid understanding of the nature of the problem in order to have good non-invasive solution for CGM.

As the proposed system consisted of two parts: hardware part and software one, **Chapter 2** showed the hardware part, which is the microwave sensors used to collect the glucose data. I discussed the process of designing the microwave sensors in order to analyze and predict the glucose levels. To design a microwave sensor, you need to consider many factors carefully, depending on the type of application. Some of those factors include: the operating or resonance frequency of the sensor, the penetration depth and the size of the sensor. Microwave-based CGM modalities are mainly based on the observations of the

variation in the dielectric properties caused by changing in the glucose levels in the tissues. Microwaves, reflected off or scattered from, these tissues are then expected to help in measuring the glucose levels in these tissues. In this thesis, two microwave sensors were used: spiral and dipole sensors. I showed the details of designing each sensor and their responses in simulation and experiment. Finally, I explained the S-parameters, which are used to describe the interactions between the microwave sensors and the material under test (MUT).

Chapter 3 investigated the software part of the system, which involves machine learning algorithms and techniques. I gave general information about machine learning along with some of the very important algorithms and techniques. I showed how we can apply machine learning for the CGM application and how we can calculate the performance metrics. The main application of machine learning algorithms for the CGM is regression and I gave more emphasis on the different regression types and algorithms.

In **chapter 4**, I showed the experimental and simulation results of the thesis. First, experimental results were shown starting from how to prepare the water-glucose solutions which were used in the experiments using both microwave sensors, the dipole and the spiral. Next, the results of the detailed steps of the research methodology were investigates for every sensor separately. After that, simulation results were shown starting from designing the hand model with four different layers in order to change the dielectric properties of those layers and then calculate the sensor's reflected signals due to the interaction of the microwave sensor with the tissues having those dielectric properties. Results of each layers were presented individually by changing of both permittivity and conductivity of each layer.

The main contribution of this research is simplicity. I tried my best in every step of the methodology of the research to be very simple and with the easiest techniques and tools, taking into account the final goals of the system which are: a non-invasive, low-cost to enable over-the-counter availability. The device has to be small and easy to use by a wide population especially children. The results have to be accurate and robust. The system has to be comfortable to use and does not interfere with daily human functions.

5.2 List of Publications

1. **Saeed M. Bamatraf**, Maged A. Aldhaeebi, and Omar M. Ramahi. Non-invasive continuous glucose monitoring on aqueous solutions using microwave sensor with machine learning. Progress In Electromagnetics Research Letters, 102:127–134, 2022.

2. **Saeed M. Bamatraf**, Omar M. Ramahi and Maged A. Aldhaeebi. Non-invasive Aqueous Glucose Monitoring using Microwave Sensor with Machine Learning . 2021 IEEE International Symposium on Antennas and Propagation and USNC-URSI Radio Science Meeting (APS/URSI), 2021, pp. 1875-1876, doi:10.1109/APS/URSI47566.2021.9704583.

3. **Saeed M. Bamatraf** and Omar M. Ramahi. Review of Non-Invasive Continuous Glucose Monitoring Techniques and Challenges. (Under Review by Sensors and Actuators Reports).

4. Maged A Aldhaeebi, Khawla Alzoubi, Thamer S Almoneef, **Saeed M Bamatraf**, Hussein Attia, and Omar M Ramahi. Review of microwaves techniques for breast cancer detection.Sensors, 20(8):2390, 2020.

5. Maged A Aldhaeebi, **Saeed Bamatraf**, Thamer S Almoneef, and Omar M Ramahi. Near-field electrically small sensors array with PCA for microwave breast tumor detection. In 2019 IEEE International Symposium on Antennas and Propagation and USNC-URSI Radio Science Meeting, pages 1007–1008. IEEE, 2019.

6. Maged Aldhaeebi, **Saeed Bamatraf**, Omar Ramahi, and Saeed A Binajjaj. Breast tumor diagnosis using machine learning with microwave probes. In 2019 First International Conference of Intelligent Computing and Engineering (ICOICE), pages1–4. IEEE, 2019.

7. Maged A Aldhaeebi, **Saeed M Bamatraf**, and Omar M Ramahi. Challenges with machine learning for microwave breast tumor detection. Journal of Computational Vision and Imaging Systems, 5(1):1–1, 2019.

5.3   Future Work

- Build a model which will relate the dielectric properties of the human body with the glucose levels which will be conducted using human volunteers from different age groups, females and males.

- Use the trained regression systems to test the data collected from human after mapping S11 with glucose levels.

- Finalize the setup of the system to be commercialized by putting all the parts together in one system taking into considerations the size, cost and safety factors of the proposed system.

References

[1] Bushra Alsunaidi, Murad Althobaiti, Mahbubunnabi Tamal, Waleed Albaker, and Ibraheem Al-Naib. A review of non-invasive optical systems for continuous blood glucose monitoring. *Sensors*, 21(20):6820, 2021.

[2] Ruochong Zhang, Siyu Liu, Haoran Jin, Yunqi Luo, Zesheng Zheng, Fei Gao, and Yuanjin Zheng. Noninvasive electromagnetic wave sensing of glucose. *Sensors*, 19(5):1151, 2019.

[3] Maged A Aldhaeebi, Thamer S Almoneef, Abdulbaset Ali, Zhao Ren, and Omar M Ramahi. Near field breast tumor detection using ultra-narrow band probe with machine learning techniques. *Scientific reports*, 8(1):12607, 2018.

[4] Ralph A DeFronzo, Eleuterio Ferrannini, Kurt George Matthew Mayer Alberti, Paul Zimmet, and George Alberti. *International Textbook of Diabetes Mellitus, 2 Volume Set*, volume 1. John Wiley & Sons, 2015.

[5] International diabetes federation. idf diabetes atlas,10th edn. Brussels,Belgium, https://www.diabetesatlas.org/, 2021.

[6] Vera Bril, Ari Breiner, Bruce A. Perkins, and Douglas Zochodne. Diabetes canada clinical practice guidelines expert committee. *Can J Diabetes*, 42:S217–S221, 2018.

[7] Wenjun Zhang, Yunqing Du, and Ming L Wang. Noninvasive glucose monitoring using saliva nano-biosensor. *Sensing and Bio-Sensing Research*, 4:23–29, 2015.

[8] Dong Geon Jung, Daewoong Jung, and Seong Ho Kong. A lab-on-a-chip-based non-invasive optical sensor for measuring glucose in saliva. *Sensors*, 17(11):2607, 2017.

[9] Oscar Olarte, Jose Chilo, José Pelegri-Sebastia, Kurt Barbe, and Wendy Van Moer. Glucose detection in human sweat using an electronic nose. In *Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE*, pages 1462–1465. IEEE, 2013.

[10] Jason Heikenfeld. Non-invasive analyte access and sensing through eccrine sweat: Challenges and outlook circa 2016. *Electroanalysis*, 28(6):1242–1249, 2016.

[11] Eunyoung Cho, Maedeh Mohammadifar, and Seokheun Choi. A single-use, self-powered, paper-based sensor patch for detection of exercise-induced hypoglycemia. *Micromachines*, 8(9):265, 2017.

[12] Jia Zhu, Shangbin Liu, Zhihui Hu, Xianzhe Zhang, Ning Yi, Kairui Tang, Michael Gregory Dexheimer, Xiaojun Lian, Qing Wang, Jian Yang, et al. Laser-induced graphene non-enzymatic glucose sensors for on-body measurements. *Biosensors and Bioelectronics*, 193:113606, 2021.

[13] Peck Shen Mun, Hua Nong Ting, Yip Boon Chong, and Teng Aik Ong. Dielectric properties of glycosuria at 0.2-50 ghz using microwave spectroscopy. *Journal of Electromagnetic Waves and Applications*, 29(17):2278–2292, 2015.

[14] Qinyi Yan, Bo Peng, Gang Su, Bruce E Cohan, Terry C Major, and Mark E Meyerhoff. Measurement of tear glucose levels with amperometric glucose biosensor/capillary tube configuration. *Analytical chemistry*, 83(21):8341–8346, 2011.

[15] Huanfen Yao, Angela J Shum, Melissa Cowan, Ilkka Lähdesmäki, and Babak A Parviz. A contact lens with embedded sensor for monitoring tear glucose level. *Biosensors and Bioelectronics*, 26(7):3290–3296, 2011.

[16] Yu-Te Liao, Huanfen Yao, Andrew Lingley, Babak Parviz, and Brian P Otis. A 3-cmos glucose sensor for wireless contact-lens tear glucose monitoring. *IEEE Journal of Solid-State Circuits*, 47(1):335–344, 2012.

[17] You-Rong Lin, Chin-Chi Hung, Hsien-Yi Chiu, Po-Han Chang, Bor-Ran Li, Sheng-Jen Cheng, Jia-Wei Yang, Shien-Fong Lin, and Guan-Yu Chen. Noninvasive glucose monitoring with a contact lens and smartphone. *Sensors*, 18(10):3208, 2018.

[18] David Bamgboje, Iasonas Christoulakis, Ioannis Smanis, Gaurav Chavan, Rinkal Shah, Masoud Malekzadeh, Ioannis Violaris, Nikolaos Giannakeas, Markos Tsipouras, Konstantinos Kalafatakis, et al. Continuous non-invasive glucose monitoring via contact lenses: Current approaches and future perspectives. *Biosensors*, 11(6):189, 2021.

[19] Dongmin Guo, David Zhang, Lei Zhang, and Guangming Lu. Non-invasive blood glucose monitoring for diabetics by means of breath signal analysis. *Sensors and Actuators B: Chemical*, 173:106–113, 2012.

[20] Masoud Baghelani, Zahra Abbasi, Mojgan Daneshmand, and Peter E Light. Non-invasive continuous-time glucose monitoring system using a chipless printable sensor based on split ring microwave resonators. *Scientific Reports*, 10(1):1–15, 2020.

[21] Tzu-Ting Wei, Hsin-Yi Tsai, Ching-Ching Yang, Wen-Tse Hsiao, and Kuo-Cheng Huang. Noninvasive glucose evaluation by human skin oxygen saturation level. In *Instrumentation and Measurement Technology Conference Proceedings (I2MTC), 2016 IEEE International*, pages 1–5. IEEE, 2016.

[22] David C Klonoff. Overview of fluorescence glucose sensing: a technology with a bright future, 2012.

[23] Jitendra Solanki, Pratima Sen, Joseph Thomas Andrews, and Kamal Kishore Thareja. Blood glucose monitoring in human subjects using optical coherence tomography. *Journal of Optics*, 41(3):127–133, 2012.

[24] Jessica Hanna, Moussa Bteich, Youssef Tawk, Ali H Ramadan, Batoul Dia, Fatima A Asadallah, Aline Eid, Rouwaida Kanj, Joseph Costantine, and Assaad A Eid. Noninvasive, wearable, and tunable electromagnetic multisensing system for continuous glucose monitoring, mimicking vasculature anatomy. *Science Advances*, 6(24):eaba5320, 2020.

[25] Yang Zhang, Jian-ming Zhu, Yong-bo Liang, Hong-bo Chen, Shi-min Yin, and Zhencheng Chen. Non-invasive blood glucose detection system based on conservation of energy method. *Physiological measurement*, 38(2):325, 2017.

[26] Luca Lipani, Bertrand GR Dupont, Floriant Doungmene, Frank Marken, Rex M Tyrrell, Richard H Guy, and Adelina Ilie. Non-invasive, transdermal, path-selective and specific glucose monitoring via a graphene-based platform. *Nature nanotechnology*, page 1, 2018.

[27] Ilana Harman-Boehm, Avner Gal, Alexander M Raykhman, Eugene Naidis, and Yulia Mayzel. Noninvasive glucose monitoring: increasing accuracy by combination of multi-technology and multi-sensors. *Journal of diabetes science and technology*, 4(3):583–595, 2010.

[28] Abu Asaduzzaman, Soumyashree Samadarsinee, and Kishore K Chidella. Simulating multisensor noninvasive blood glucose monitoring systems. In *SoutheastCon, 2016*, pages 1–7. IEEE, 2016.

[29] Linh Lan Nguyen, Steven Su, and Hung T Nguyen. Neural network approach for non-invasive detection of hyperglycemia using electrocardiographic signals. In *Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE*, pages 4475–4478. IEEE, 2014.

[30] Sai Ho Ling, Phyo Phyo San, and Hung T Nguyen. Non-invasive hypoglycemia monitoring system using extreme learning machine for type 1 diabetes. *ISA transactions*, 64:440–446, 2016.

[31] Maryamsadat Shokrekhodaei and Stella Quinones. Review of non-invasive glucose sensing techniques: optical, electrical and breath acetone. *Sensors*, 20(5):1251, 2020.

[32] Benjamin Jasha Van Enter and Elizabeth Von Hauff. Challenges and perspectives in continuous glucose monitoring. *Chemical Communications*, 54(40):5032–5045, 2018.

[33] Dixit N Sathyanarayana. *Vibrational spectroscopy: theory and applications*. New Age International, 2015.

[34] Chi-Fuk So, Kup-Sze Choi, Thomas KS Wong, and Joanne WY Chung. Recent advances in noninvasive glucose monitoring. *Medical Devices (Auckland, NZ)*, 5:45, 2012.

[35] Vishnu Dantu, Jagannadh Vempati, and Srinivasan Srivilliputhur. Non-invasive blood glucose monitor based on spectroscopy using a smartphone. In *Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE*, pages 3695–3698. IEEE, 2014.

[36] Zhe Li, Gang Li, Wen-Juan Yan, and Ling Lin. Classification of diabetes and measurement of blood glucose concentration noninvasively using near infrared spectroscopy. *Infrared Physics & Technology*, 67:574–582, 2014.

[37] Xiaoli Li and Chengwei Li. Research on non-invasive glucose concentration measurement by nir transmission. In *Computer and Communications (ICCC), 2015 IEEE International Conference on*, pages 223–228. IEEE, 2015.

[38] Ghozzi Dorsaf, Manai Yacine, and Nouri Khaled. Non-invasive glucose monitoring: Application and technologies. *Curr Trends Biomedical Eng Biosci*, 14, 2018.

[39] Xinxin Guo, Andreas Mandelis, and Bernard Zinman. Noninvasive glucose detection in human skin using wavelength modulated differential laser photothermal radiometry. *Biomedical optics express*, 3(11):3012–3021, 2012.

[40] Cnoga. Combo glucometer. https://cnogacare.co/hybridglucometer/, 2018.

[41] Yosef Segman. Device and method for noninvasive glucose assessment. *Journal of diabetes science and technology*, 12(6):1159–1168, 2018.

[42] Tamar Lin, Avner Gal, Yulia Mayzel, Keren Horman, and Karnit Bahartan. Non-invasive glucose monitoring: a review of challenges and recent advances. *Curr Trends Biomed Eng Biosci*, 6:1–8, 2017.

[43] Armin Schneider and Hubertus Feussner. *Biomedical engineering in gastrointestinal surgery*. Academic Press, 2017.

[44] Naoki Wadamori. Behavior of long-period measurements using a small-sized photoacoustic cell for aqueous glucose monitoring. In *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*, pages 1267–1270. IEEE, 2015.

[45] Mohamad G Ghosn, Narendran Sudheendran, Mark Wendt, Adrian Glasser, Valery V Tuchin, and Kirill V Larin. Monitoring of glucose permeability in monkey skin in vivo using optical coherence tomography. *Journal of biophotonics*, 3(1-2):25–33, 2010.

[46] Said Amrane, Nawfel Azami, and Youssef Elboulqe. Optimized algorithm of dermis detection for glucose blood monitoring based on optical coherence tomography. In *Intelligent Systems: Theories and Applications (SITA), 2015 10th International Conference on*, pages 1–5. IEEE, 2015.

[47] YT Lan, YP Kuang, LP Zhou, GY Wu, PC Gu, HJ Wei, and K Chen. Noninvasive monitoring of blood glucose concentration in diabetic patients with optical coherence tomography. *Laser Physics Letters*, 14(3):035603, 2017.

[48] Qingqing Ke, Yu Zheng, Fan Yang, Hanchang Zhang, and Xiurong Yang. A fluorescence glucose sensor based on ph induced conformational switch of i-motif dna. *Talanta*, 129:539–544, 2014.

[49] Erdem Topsakal, Tutku Karacolak, and Elaine C Moreland. Glucose-dependent dielectric properties of blood plasma. In *General Assembly and Scientific Symposium, 2011 XXXth URSI*, pages 1–4. IEEE, 2011.

[50] Margarita Puentes Vargas. *Planar Metamaterial Based Microwave Sensor Arrays for Biomedical Analysis and Treatment*. Springer Science & Business Media, 2014.

[51] Shizhen Hu, Seko Nagae, and Akira Hirose. Millimeter-wave adaptive glucose concentration estimation with complex-valued neural networks. *IEEE Transactions on Biomedical Engineering*, 2018.

[52] Wilbert Villena Gonzales, Ahmed Toaha Mobashsher, and Amin Abbosh. The progress of glucose monitoring—a review of invasive to minimally and non-invasive techniques, devices and sensors. *Sensors*, 19(4):800, 2019.

[53] Arsen Bababjanyan, Harutyun Melikyan, Seungwan Kim, Jongchel Kim, Kiejin Lee, and Barry Friedman. Real-time noninvasive measurement of glucose concentration using a microwave biosensor. *Journal of Sensors*, 2010, 2010.

[54] Shiv Kumar and Jaspal Singh. Measuring blood glucose levels with microwave sensor. *International Journal of Computer Applications*, 72(15), 2013.

[55] G Guarin, M Hofmann, R Weigel, G Fischer, and D Kissinger. Determination of sugar concentration in aqueous solutions using ultra-wideband microwave impedance spectroscopy. In *Microwave Symposium Digest (IMS), 2013 IEEE MTT-S International*, pages 1–4. IEEE, 2013.

[56] A Elkady, M El-Hadidy, A Medhat, A Khorshid, and A Darwish. Microwave power absorption in human body for non-invasive glucose monitoring. *Prog Electromag Res Sym Proc*, pages 109–113, 2013.

[57] Rasool Baghbani, Masoomeh Ashoori Rad, and Ali Pourziad. Microwave sensor for non-invasive glucose measurements design and implementation of a novel linear. *IET Wireless Sensor Systems*, 5(2):51–57, 2015.

[58] Maximilian Hofmann, Georg Fischer, Robert Weigel, and Dietmar Kissinger. Microwave-based noninvasive concentration measurements for biomedical applications. *IEEE Transactions on Microwave Theory and Techniques*, 61(5):2195–2204, 2013.

[59] Santiago Pimentel, Pablo Daniel Agüero, Alejandro José Uriz, Juan Carlos Bonadero, Mónica Liberatori, and Jorge Castiñeira Moreira. Simulation of a non-invasive glucometer based on a microwave resonator sensor. In *Journal of Physics: Conference Series*, volume 477, page 012020. IOP Publishing, 2013.

[60] Anoop Adhyapak, Matthew Sidley, and Jayanti Venkataraman. Analytical model for real time, noninvasive estimation of blood glucose level. In *Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE*, pages 5020–5023. IEEE, 2014.

[61] Tuba Yilmaz, Robert Foster, and Yang Hao. Broadband tissue mimicking phantoms and a patch resonator for evaluating noninvasive monitoring of blood glucose levels. *IEEE transactions on Antennas and Propagation*, 62(6):3064–3075, 2014.

[62] Apurva A Muley and Rajesh B Ghongade. Design and simulate an antenna for aqueous glucose measurement. In *India Conference (INDICON), 2014 Annual IEEE*, pages 1–6. IEEE, 2014.

[63] Heungjae Choi, Jack Nylon, Stephen Luzio, Jan Beutler, and Adrian Porch. Design of continuous non-invasive blood glucose monitoring sensor based on a microwave split ring resonator. In *RF and Wireless Technologies for Biomedical and Healthcare Applications (IMWS-Bio), 2014 IEEE MTT-S International Microwave Workshop Series on*, pages 1–3. IEEE, 2014.

[64] Tuba Yilmaz, Robert Foster, and Yang Hao. Towards accurate dielectric property retrieval of biological tissues for blood glucose monitoring. *IEEE transactions on Microwave Theory and Techniques*, 62(12):3193–3204, 2014.

[65] Ali A Abduljabar, David J Rowe, Adrian Porch, and David A Barrow. Novel microwave microfluidic sensor using a microstrip split-ring resonator. *IEEE Transactions on Microwave Theory and Techniques*, 62(3):679–688, 2014.

[66] Heungjae Choi, Jack Naylon, Steve Luzio, Jan Beutler, James Birchall, Chris Martin, and Adrian Porch. Design and in vitro interference test of microwave noninvasive blood glucose monitoring sensor. *system*, 19:20, 2015.

[67] Jan Vrba, Jakub Karch, and David Vrba. Phantoms for development of microwave sensors for noninvasive blood glucose monitoring. *International Journal of Antennas and Propagation*, 2015, 2015.

[68] Nam-Young Kim, Kishor Kumar Adhikari, Rajendra Dhakal, Zorigt Chuluunbaatar, Cong Wang, and Eun-Soo Kim. Rapid, sensitive, and reusable detection of glucose by a robust radiofrequency integrated passive device biosensor chip. *Scientific reports*, 5:7807, 2015.

[69] Berk Camli, Emre Kusakci, Berkan Lafci, Seyhan Salman, Hamdi Torun, and Arda Yalcinkaya. A microwave ring resonator based glucose sensor. *Procedia Eng*, 168:465–468, 2016.

[70] Jinjin Shao, Fan Yang, Fen Xia, Qingfeng Zhang, and Yifan Chen. A novel miniature spiral sensor for non-invasive blood glucose monitoring. In *Antennas and Propagation (EuCAP), 2016 10th European Conference on*, pages 1–2. IEEE, 2016.

108

[71] Shimul Saha, Ioannis Sotiriou, Ioannis Gouzouasis, Helena Cano-Garcia, George Palikaras, Panagiotis Kosmas, and Efthymios Kallos. Evaluation of the sensitivity of transmission measurements at millimeter waves using patch antennas for non-invasive glucose sensing. In *Antennas and Propagation (EuCAP), 2016 10th European Conference on*, pages 1–4. IEEE, 2016.

[72] Vidya V Deshmukh and Rajesh B Ghongade. Measurement of dielectric properties of aqueous glucose using planar ring resonator. In *Microelectronics, Computing and Communications (MicroCom), 2016 International Conference on*, pages 1–5. IEEE, 2016.

[73] Md Shawkat Ali, NJ Shoumy, S Khatun, LM Kamarudin, and V Vijayasarveswari. Non-invasive blood glucose measurement performance analysis through uwb imaging. In *Electronic Design (ICED), 2016 3rd International Conference on*, pages 513–516. IEEE, 2016.

[74] Heungjae Choi, Steve Luzio, Jan Beutler, and Adrian Porch. Microwave noninvasive blood glucose monitoring sensor: human clinical trial results. In *Microwave Symposium (IMS), 2017 IEEE MTT-S International*, pages 876–879. IEEE, 2017.

[75] Debasish Mondal, Nilesh Kumar Tiwari, and M Jalcel Akhtar. Microwave assisted non-invasive microfluidic biosensor for monitoring glucose concentration. In *2018 IEEE SENSORS*, pages 1–4. IEEE, 2018.

[76] Ratnesh Kumari, Piyush N Patel, and Rahul Yadav. An eng resonator-based microwave sensor for the characterization of aqueous glucose. *Journal of Physics D: Applied Physics*, 51(7):075601, 2018.

[77] Ala Eldin Omer, George Shaker, Safieddin Safavi-Naeini, Hamid Kokabi, Georges Alquié, Frédérique Deshours, and Raed M Shubair. Low-cost portable microwave sensor for non-invasive monitoring of blood glucose level: Novel design utilizing a four-cell csrr hexagonal configuration. *Scientific Reports*, 10(1):1–20, 2020.

[78] Ala Eldin Omer, George Shaker, Safieddin Safavi-Naeini, Georges Alquié, Frédérique Deshours, Hamid Kokabi, and Raed M Shubair. Non-invasive real-time monitoring of glucose level using novel microwave biosensor based on triple-pole csrr. *IEEE Transactions on Biomedical Circuits and Systems*, 14(6):1407–1420, 2020.

[79] Andreas Caduff, Etienne Hirt, Yu Feldman, Zulfiqur Ali, and Lutz Heinemann. First human experiments with a novel non-invasive, non-optical continuous glucose monitoring system. *Biosensors and Bioelectronics*, 19(3):209–217, 2003.

[80] JH DeVries, IME Wentholt, A Zwart, and JB Hoekstra. Pendra goes dutch; lessons for the ce mark in europe. *Diabetes Research and Clinical Practice*, 74:S93–S96, 2006.

[81] Christopher McCormick, David Heath, and Patricia Connolly. Towards blood free measurement of glucose and potassium in humans using reverse iontophoresis. *Sensors and Actuators B: Chemical*, 166:593–600, 2012.

[82] Nemaura Medical. Sugarbeat nemaura medical. https://nemauramedical.com/sugarbeat/, 2016.

[83] Md Koushik Chowdhury, Anuj Srivastava, Neeraj Sharma, and Shiru Sharma. Non-invasive blood glucose measurement utilizing a newly designed system based on modulated ultrasound and infrared light. *International Journal of Diabetes in Developing Countries*, 36(4):439–448, 2016.

[84] Lili Zhu, Jieqing Lin, Baiqing Lin, and Hui Li. Noninvasive blood glucose measurement by ultrasound-modulated optical technique. *Chinese Optics Letters*, 11(2):021701, 2013.

[85] Anuj Srivastava, Md Koushik Chowdhury, Shiru Sharma, and Neeraj Sharma. Measurement of glucose by using modulating ultrasound with optical technique in normal and diabetic human blood serum. In *2014 International Conference on Advances in Engineering & Technology Research (ICAETR-2014)*, pages 1–5. IEEE, 2014.

[86] Andreas Caduff, Mark S Talary, Martin Mueller, Francois Dewarrat, Jelena Klisic, Marc Donath, Lutz Heinemann, and Werner A Stahel. Non-invasive glucose monitoring in patients with type 1 diabetes: A multisensor system combining sensors for dielectric and optical characterisation of skin. *Biosensors and Bioelectronics*, 24(9):2778–2784, 2009.

[87] Thomas Chretiennot, David Dubuc, and Katia Grenier. Microwave-based microfluidic sensor for non-destructive and quantitative glucose monitoring in aqueous solution. *Sensors*, 16(10):1733, 2016.

[88] Zhanxiao Geng, Fei Tang, Yadong Ding, Shuzhe Li, and Xiaohao Wang. Noninvasive continuous glucose monitoring using a multisensor-based glucometer and time series analysis. *Scientific reports*, 7(1):12650, 2017.

[89] Ilana Harman-Boehm, Avner Gal, Alexander M Raykhman, Jeffrey D Zahn, Eugene Naidis, and Yulia Mayzel. Noninvasive glucose monitoring: a novel approach, 2009.

[90] K Horman, Y Mayzel, A Gal, K Bahartan, A Drexler, and T Lin. Performance and user experience evaluation of a non-invasive glucose monitoring device. *Int J Diabetes Metab Disord*, 1(2):1–7, 2016.

[91] Liu Tang, Shwu Jen Chang, Ching-Jung Chen, and Jen-Tsai Liu. Non-invasive blood glucose monitoring technology: a review. *Sensors*, 20(23):6925, 2020.

[92] Permittivity. Britannica, The Editors of Encyclopaedia, "https://www.britannica.com/science/permittivity.". Accessed: 2022-1-9.

[93] Camelia Gabriel, Sami Gabriel, and y E Corthout. The dielectric properties of biological tissues: I. literature survey. *Physics in medicine & biology*, 41(11):2231, 1996.

[94] Jiri Polivka et al. An overview of microwave sensor technology. *High Frequency Electronics*, 6(4):32–42, 2007.

[95] Rammah Ali Alahnomi, Zahriladha Zakaria, Zulkalnain Mohd Yussof, Ayman Abdulhadi Althuwayb, Ammar Alhegazi, Hussein Alsariera, and Norhanani Abd Rahman. Review of recent microwave planar resonator-based sensors: Techniques of complex permittivity extraction, applications, open challenges and future research directions. *Sensors*, 21(7):2267, 2021.

[96] Roger M Langdon. Resonator sensors-a review. *Journal of Physics E: Scientific Instruments*, 18(2):103, 1985.

[97] S Hensley and T Farr. 3.3 microwave remote sensing and surface characterization. 2013.

[98] Arthur R von Hippel and SO Morgan. Dielectric materials and applications. *Journal of The Electrochemical Society*, 102(3):68Ca, 1955.

[99] David M Pozar. *Microwave engineering*. John wiley & sons, 2011.

[100] CST. Computer simulation technology cst. http://www.CST.com/, 7 2019.

[101] Kevin P Murphy. *Machine learning: a probabilistic perspective*. MIT press, 2012.

[102] Steve R Gunn et al. Support vector machines for classification and regression. *ISIS technical report*, 14(1):5–16, 1998.

[103] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine learning*, 20(3):273–297, 1995.

[104] Bernhard Schölkopf, Alexander J Smola, Francis Bach, et al. *Learning with kernels: support vector machines, regularization, optimization, and beyond.* MIT press, 2002.

[105] Douglas C Montgomery, Elizabeth A Peck, and G Geoffrey Vining. *Introduction to linear regression analysis.* John Wiley & Sons, 2021.

[106] William L Clarke, Daniel Cox, Linda A Gonder-Frederick, William Carter, and Stephen L Pohl. Evaluating clinical accuracy of systems for self-monitoring of blood glucose. *Diabetes care*, 10(5):622–628, 1987.

[107] William L Clarke. The original clarke error grid analysis (ega). *Diabetes technology & therapeutics*, 7(5):776–779, 2005.

[108] Florian Reiterer, Philipp Polterauer, Michael Schoemaker, Guenther Schmelzeisen-Redecker, Guido Freckmann, Lutz Heinemann, and Luigi Del Re. Significance and reliability of mard for the accuracy of cgm systems. *Journal of diabetes science and technology*, 11(1):59–67, 2017.

[109] Anne-Marie Brady, Catherine McCabe, and Margaret McCann. *Fundamentals of Medical-Surgical Nursing: A Systems Approach.* John Wiley & Sons, 2013.

[110] R Pethig. Dielectric properties of body tissues. *Clinical Physics and Physiological Measurement*, 8(4A):5, 1987.

[111] David Harvey. *Modern analytical chemistry*, volume 1. McGraw-Hill New York, 2000.

[112] Starting Matlab. Matlab. *The MathWorks, Natick, MA*, 2012.

[113] Ian T Jolliffe. Principal components in regression analysis. *Principal component analysis*, pages 167–198, 2002.

[114] Carl Edward Rasmussen. Gaussian processes in machine learning. In *Summer school on machine learning*, pages 63–71. Springer, 2003.

[115] Christopher K Williams and Carl Edward Rasmussen. *Gaussian processes for machine learning*, volume 2. MIT press Cambridge, MA, 2006.

[116] Sami Gabriel, RW Lau, and Camelia Gabriel. The dielectric properties of biological tissues: Iii. parametric models for the dielectric spectrum of tissues. *Physics in Medicine & Biology*, 41(11):2271, 1996.

[117] Syaiful Redzwan Mohd Shah, Jacob Velander, Parul Mathur, Mauricio D Perez, Noor Badariah Asan, Dhanesh G Kurup, Taco J Blokhuis, and Robin Augustine. Split-ring resonator sensor penetration depth assessment using in vivo microwave reflectivity and ultrasound measurements for lower extremity trauma rehabilitation. *Sensors*, 18(2):636, 2018.

[118] D Andreuccetti. An internet resource for the calculation of the dielectric properties of body tissues in the frequency range 10 hz-100 ghz. *http://niremf.ifac.cnr.it/tissprop/*, 2012.