Received April 16, 2021, accepted May 2, 2021, date of publication May 11, 2021, date of current version May 24, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3079182

Non-Invasive Glucose Monitoring Using Optical Sensor and Machine Learning Techniques for Diabetes Applications

MARYAMSADAT SHOKREKHODAEI[1], (Graduate Student Member, IEEE), DAVID P. CISTOLA[2], ROBERT C. ROBERTS[1], (Member, IEEE), AND STELLA QUINONES[3]

[1]Department of Electrical and Computer Engineering, The University of Texas at El Paso, El Paso, TX 79968, USA
[2]Center of Emphasis in Diabetes and Metabolism, Paul L. Foster School of Medicine, Texas Tech University Health Sciences Center El Paso, El Paso, TX 79905, USA
[3]Department Metallurgical, Materials and Biomedical Engineering, The University of Texas at El Paso, El Paso, TX 79968, USA

Corresponding author: Maryamsadat Shokrekhodaei (mshokrekhod@miners.utep.edu)

ABSTRACT Diabetes is a major public health challenge affecting more than 451 million people. Physiological and experimental factors influence the accuracy of non-invasive glucose monitoring, and these need to be overcome before replacing the finger prick method. Also, the suitable employment of machine learning techniques can significantly improve the accuracy of glucose predictions. One aim of this study is to use light sources with multiple wavelengths to enhance the sensitivity and selectivity of glucose detection in an aqueous solution. Multiple wavelength measurements have the potential to compensate for errors associated with inter- and intra-individual differences in blood and tissue components. In this study, the transmission measurements of a custom built optical sensor are examined using 18 different wavelengths between 410 and 940 nm. Results show a high correlation value (0.98) between glucose concentration and transmission intensity for four wavelengths (485, 645, 860 and 940 nm). Five machine learning methods are investigated for glucose predictions. When regression methods are used, 9% of glucose predictions fall outside the correct range (normal, hypoglycemic or hyperglycemic). The prediction accuracy is improved by applying classification methods on sets of data arranged into 21 classes. Data within each class corresponds to a discrete 10 mg/dL glucose range. Classification based models outperform regression, and among them, the support vector machine is the most successful with F1-score of 99%. Additionally, Clarke error grid shows that 99.75% of glucose readings fall within the clinically acceptable zones. This is an important step towards critical diagnosis during an emergency patient situation.

INDEX TERMS Classification, decision tree, diabetes, k-nearest neighbor, machine learning, neural network, non-invasive, optimization, spectroscopy, support vector machine.

I. INTRODUCTION

Diabetes mellitus is a metabolic disorder that affects the body's ability to process blood glucose. Diabetic patients monitor their blood glucose levels in an effort to keep them in the normal range (approximately 70 to 180 mg/dL, depending on food intake) by medication, exercise, proper diet, etc. Blood glucose levels below 70 mg/dL serve as an alert for possible life-threatening hypoglycemia, and glucose levels higher than 180 mg/dL can indicate clinically significant hyperglycemia [1], [2]. Frequent monitoring of blood glucose levels informs diabetic patients to take appropriate action to adjust their blood glucose levels, and thus avoid risks that are associated with hypo and hyperglycemia.

Outside of the reference method for plasma glucose measurements used in clinical laboratories, the conventional finger-prick method using glucose strips and accompanying meter is the most reliable method for patient self-monitoring. A finger prick based glucose meter has two essential parts: a testing strip coated with enzymes (e.g. glucose oxidase (GOx), glucose dehydrogenase (GDH), and hexokinase (HK)), and a detector composed of electronics. When a drop of blood is applied to a testing strip, the glucose within the blood sample reacts with the enzymes and the resulting

The associate editor coordinating the review of this manuscript and approving it for publication was Pasquale De Meo.

electrochemical reaction produces a current signal which is linearly proportional to the glucose concentration [3].

Calibration methods such as simple linear regression or Deming regression are used to convert the current signal (in order of nano amps) generated by the meter electronics to a digitized blood glucose value. Linear regression is a technique used to fit a straight line to two-dimensional data. One of the variables is the current signal from the sensor based on standard solutions of known glucose concentration (x), and the second variable is the reference glucose concentration (y). Optimal values for the regression coefficients are determined by minimizing the sum of the squared error in the y direction (i.e. glucose reading) [4]. The Deming regression fits a line for two dimensional data where measurement errors are assumed for both x and y values. Deming's regression coefficients are determined by minimizing the sum of the squared errors in both the x (i.e. current signal readings) and y directions (i.e. glucose readings) [5].

The accuracy of blood glucose meter readings are dependent on the test strip material, fabrication process, operating procedures by patients, environmental conditions, and patient medication. The technical accuracy of a glucose meter is determined by comparing the glucose readings from the blood samples analyzed using a glucose meter against the blood plasma samples analyzed by laboratory methods at the same time [3]. It is well established that the finger-prick method is a reliable method for accurate glucose measurements. However, consistent penetration of the skin is painful, inconvenient and carries a risk of infection. Non-invasive glucose measurement methods have the potential to ease glucose detection and can result in greater patient comfort and more effective treatment options.

Research groups have been trying for decades to develop a reliable commercial non-invasive glucose device with long-term accuracy. However, none of the current non-invasive devices exhibit the long-term accuracy required to replace the finger prick methods. There is room to improve the performance of non-invasive glucose sensors for public use under regular home conditions. These non-invasive glucose measurement techniques are based on infrared spectroscopy [6]–[8], Raman spectroscopy [9], [10], polarimetry [11], photoacoustic spectroscopy [12], millimeter wave/microwave spectroscopy [13]–[15], bio-impedance spectroscopy [16], optical coherence tomography [17], and hybrid techniques [18], [19].

This study focuses on the custom design of an optical glucose sensor with light sources and detectors within the visible (VIS) and near-infrared (NIR) range. The optimal wavelengths of light are analyzed to accurately predict the concentration of glucose in solution. The intensity of the signal transmitted through the sample is dependent on the absorption and scattering behavior of light within the sample, which depends on the glucose concentration in solution [20]. The use of light sources in the VIS-NIR range is preferred over light sources in the mid-infrared (MIR) range because of the lower absorption of VIS-NIR light by tissue and water content in the blood. The strong absorption of MIR light by water decreases the intensity of light that is able to penetrate the blood and tissue samples and thus decreases the corresponding intensity of light that is picked up by detectors. This means that MIR light can only penetrate the interstitial fluid of tissue whereas VIS-NIR light can penetrate deeper into the tissue and interact with glucose molecules within both the interstitial fluid and the blood vessels [20]. Since it takes time for glucose to diffuse from blood vessels into the interstitial fluid, there is a difference in glucose levels within the blood vs the interstitial fluid at any given time [20]. VIS-NIR light can interact with glucose molecules within the blood in real time, and this increases the sensitivity of the sensor device to read real time glucose concentration changes. Thus, this study focuses on VIS-NIR light sources.

The low concentrations of glucose in the blood makes it challenging to detect glucose in the body by non-invasive methods. In addition, signals associated with other species and tissue components may obscure the already low signal from glucose molecules. Examples of dominant interfering species include water content in the blood, protein (albumin and hemoglobin), fatty tissue, etc. The variation in concentration of these components between individuals makes it challenging to detect and measure fluctuations of glucose concentrations. A possible solution to overcome this challenge is to use a glucose sensor with multiple wavelength sources. The light attenuation coefficient is unique for each species and is dependent on the wavelength of the incident light interacting with each molecule.

The VIS-NIR optical sensor that is developed for this work supports light intensity measurements at multiple wavelengths. Multi-wavelength analysis helps to extract more quantitative information from light interactions with the different species in aqueous glucose solutions. In this study, the intensity data measured using 485, 645 and 860 and 940 nm light sources are used collectively to achieve a higher sensitivity to glucose concentration.

The data analysis of intensity data and associated glucose prediction models have an effect on the accuracy of glucose readings for any kind of measurement device. The complexity of non-invasive glucose measurements necessitates the use of more effective algorithms capable of dealing with higher dimensional data in order to improve the accuracy of non-invasive glucose concentration measurements. Various machine learning techniques are used for glucose predictions, such as multiple linear regression [21], partial least square regression [22], [23], principal component regression [22], feed forward neural networks [24], [25], deep neural networks [26], [27], support vector machines [27], [28], random forest regression [18], etc. These techniques are used to build regression based models that predict continuous values of glucose concentration. The goal of regression techniques is to train a model to minimize the error between an actual and predicted glucose concentration value, but it is not directly involved in reducing the error associated with the detection of hypoglycemia and hyperglycemia. Thus, predictions made by

regression models may be limited in their ability to accurately predict hyperglycemia and hypoglycemia, especially in real life applications where various factors affect the measured signal.

In this study, machine learning classification techniques are compared to regression models and are used to improve the accuracy of glucose predictions. The data samples measured at the four optimal wavelengths are subdivided into 21 classes. Each class includes a series of data samples (∼95 samples) within a discrete 10 mg/dL glucose range, and each class belongs to one of the three ranges: hypoglycemic, normal, or hyperglycemic ranges. Thus, with 21 classes and ∼95 samples per class, there are ∼2,000 data points in this study. The class labels and measured intensity data are then used as input for classification algorithms that result in a classifier with (1) maximum accuracy in classification of data samples into one of the 21 classes, and (2) greatest capability in the identification of hypoglycemia, normal or hyperglycemia conditions for each input data point.

The rest of this paper includes Sections II-VI. In Section II, the focus is on the design and implementation of the VIS-NIR optical glucose sensor. The selection process of 4 optimal wavelengths from 18 original wavelengths ranging between 410 and 940 nm is described in Section III. Machine learning techniques and how they are used to train 5 glucose prediction models is discussed in Section IV. Measurement results along with a comparison between each prediction model are provided in Section V. Finally, the results of this research are discussed in Section VI.

II. EXPERIMENTAL DESIGN AND IMPLEMENTATION

A. GLUCOSE SENSOR EXPERIMENTAL SET UP

Fig. 1 illustrates a schematic of the optical based glucose sensor. The basic components consist of (1) several VIS-NIR light sources that are soldered on a designed PCB; (2) a Triad AS7265x Spectroscopy Sensor (SparkFun Electronics, Colorado, USA) that serves as a photo detector in the VIS-NIR range; (3) an optical cuvette (Starna Cells Inc., CA, USA) to hold the glucose sample, (4) a peristaltic fluidic pump (PE-WX14-S21-1, REET Corp., Connecticut, USA) to control flow rates between 2.5 μL/min and 24 mL/min; and (5) several 3D printed parts designed and printed for the PCB holder and platform. The cuvette, light sources and the Triad Sensor are secured on the 3D printed platform to ensure stability and minimize measurement error due to movement. The peristaltic pump with stepper motor and driver automatically and precisely changes the concentration of glucose in solution. This set-up eliminates measurement errors from manual changes in the glucose concentration and results in corresponding continuous changes in glucose concentration measurements. The distances between (1) the light sources and cuvette and (2) the cuvette and detectors are adjustable, from 1 cm to 5 cm, and this facilitates the selection of the optimal distance to ensure maximum sensitivity to glucose changes. The Triad AS7265x spectroscopy sensor measures

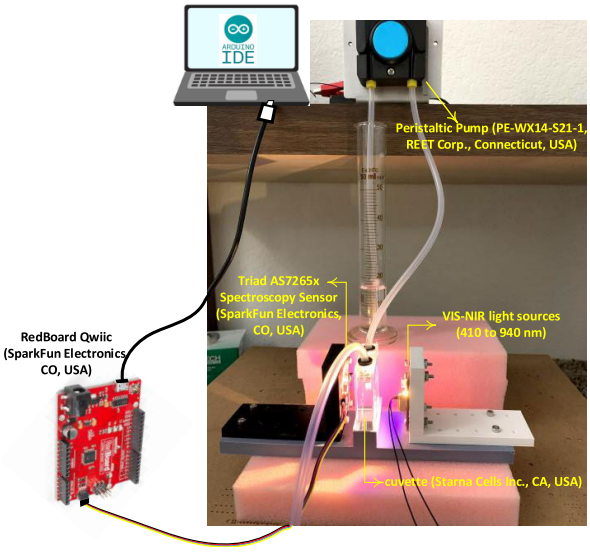

FIGURE 1. VIS-NIR based optical glucose sensor set up.

light in a range between 410 nm and 940 nm with 18 individual frequency bands (each with 20 nm full width at half maximum (FWHM)). A SparkFun RedBoard Qwiic (SparkFun Electronics, Colorado, USA) is used to collect data from the Triad. The RedBoard includes an Arduino-compatible microcontroller that is programmed using the Arduino Integrated Development Environment (IDE) Software written in C++ code in order to read the Triad signal at 18 wavelengths. The measurement integration time and gain can be controlled using command-line programming.

The peristaltic fluidic pump flows distilled water into a glucose solution in an optical glass cuvette with a volume of 14 mL. This lowers the glucose concentration in a controlled manner. The VIS-NIR light source produces light with a range of wavelengths between 410 nm and 940 nm, and the light transmits through the sample. When light impinges on the sample, physiological factors within the glucose sample absorb a portion of the light, and the rest of the light passes through the sample. Light also scatters at different angles from the incident light. When there is no change in the angle, transmission of light through the sample occurs. Absorption, scattering and transmission of light through the sample is a function of the optical properties of the sample, and depends on the concentration of components in solution. Triad sensors collect the light passing through the sample and the IDE interface software displays the data on a laptop screen. Then, the intensity data is processed using machine learning methods to determine glucose concentration.

B. EXPERIMENTAL METHODS

Plasma is primarily made up of water (∼92%). Blood plasma also includes proteins (7%), inorganic salts (0.5%), and glucose (0.07-0.1%) [29]. The most abundant protein, by far, is serum albumin (3.5-5.5 g/dL or ∼60% of total plasma protein), and the second most abundant is immunoglobulin

G (0.7-1.6 g/dL or ∼20% of total plasma protein), followed by transferrin and fibrinogen [29], [30]. There are two forms of glucose known as L-glucose and D-glucose (or dextrose). L-glucose and D-glucose have the same chemical formula ($C_6H_{12}O_6$), but the molecular arrangement is different. The amount of L-glucose in a complex living organism such as human blood is very negligible (in order of nm/dL) compared to the amount of D-glucose (in order of mg/dL) [31], and commercially available glucose meters rely on measuring D-glucose levels within the blood [31].

Among the components within the plasma, water acts as the main influencer affecting the accuracy of non-invasive glucose measurements. Water is the largest component in plasma, absorbs a significant percentage of the incident light, and this reduces the sensitivity of the glucose sensor to track changes in glucose concentration. A comparable in-vitro solution can be made by dissolving D-glucose in water to simulate the adverse effect of the water signal coming from blood plasma while trying to measure glucose. Although the in-vitro testing of this type of solution is different from in-vivo sampling, in-vitro experiments make it possible to measure the influence of obstructive factors, and to identify methodological or theoretical approaches to improve the accuracy of glucose measurements. We can consider in-vitro experiments as a first and important step to examine the feasibility and accuracy of non-invasive glucose sensor measurements before applying the sensor to a body measurement site (e.g. ear, arm, and finger).

A 250 mg/dL glucose solution is made by dissolving 250 mg of D-Glucose (Carolina Biological Supply, NC, USA) in 100 mL of distilled water and an initial volume of 14 mL is poured into the cuvette. A peristaltic pump is connected to one of the two holes on top of the cuvette and is used to transfer distilled water into the initial volume at a rate of 7 mL/min. The flow of water into the cuvette through the first hole causes the same volume of solution to be pushed out through the second hole. It takes about 3.5 minutes for the pump to add 24 mL of distilled water into the cuvette, and this gradually lowers the glucose concentration (from 250 to 40 mg/dL) in a controlled manner. During each test, as the glucose concentration is diluted from 250 mg/dL to 40 mg/dL, the SparkFun Triad measures the intensity of light that passes through the sample at 18 different wavelengths ranging from 410 nm to 940 nm. The time interval between each reading is about 1 second, resulting in the collection of approximately 200 readings or data points for each test for a specific wavelength.

After measuring the light intensity transmitted through the glucose-distilled water solution, the same experiment is repeated for distilled water without glucose. The entire process described here is repeated 10 times during different days, resulting in 10 experiments for glucose within distilled water and 10 experiments for distilled water alone. Table 1 is an abbreviated table that illustrates the data collected for experiment 1 out of 10 for the glucose in distilled water solution. The data in Table 1 is paired with corresponding data for the distilled water solution with no glucose. Since approximately 200 intensity measurements are made for each wavelength, and each test is repeated 10 times, there are approximately 2,000 total tests per wavelength for glucose-distilled water, and 2,000 tests for distilled water. The measured intensity data for the glucose-distilled water solution is normalized to the measured intensity data for distilled water alone, thus removing the effects of any possible factors that have an effect on the intensity measurement results for glucose. The data processed in this manner represents a more accurate measurement of the glucose concentration. Note that for each experiment, the intensity measurements for the glucose solutions between 40 mg/dL and 250 mg/dL is collected for 18 wavelengths. Table 1 illustrates how the data is classified into 21 classes within the glucose concentration range between 40 mg/dL and 250 mg/dL. Each class includes data associated with a glucose concentration range of 10 mg/dL.

TABLE 1. Abbreviated table illustrating the data collected for the glucose-distilled water solution in one experiment.

Index	Glucose level (mg/dL)	Wave1	Wave2	…	Wave18	Class label
1	40	701	305	…	193	1
2	40.9	701	305	…	193	1
. .	. .	. .	. .	…	. .	. .
11	49.8	699	305	…	192.7	1
12	50.3	695	305	…	192.7	2
. .	. .	. .	. .	…	. .	. .
20	59.7	698	304		192.1	2
21	61.3	698	304	…	192.1	3
. . .	. . .	. . .	. . .	…	. . .	. . .
200	250	690	297	…	178	21

III. IDENTIFYING OPTIMAL WAVELENGTH FOR GLUCOSE MEASUREMENTS

The optimal wavelengths suitable for glucose prediction are identified by analyzing the intensity data measured by the Triad Sensor for 18 wavelengths. The intensity data is converted from analog to digital form using an AS7265x integrated analog to digital converter. The intensity data for each of the 10 experiments is read by programs in Arduino IDE software and input into MATLAB for signal processing. The data set is filtered using a finite impulse response low pass filter with an order of 34 and a bandwidth of 100 mHz. For each experiment and at each wavelength, each measured data point for distilled water is divided by each data point for the glucose-distilled water solution, resulting in normalized data,

in the form of a matrix similar to Table 1. The logarithmic of the normalized data is then calculated for each of the 10 experiments.

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

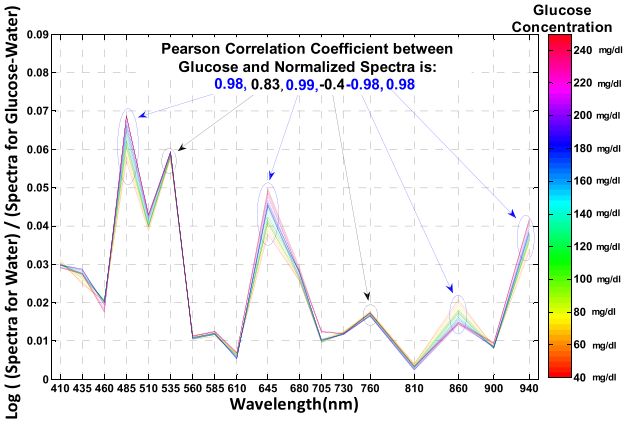

FIGURE 2. Average of the normalized spectra for different glucose concentrations.

The Pearson correlation coefficient measures the linear relationship between the glucose concentration values and the average values of the normalized intensity at each wavelength. The intensity of light measured at four wavelengths of 485, 645, 860 and 940 nm results in the highest correlation coefficient of 0.98. These wavelengths are the optimal wavelengths to detect glucose concentration molecules in aqueous solutions.

The light intensity measured by the detector is influenced by (1) light scattering as well as (2) light absorption by the glucose and water solution, and both properties are dependent on the wavelength of the incident light. A positive correlation coefficient for the 485, 645, and 940 nm wavelengths indicates that an increase in the glucose concentration leads to an increase in the absorption of light at these wavelengths, and this causes a decrease in the intensity of light measured by the detectors. In this case, the values associated with the vertical axis of Fig. 2 have a positive correlation with glucose concentration and thus have positive Pearson correlation coefficients.

An increase in the glucose concentration can lead to a decrease in the scattering coefficient of the sample depending on the wavelength of the incident light [32]. The scattering coefficient has an influence on the overall deviation of the light from its original path. An increase in the glucose concentration can lead to a change in the direction/angle of the scattered light so that the light intensity received by the detector increases, and this results in a negative Pearson correlation between the glucose concentration and the normalized light intensity (vertical access in Fig. 2). The intensity measured by the detector is also dependent on the position of the detector with respect to the light source, and this can vary from one measurement set-up to another.

Another factor that may have an effect on the sign of the Pearson correlation coefficient is the water displacement effect. The addition of glucose to the water solution displaces a fraction of the absorbing water molecules and reduces the molar concentration of the water molecules (this is known as the water displacement effect) [33]. This increase in glucose concentration can lead to a decrease in the total absorption of light by the solution depending on the wavelength of the incident light, and this may result in a higher light intensity at the detector.

IV. GLUCOSE PREDICTION MODELS USING MACHINE LEARNING METHODS

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

Five different machine learning models are trained using classification and regression methods as outlined in Figure 3, and are compared in terms of their ability to generate accurate glucose predictions. The regression methods consist of multiple linear regression (MLR) and feed forward neural network (FFNN) and the classification methods include k-nearest neighbor (KNN), decision tree (DT) and support vector machine (SVM).

In machine learning, hyperparameters need to be optimized to improve the model performance. In this study, a finite set of values are specified to a given hyperparameter, and

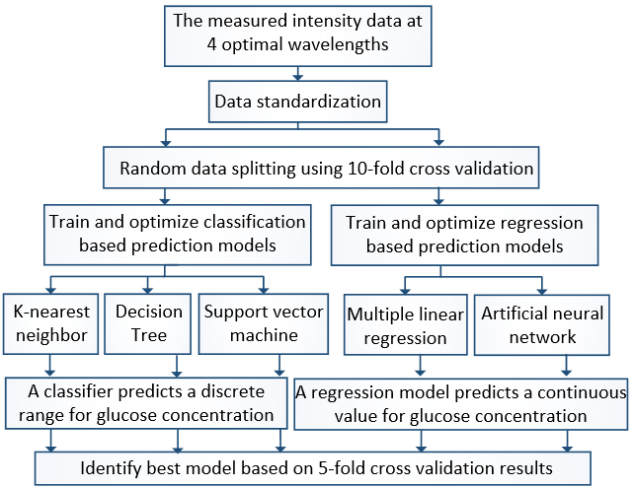

FIGURE 3. Block diagram representing glucose prediction approaches.

a 10-fold cross validation (10-fold CV) method is used to find the best value for the hyperparameter. Based on the 10-fold CV, the data set is randomly split up into 10 ''folds,'' where each is roughly the same size. A machine learning model is trained using data from the first 9 folds (the training set), and the performance of the model is evaluated on the 10th fold (the validation set). This process is repeated until each of the ten folds has served as a validation set, while the rest have served as a training set. The average prediction accuracy from the 10 different validation sets is calculated for each value of the hyperparameter. The averaged results are then interpreted to select the best value of the hyperparameter that results in the best fit to the data set. This process is repeated for each of the 5 models in Figure 3.

After the performance of each model is improved by finding the optimum hyperparameter, the 5 models are compared to identify the best model. A 5-fold cross validation method is used to evaluate the performance of each model by measuring the average of the prediction accuracy achieved on 5 different validation sets. The size of each validation set is 20% of the total data points and the remaining 80% of the data is used as a training set. The model is evaluated using a validation set (or unseen data) in order to estimate the ability of the trained model to generalize from a training set to previously unseen data.

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

A. GLUCOSE PREDICTION MODELS USING REGRESSION METHODS

1) MULTIPLE LINEAR REGRESSION

The multiple linear regression (MLR) method is used to predict the value of an output target based on a linear combination of scaled input variables. The input variables for this work are the measured intensity at the four optimal wavelengths (denoted by x1, x2, x3 and x4), and the output is the glucose concentration predicted value (denoted by $\hat{y}$). The relationship between the input and output variables is represented using a regression equation $\hat{y}^{(i)} = \theta_0 + \theta_1 x1^{(i)} + \theta_2 x2^{(i)} + \theta_3 x3^{(i)} + \theta_4 x4^{(i)}$, where $\hat{y}^{(i)}$ refers to the glucose predicted value for the i^{th} data sample. The objective of the model is to find the best parameters (θ) that will minimize the mean squared error (MSE) value. The error term, or residual, is the difference between the predicted ($\hat{y}$) and the actual (y) glucose concentration values. The most common methods used to find the optimal values of the parameters that will minimize the error are (1) ordinary least squares which uses linear algebra operations to compute the model parameters, and (2) iterative optimization, such as gradient descent, which gradually updates the model parameters until the prediction error is at a minimum value. The accuracy of a trained MLR model is evaluated using the performance metrics such as root mean squared error (RMSE) and coefficient of determination (R^2).

The MLR model can be an appropriate model for accurate predictions if the following requirements are met: (1) a linear relationship between the input and output variables exists, (2) there is little or no multicollinearity between the input variables, (3) there is a normal distribution of the residuals, and (4) homoscedasticity of the residuals exists [34], [35]. Otherwise, non-linear regression is an alternative. The first requirements can be checked by (1) creating a scatter plot of the output vs each input variable and assessing the linearity (1^{st} requirement), (2) computing the correlation coefficients between any two input variables to assess multicollinearity (2^{nd} requirement), and (3) creating a residual histogram to assess the normality of the distribution (3^{rd} requirement). The last requirement, homoscedasticity, can be checked by creating a scatter plot of the residuals vs predicted outputs (4^{th} requirement). The residuals should be symmetrically and randomly distributed around the zero line on the plot of residuals vs predicted outputs, and a deviating trend to the data should not exist.

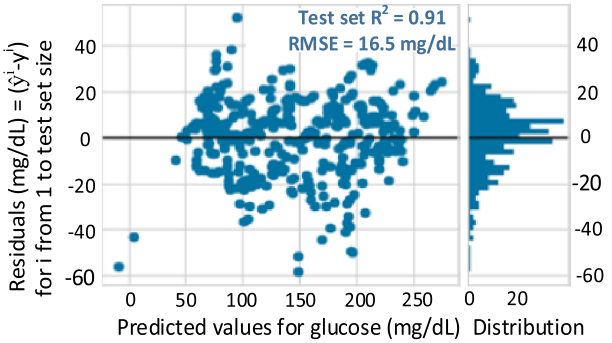

FIGURE 4. MLR residuals vs glucose predictions (left side plot) to check for homoscedasticity, and the residual histogram (right side plot) to check for the normality of the distribution.

Further clarification of the two last requirements is illustrated in Fig. 4, which includes the residual plots for the MLR

model that is trained using the training set (80% of the total data) and is evaluated using the test set (20% of the total data). Fig. 4 shows that the residuals have an apparent random dispersion around the x-axis. Although there is not a clear trend, it appears that the residuals associated with the higher glucose predictions have a slight tendency to spread farther away from the x-axis. This indicates that the relationship between the input variables and the output target is not perfectly linear. From the right side of the plot, the histogram of the residuals has the shape of a normal distribution with the mean close to zero, but the R^2 value is 0.91, which indicates that there is room to improve the accuracy of the glucose model by using non-linear regression algorithms such as neural networks. Neural network is generally expected to give a better fit to the data set compared to MLR, and is evaluated in the next sub-section.

2) FEEDFORWARD NEURAL NETWORK FOR REGRESSION

Artificial neural networks is a subset of machine learning and uses a group of algorithms to solve both linear and non-linear problems. There are different types of neural networks, and each has its own structure and application, such as feed forward neural network (FFNN), recurrent neural network (RNN), convolutional neural network (CNN), hybrid neural network, etc. [36], [37]. In this study, FFNN is used for the purpose of predicting glucose concentrations.

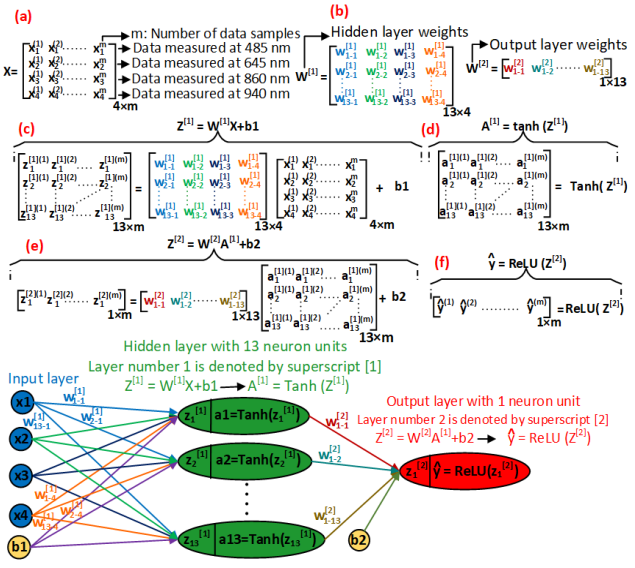

FIGURE 5. Feedforward neural network with one hidden layer.

FFNN consists of multiple processing units called neurons which are arranged in different layers. The network generally consists of an input layer, one or more hidden layers and an output layer. Each layer contains neurons which are linked together to model the data and make a prediction at the output layer. In this study, a neural network with one hidden layer is trained as illustrated in Fig. 5. The input layer is denoted as matrix X (see Fig. 5a) and includes a total of 'm' data samples and four variables (x1, x2, x3, and x4). The four variables coincide with the measured intensity at each of the four optimal wavelengths. The number of neurons in the hidden layer is tuned using a grid search method as explained at the end of this subsection, and is determined to be 13. The number of neurons at the output layer is equal to 1, and is the predicted glucose concentration, denoted by the vector $\hat{y}$.

The weight matrix for the hidden layer and the output layer are denoted by $W^{[1]}$ and $W^{[2]}$, respectively (see Fig. 5b). The training of the neural network model starts with a random initialization of the weights ($W^{[1]}$ and $W^{[2]}$) and biases (b1 and b2). The values for $W^{[1]}$ and $W^{[1]}$ are set between -0.5 and 0.5, and the values for b1 and b2 are both set to zero. The input data in the form of the matrix X is multiplied by the weights ($W^{[1]}$) that are assigned to the hidden layer. The resulting matrix from this step is added to the b1 bias and this creates a matrix in the hidden layer that is denoted by $Z^{[1]}$ (see Fig. 5c). Then, each neuron in the hidden layer takes part in the activation function calculations. In this network, the hyperbolic tangent (Tanh) function is used as the activation function for the hidden layer. The Tanh function performs a mathematical operation on $Z^{[1]}$ and transforms it into the matrix $A^{[1]}$ which is the output of the neurons in the hidden layer ($A^{[1]} = \text{Tanh}\,(Z^{[1]}) = (1-e^{-Z[1]}) / (1+e^{-Z[1]})$) as illustrated in Fig. 5d.

Similar calculations are required at the output layer. The output of the hidden layer $A^{[1]}$ is multiplied by $W^{[2]}$ and added to the bias b2 to create a matrix denoted by $Z^{[2]}$ as illustrated in Fig. 5e. The matrix $Z^{[2]}$ is then plugged into the rectifier linear unit ReLU, an activation function for the output layer, which results in $\hat{y}$ (see Fig. 5f). The ReLU function returns 0 if $Z^{[2]}$ is negative and for any positive value of Z [2] it returns that value back ($\hat{y} = \text{ReLU}\,(Z^{[2]}) = \max(0, Z^{[2]})$).

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

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

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

The mini batch gradient descent method splits up the training set into smaller sets that are called mini batches. We use mini batches with a size of 64, and the total number of mini

batches (denoted by n) is calculated by dividing the number of training samples by 64. When the algorithms run through all the mini batches, this is called one epoch, and for every epoch, the weights and biases are updated per each mini batch. Therefore, there will be a total of n updates for weights and biases at each epoch. The process of updating weights and biases are repeated during multiple epochs. In this work, the number of epochs is set to 5000 since after 5000 epochs there is no significant change in the error and thus the training can be stopped.

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

To speed up the training process for MBGD, the momentum technique is also used. In MBGD with momentum, the weights and biases are updated using the exponentially weighted average of the gradients (denoted by V_{dw} and V_{db}). The momentum parameter (denoted by beta) is usually given a value of 0.9. The process of training the neural network using MBGD with momentum, is shown as a flowchart in Fig. 6.

Different methodologies have been used to optimize neural networks, and to identify the optimal set of hyperparameters such as "learning rate" and "hidden unit size," as reviewed in [38], [39]. We use a 10-fold cross validation method and a grid search method to tune the learning rate and the hidden unit size. The grid search is a process in which a set of values are given to each parameter and then different models are trained for different combinations of the two parameters. Each trained model is then evaluated using 10 different validation sets to identify the best combination of the two parameters that results in a more accurate prediction. Fig. 7 illustrates a 4×7 optimization grid in which each cell is the average of the root mean square error (RMSE) achieved on 10 different validation sets. For a given learning rate, the average value of RMSE decreases with an increase in the size of the hidden layer units. We select 13 and 7×10^{-3} as the optimal values for the hidden unit size and the learning rate, respectively. This combination of the two parameters results in a neural network architecture with the smallest possible size of hidden units and a low enough RMSE.

B. GLUCOSE PREDICTION MODELS USING CLASSIFICATION METHODS

The measured intensity data at four optimal wavelengths is classified into 21 classes within the glucose concentration range from 40 mg/dL to 250 mg/dL. The first three classes

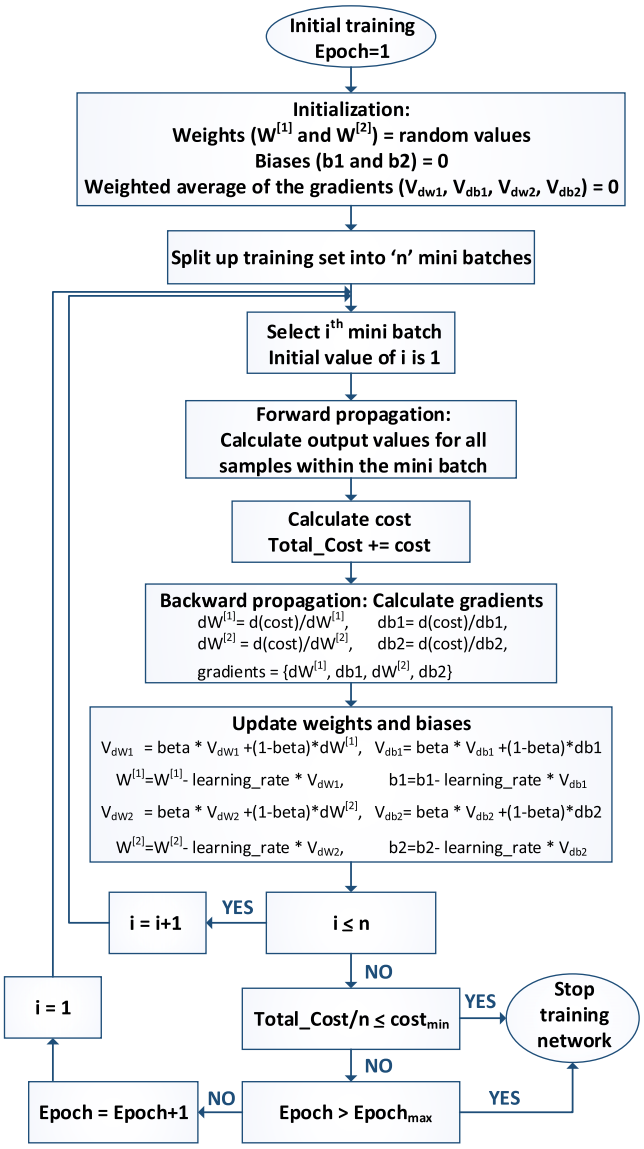

FIGURE 6. Training flowchart for feedforward neural network based on MBGD with momentum.

(1-3) include data points with glucose concentrations less than 70 mg/dL (hypoglycemic range), the next 11 classes (4-14) include data points with glucose concentrations between 70 and 180 mg/dL (normal range), and the last 7 classes include data points for glucose concentrations greater than 180 mg/dL (hyperglycemic range). The data points and class labels serve as input for the machine learning classification algorithms, and the classification models are trained to accurately predict the class labels (identification of the glucose range) of new cases (new measured data points). In this study, three classification techniques, the k-nearest neighbor, decision tree and support vector machine, are investigated. The performance of each classification model is evaluated using the jaccard Index, which is defined as the number of labels predicted correctly divided by the total number of predictions.

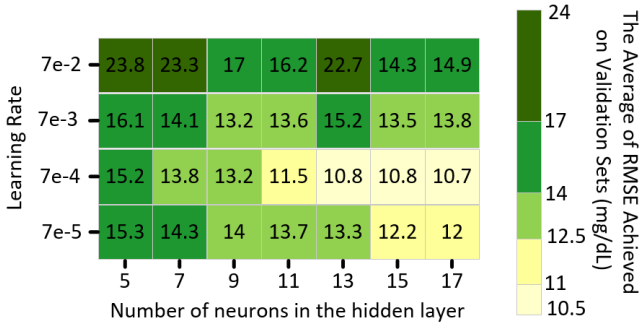

FIGURE 7. **Tuning hyperparameters of the neural network model using 10-fold cross validation and grid search methods.**

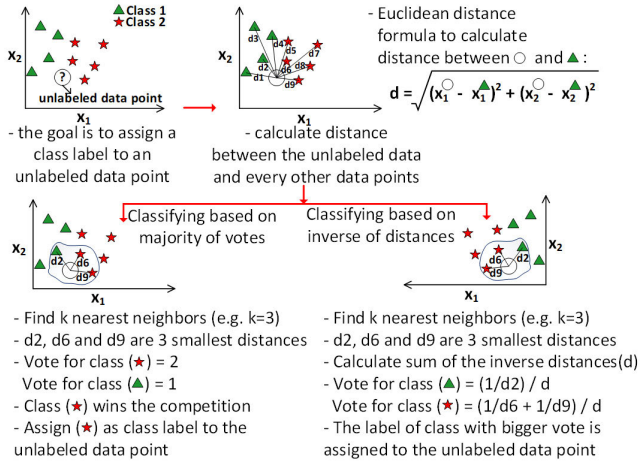

FIGURE 8. **Classification of unlabeled data points using k-nearest neighbor classifier based on two different rules: majority voting rule and inverse distance weighting rule.**

1) K-NEAREST NEIGHBOR CLASSIFIER

K-nearest neighbor (KNN) is a method that classifies unlabeled data points based on their similarity to other data points in close proximity. The similarity between two data points can be measured using different distance metrics such as Euclidean, Minkowsky, Manhattan, city-block, and Chebyshev [40], [41]. Fig. 8 illustrates the idea of classifying an unlabeled data point based on the KNN method for an arbitrary data set with two features (denoted by x1 and x2) and only 2 classes (depicted by star and triangle). To classify the unlabeled data point (depicted by a circle), the distances between the unlabeled data point and the labeled data points are computed. According to the distances, the number of k observations in the data set that are nearest to the unlabeled data point are identified. Then, a voting rule is used to classify the unlabeled data.

Two common and simple voting rules to classify the unlabeled data points are (1) the majority voting rule and (2) the inverse distance weighting rule. Both voting methods are illustrated in Fig. 8. In the majority voting rule, the most popular class labels from the k nearest neighbors (e.g. $k = 3$) are identified, where the distance is calculated based on the Euclidean formula. The votes of the k nearest neighbors are based on the class label and each vote has equal weight. This means there exists equality in the k nearest neighbors in the process of voting regardless of their distance from the unlabeled data point [42] (e.g. 2 votes for star and 1 vote for triangle). When using the majority voting rule, it is difficult to accurately classify the unlabeled data point if the value of k is large and if the neighbors are from multiple classes. To address this, the inverse distance weighting rule can be used to improve the prediction accuracy of the classifier. In this case, the votes of the k nearest neighbors are weighed according to their distance from the unlabeled data point. This is done by calculating the inverse of the Euclidean distance between the unlabeled data and each of the k nearest neighbors, and then normalizing the value with the summation of all the inverse distances. The class with the maximum normalized value wins the vote and the associated label is assigned to the unlabeled data point. Thus, the closest neighbors have much more decision power in determining the classification for the unlabeled data point.

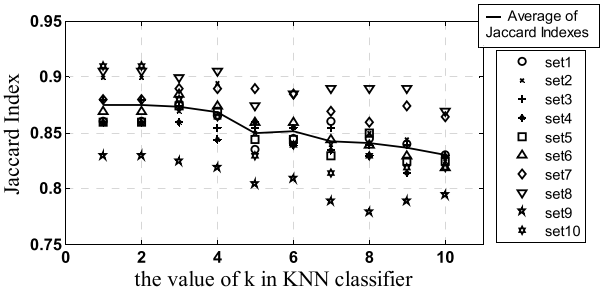

FIGURE 9. **Jaccard index vs k parameter in KNN for 10 different validation sets.**

The optimal value of k is identified using the ten-fold cross validation method. Fig. 9 illustrates the jaccard index as a function of k for 10 different validation sets. The average jaccard index for all ten validation sets is plotted as a solid line for each value of k nearest neighbors.

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

2) DECISION TREE CLASSIFIER

The decision tree (DT) model is a tree shaped diagram that represents all possible decision paths to classify data points, with a structure that includes decision nodes, branches and leaf nodes. A decision node represents a condition that is applied to an attribute, a branch corresponds to the result of a condition on the attribute, and a leaf node (or terminal node) represents a class label as the final prediction. Based

on our data, there are four attributes (x1, x2, x3, and x4) that correspond to the intensity data measured at the four optimal wavelengths.

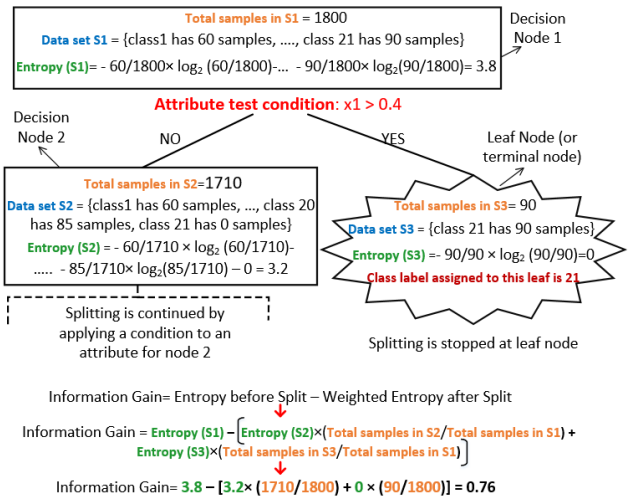

FIGURE 10. Decision tree classification.

The decision tree in Fig. 10 is used to illustrate the data classification process. The top node in the decision tree represents the entire data set S1 which is partitioned to create more nodes and leaves. The data set S1 is partitioned into S2 and S3 based on a condition that is applied on an attribute, such as x1>0.4, where 0.4 is an intensity value for the attribute x1. Attribute conditions need to be evaluated to identify the best condition. An optimal attribute condition is one that divides a single node into two nodes that are as pure as possible. The purity of a node can be measured by computing the entropy of the data set S within that node [43] as defined in (2).

$$Entropy(S) = -\sum_{i=1}^{c} p_i \log_2(p_i) \tag{2}$$

where p_i is the probability of data points that belong to the i^{th} class and are within the subset S, and c is the total number of classes. The purest node has an entropy equal to 0 and this will occur if all of the data points within the node all belong to the same class. The maximum entropy value is dependent on the value of c, and a node with a diverse class label distribution has the highest impurity or maximum entropy. Entropy measures the purity of a single node, and to select the best attribute, the purity of the entire split needs to be computed using metrics such as Information Gain. Information Gain is the entropy before the split minus the weighted entropy after the split. The weighted entropy represents the entropy for each node multiplied by a proportional factor based on the number of data points associated with each node. Fig. 10 includes the information gain calculations for the split based on the attribute condition x1>0.4.

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

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

The predictive performance of a DT model also depends on the DT hyperparameter values. These include (1) the maximum depth of a tree, denoted as max_depth, (2) the minimum number of samples required to split a decision node, denoted as min_samples_split, and (3) the maximum number of leaves in a tree, denoted as max_leaf_nodes [44]. These hyperparameters restrict the depth and width of a tree, and thus prevent the tree from extending beyond a certain level of complexity. More importantly, they improve the generalizability of the DT classifier to make an accurate prediction based on the test data.

TABLE 2. Hyperparameter tuning for decision tree using the results of 10-fold cross validation.

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

The three hyperparameters for the decision tree classifier are included in Table 2, along with a list of test values. The decision tree classifier is built for each possible combination of the three hyperparameter values, and is evaluated using the 10-fold CV method. The assessment of the prediction accuracy for each classifier is computed by averaging the values of the jaccard index for the 10 validation sets. The Pearson correlation between each possible hyperparameter and the average jaccard index is also calculated. A *max_leaf_node* resulting in the highest correlation coefficient of 0.7 has a significant impact on the prediction accuracy of the DT classifier. Based on the results in Table 2, a prediction accuracy of more than 75% occurs when the *min_sample_split* is not more than 10, the *max_depth* is at least 10, and the *max-leaf-node* is at least 200. By zooming in on these hyperparameter critical limits, the optimal values for the *min_sample_split*, *max_depth*, and *max_leaf_node* parameters are determined to be 8, 12, and 210, respectively. These values help to ensure a DT classifier with a relatively small prediction error for both the training

sets and the validation sets while limiting the complexity of the DT classifier.

3) SUPPORT VECTOR MACHINE CLASSIFIER

Support vector machine (SVM) algorithms classify data by mapping the data into high dimensional feature space and by finding a separating hyperplane (or decision boundary) that partitions the samples into different classes. SVM algorithms use a kernel function such as linear, polynomial, sigmoid, and radial basis function (RBF) to create a decision boundary. In this study, RBF is used as the kernel function due to its popularity as an effective kernel since it allows any shape associated with a decision boundary to make the distinction between different classes [45]. The two hyperparameters associated with the SVM classifier are (1) the kernel scale parameter denoted by γ (with values between 10^{-6} and 10^{2}), and (2) the regularization parameter denoted by C (with values between 10^{-2} and 10^{6}).

Increasing the value of C increases the risk for overfitting, which means that the classifier model cannot generalize from the training set to previously unseen data (test set). When overfitting, the classification model fits the training set perfectly, but has a poor fit on the test set, and this results in an unreliable prediction of class labels for the previously unseen data. On the other hand, lowering the value of C leads to more regularization which means that a more simple decision function is applied on the training set. When the model is highly regularized, there is a risk of underfitting. When underfitting, the classification model is not fitting well to the training set and performs poorly on the test set [46].

Increasing the value of γ leads to a greater curvature of the decision boundary (or separating hyperplane) and this increases the possibility of overfitting. Whereas, when the decision boundary is nearly linear (low curvature and small values of γ), the SVM model is simplified, resulting in a greater possibility of underfitting. Thus, when γ is large, the prediction accuracy on a test set can be improved by decreasing the value of C (because decreasing the value of C minimizes the risk of overfitting). Likewise, for very small values of γ, the prediction accuracy on the test set can be improved by increasing the value of C (because increasing the value of C minimizes the risk of underfitting) [46]. The optimization of C and γ can be visualized by examining the results in Fig. 11.

The optimal value of C and γ are identified using the grid search and 10-fold cross validation methods from the 6×6 optimization grid as illustrated in Fig. 11. Each cell represents the average of the ten jaccard indexes achieved on the validation sets (Fig. 11 (a)) and training sets (Fig. 11 (b)). The values of $C = 100$ and $\gamma = 1$ are identified as optimal values to ensure that a decision function is appropriately fitted on the training set and overfitting or underfitting is avoided on the training set. This combination of C and γ results in an accurate classification of data points within both the training and validation sets without adding unnecessary complexity to the SVM classifier.

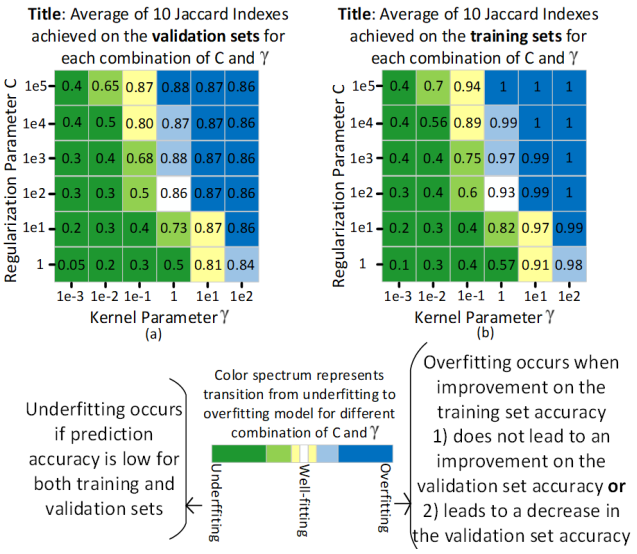

FIGURE 11. Average of ten jaccard indexes achieved on the (a) validation sets and (b) training sets for each combination of C and γ.

V. RESULTS

As described in section IV, hyperparameter tuning is necessary to optimize the performance of each model and improve the model's ability to generalize to unseen data. The focus of this section is to select the model that is best suited for the prediction of glucose concentrations by evaluating the predictive performance of each model using the 5-fold CV method.

A. EVALUATING THE ACCURACY OF GLUCOSE PREDICTION MODELS USING F1-SCORE AND CONFUSION MATRIX

Figure 12(a) illustrates the evaluation of the SVM classifier for one out of the five validation sets using the confusion matrix. The confusion matrix represents the actual class labels vs the predicted labels and is used to visualize the performance of the classifier. If an element in row i and column j of a confusion matrix is denoted by n_{ij}, then n_{ij} is the number of data samples that belong to i^{th} class and predicted to be in j^{th} class. Thus, each element on the main diagonal of the confusion matrix counts the number of data samples that are correctly classified.

The confusion matrix in Fig. 12(a) with 21 classes is transformed into a smaller matrix and reclassified into 3 classes as illustrated in Fig. 12 (b). Each of the three classes contains data samples that fall into one of the glycemic ranges: hypoglycemic, normal, or hyperglycemic. As illustrated in Fig. 12(a), the data samples within the first three classes are associated with the glucose concentrations between 40 and 70 mg/dL, and this range of glucose serves as an alert for hypoglycemia. By merging the data samples within the first three classes, we create a single class containing data samples that fall in the hypoglycemic range (the yellow cell in Fig.12 (b)). Merging the last 7 classes in Fig. 12(a)

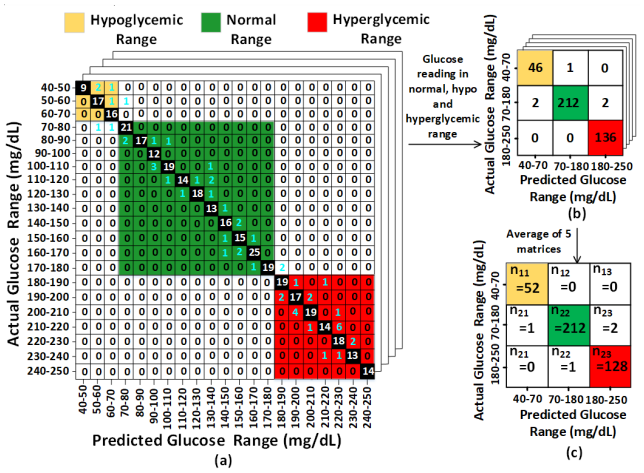

FIGURE 12. (a) The confusion matrix resulting from the SVM classifier that is evaluated using one of the validation sets (b) Actual vs predicted glucose within hypoglycemic, normal, and hyperglycemic ranges, and (c) average of 5 matrices resulting from the 5 validation sets.

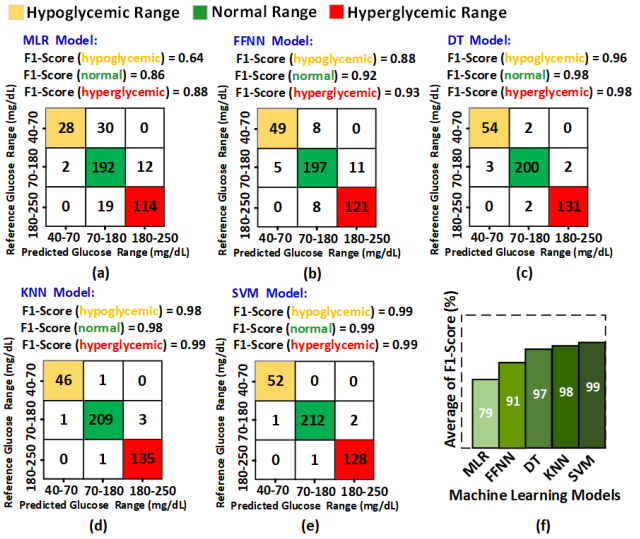

FIGURE 13. Visualization of the accuracy of glucose predictions in the hypoglycemic, normal, and hyperglycemic ranges when using (a) MLR, (b) FFNN, (c) DT, (d) KNN, (e) SVM based models, and (f) a bar graph comparing the different models based on the average of the F1-scores.

results in a single class that contains data samples associated with high levels of glucose concentrations between 180 and 250 mg/dL which are representative of cases with hyperglycemia (the red cell in Fig. 12 (b)). A class for the normal glucose range between 70 and 180 mg/dL consists of the reclassification of 11 classes into one (the green cell in Fig. 12 (b)).

The matrix in Fig.12 (b) helps visualize the accuracy of data classification for the hypoglycemic, normal and hyperglycemic ranges. Since there are 5 different validation sets, the average of 5 matrices resulting from the 5 validation sets is computed to create a single matrix labeled as Fig. 12(c). The average value for each cell of this matrix is rounded to the nearest integer. The F1-score is used to measure the accuracy of glucose predictions for each of the three classes ($class(i), i = 1, 2, or\ 3$), and is defined as the weighted average of the precision and recall for each class using the formula in (3).

$$\begin{aligned} \Pr ecision(class(i)) &= n_{ii} \Big/ \sum_{j=1}^{c} n_{ji} \\ Recall(class(i)) &= n_{ii} \Big/ \sum_{j=1}^{c} n_{ij} \\ F1\_score(class(i)) &= 2 \times \frac{\Pr ecision \times \text{Recall}}{\Pr ecision + \text{Recall}} \end{aligned} \tag{3}$$

where n_{ij} is the number of samples that belong to i^{th} class and are predicted to be in j^{th} class, and c is the total number of classes. Based on Fig. 12 (c), the value of c is 3, which is representative of the three glucose ranges (or classes): hypoglycemic, normal, and hyperglycemic.

The other two classifiers, DT and KNN, are evaluated in a similar fashion, resulting in a single matrix for each classifier, similar to the one in Fig.12 (c). The classifier predicts a discrete range of glucose concentrations, whereas the regression model predicts continuous values of glucose concentrations. Thus, the comparison of the classifier models to the regression models requires that every continuous prediction value from the regression models be converted to a discrete range (hypoglycemic, normal, or hyperglycemic). A matrix similar to the one in Fig. 12(c) is created for each regression model, as shown in Fig. 13. The confusion matrices resulting for the five different models along with an F1-score for each of the three classes are shown in Fig. 13(a) through (e). A bar graph in Fig. 13(f) compares the different models based on the average F1-scores. Classifiers demonstrate a higher capacity for accuracy and the ability to identify the hypoglycemia, normal and hyperglycemia conditions. Among the classifiers, SVM, with an average F1-score of 0.99, yields a better performance compared to the other prediction models. This indicates that the SVM classifier is able to predict the hypoglycemic and hyperglycemic ranges with the highest degree of accuracy.

B. EVALUATING THE ACCURACY OF GLUCOSE PREDICTION MODELS USING CLARKE ERROR GRID ANALYSIS

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

The prediction outputs from the regression models (multiple linear regression and feed forward neural network) are a series of continuous values for glucose concentration, whereas the prediction outputs from the classification models (k-nearest neighbor, decision tree, and support vector machine) consist of class labels that indicate a particular range of glucose concentrations. Thus, the class labels for the latter must first be converted to continuous values. A glucose

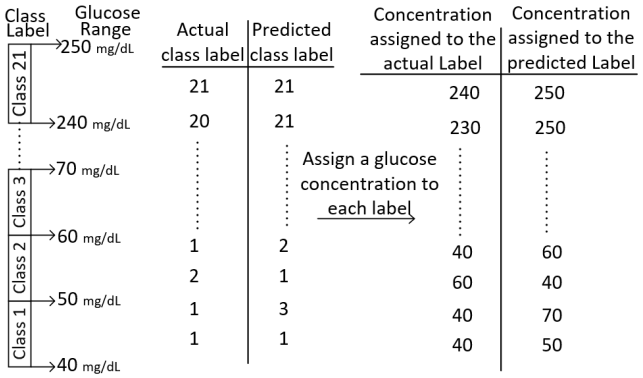

FIGURE 14. Conversion of class labels to continuous glucose concentration values.

value must be assigned to an actual class label and to a predicted class label for each sample in a validation set. The conversion of class labels to continuous values is carried out in a way that results in a maximum distance between the actual and predicted glucose concentration values based on the maximum and minimum concentrations associated with each class. For example, a data point that consists of an actual class label of 2 and a predicted class label of 1 is converted to 60 mg/dL concentration for the actual label and 40 mg/dL for the predicted label, respectively. This method re-classifies each data point to a concentration value based on a worst-case scenario.

The Clarke error grid includes 5 zones, identified as zones A-E, depending on glucose concentration prediction accuracy. A glucose prediction point will fall in zone A if a predicted value of glucose concentration is within 20% of the actual glucose level. Thus, the recommended treatment will be appropriate for the patient. Glucose predictions falling in zone B differ from the glucose reference levels by more than 20%, but the error would result in a clinically acceptable treatment for the patient. Glucose readings that fall in zone C are associated with a level of error that leads to unnecessary treatment. Glucose readings that fall in zone D include a level of error that may result in a failure in the detection and treatment of hypoglycemia and/or hyperglycemia. The worst case is when a prediction value falls in zone E since this level of error leads to a treatment that is opposite to what is actually required for the patient [1], [20]. The results of the Clarke EGA are included in Fig. 15 and quantifies the accuracy of each of the five prediction models.

The percentage of predictions falling in zone D (the zone of failure to detect) is the lowest (0.25%) for the SVM and KNN classifier models (see Fig. 15(e)). Among these two classifiers, SVM results in the highest agreement between the reference and predicted glucose level since 97.5% of glucose predictions fall within zone A (clinically accurate) and 2.25% fall within zone B (clinically acceptable).

VI. DISCUSSION

A VIS-NIR optical device is developed to measure the intensity of light transmitted through aqueous glucose solutions

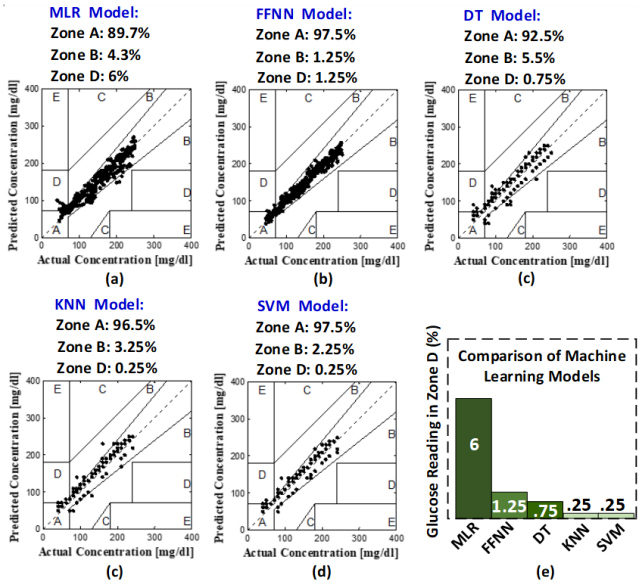

FIGURE 15. Clarke EGA of the glucose predictions when using (a) MLR, (b) FFNN, (c) DT, (d) KNN, (e) SVM based model, and (e) the Zone D bar graph representation of all 5 models.

using light sources between 410 and 940 nm. The use of multi-wavelength measurements and analysis extracts more quantitative information about the glucose concentration in solution, and improves the accuracy of non-invasive prediction of glucose values. In this study, four distinct wavelengths, 485, 645, 860 and 940 nm, result in a higher correlation between predicted and actual glucose concentration values.

A. GLUCOSE PREDICTION ACCURACY RESULTS FOR THE SOLUTION OF GLUCOSE AND WATER

Glucose prediction models using regression and classification models yield promising results. The regression models, MLR and FFNN, predict continuous glucose concentration values based on measured data, whereas classification models, KNN, SVM and DT, predict a class label. A class label encompasses a discrete 10 mg/dL glucose range. Identifying the best values of hyperparameters helps to optimize each model until it yields the most accurate prediction score. The performance of each model is evaluated using the 5-fold CV method, and the accuracy results obtained from 5 validation sets are compared in Table 3.

The FFNN regression model with an R-squared value of 0.96 and a RMSE of 11.1 mg/dL outperforms the MLR model. This is not surprising since the neural network model is able to learn complex non-linear relationships between its input and output values. Regression algorithms aim to minimize the error between actual and predicted glucose concentration values, but are not directly involved in minimizing the error in the prediction of hypo or hyperglycemia. Thus, it is possible for a regression model to make a prediction that erroneously identifies a value within the hypo or hyperglycemic ranges.

TABLE 3. The results obtained from 5-fold cross validation method for five glucose prediction models that are used to predict glucose within water.

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

On the other hand, classification algorithms learn how to classify data samples into the normal, hypo and hyperglycemic ranges. The prediction output of a classification model (or classifier) is a discrete glucose range, in comparison to the regression model that predicts continuous glucose concentration values. When the classifier predicts a correct glucose concentration range, the actual glucose concentration value lies somewhere within the predicted range. The difference between the highest and lowest value is the width of the range. The width of the range will depend on the number of classes selected for the entire range of data samples. In this study, data samples are grouped into 21 classes. Each class includes samples that correspond to a discrete 10 mg/dL glucose range, and each class falls into one of the glycemic ranges (normal, hypo or hyperglycemic). For example, the first class (class1) includes samples associated with a glucose concentration range between 40 and 50 mg/dL, and this range is associated with the hypoglycemic range. Likewise, the last class (class 21) includes samples that are associated with a glucose concentration range between 240 and 250 mg/dL, and this range lies within the hyperglycemic range. The partitioning of data samples into 21 classes enables the use of classification algorithms that have the ability to classify a new data sample into the correct glucose range. Thus, there is a high probability that a glucose prediction will fall within the glycemic range that contains the actual glucose value.

Table 3 includes a summary of the analysis of all five models. The classification models are associated with the highest F1-scores compared to regression models, and outperform the MLR model, which is the model commonly used for glucose predictions. A higher F1-score in the case of classifiers indicates an ability to correctly distinguish between the three glycemic ranges: hypoglycemic, normal, and hyperglycemic. Classifiers result in a lower percentage of glucose reading in zone D of the Clarke error grid, and this indicates a lower expectation of incorrectly detecting and treating hypoglycemia or hyperglycemia. It is very crucial to predict glucose values in the correct glycemic range, so diabetes patients can take appropriate actions to prevent life-threatening consequences.

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

B. VALIDATION OF PROPOSED METHODOLOGIES USING GLUCOSE AND ALBUMIN PBS SOLUTION

This paper presents basic systematic methods for accurately measuring glucose in water. All experiments in this preliminary work are applied to the measurements of glucose within a solution that assimilates the composition of human blood plasma. The three main components in plasma are water, protein (consisting mostly of albumin), and salt. A plasma sample can be mimicked by dissolving D-glucose in an aqueous solution containing physiological amounts of albumin and sodium chloride. In this work, 15 salt containing solutions are prepared with different albumin protein levels (4, 4.5, and 5 g/dL) by adding bovine serum albumin fraction V (Grainger Industrial Supply, USA) to a phosphate

buffered saline (PBS) solution with a pH of 7.4 (Grainger Industrial Supply, USA). A pH of 7.4 is close to the physiological pH value of human blood, and the dissolved albumin levels are within the corresponding physiological ranges (3.5-5.5 g/dL).

Fifteen experiments are set up to control the glucose concentration as it is reduced from 250 mg/dL to 40 mg/dL. The intensity data of the transmitted light at four optimal wavelengths (485, 645, 860, and 940 nm) are collected and are used to analyze the glucose with albumin PBS solutions. A SVM model is created and optimized using intensity data measured at the 4 optimal wavelengths in order to accurately predict the glucose concentration. The performance of the SVM model is evaluated using a 5-fold CV method which results in a RMSE of 12.5 mg/dL and an average F1-score of 0.99. Based on the Clarke error grid analysis, 99.55% of the glucose readings fall in zones A and B, and 0.45% in zone D.

TABLE 4. The results obtained from 5-fold cross validation method for the SVM model that is used to predict glucose within albumin PBS solutions.

Metrics / Model	RMSE (mg/dL)	Averaged F1-Score	Data percentage in the zones of the Clarke error grid
SVM	12.5	0.99	A: 98% B: 1.55% D: 0.45%

A comparison of the accuracy in the glucose predictions between the two types of solutions used in this study are shown in Tables 3 and 4 for the glucose with water compared to the glucose with albumin in the PBS solution, respectively. There is a slight increase in both the RMSE value and the percentage of readings in zone D for the predictions of glucose in the PBS solutions with albumin. It is expected that different levels of albumin in the PBS solutions can affect measured intensity, and thus affect the accuracy of glucose reading. Multiple wavelengths analysis using SVM helps to minimize the error that can result from different levels of albumin so that we see negligible changes in the accuracy results when we compare results in Table 3 (for glucose in water solution) and Table 4 (for glucose and albumin PBS solution).

In-vitro experiments performed in this study can be considered as a first and important step to demonstrate the feasibility of non-invasive glucose measurements of glucose in the presence of PBS solution with albumin. The next step is to test the sensor and methodologies on a body measurement site (e.g. finger). In-vivo experiments will require a change in the design of the VIS-NIR sensor so that it is miniaturized enough to be compatible as a wearable type sensor. Especially important in this work is to examine the relationship between in-vitro and in-vivo glucose detection by performing clinical studies and comparing the data from in-vivo to corresponding in-vitro solutions.

REFERENCES

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

[2] T. Battelino, T. Danne, R. M. Bergenstal, S. A. Amiel, R. Beck, T. Biester, E. Bosi, W. T. Cefalu, K. L. Close, and C. Cobelli, "Clinical targets for continuous glucose monitoring data interpretation: Recommendations from the international consensus on time in range," *Diabetes Care*, vol. 42, no. 8, pp. 1593–1603, Aug. 2019, doi: 10.2337/dci19-0028.

[3] K. Tonyushkina and J. H. Nichols, "Glucose meters: A review of technical challenges to obtaining accurate results," *J. Diabetes Sci. Technol.*, vol. 3, no. 4, pp. 971–980, Jul. 2009, doi: 10.1177/193229680900300446.

[4] B. W. Bequette, "Continuous glucose monitoring: Real-time algorithms for calibration, filtering, and alarms," *J. Diabetes Sci. Technol.*, vol. 4, no. 2, pp. 404–418, Mar. 2010, doi: 10.1177/193229681000400222.

[5] D. Wild, *The Immunoassay Handbook Theory and Applications of Ligand Binding, ELISA and Related Techniques*. Jordan Hill, U.K.: Elsevier, 2013.

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

[7] J. Y. Chen, Q. Zhou, G. Xu, R. T. Wang, E. G. Tai, L. Xie, Q. Zhang, Y. Guan, and X. Huang, "Non-invasive blood glucose measurement of 95% certainty by pressure regulated mid-IR," *Talanta*, vol. 197, pp. 211–217, May 2019, doi: 10.1016/j.talanta.2019.01.034.

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

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

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

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

[12] Y. Tanaka, T. Tajima, M. Seyama, and K. Waki, "Differential continuous wave photoacoustic spectroscopy for non-invasive glucose monitoring," *IEEE Sensors J.*, vol. 20, no. 8, pp. 4453–4458, Apr. 2020, doi: 10.1109/JSEN.2019.2962251.

[13] X. Xiao and Q. Li, "A noninvasive measurement of blood glucose concentration by UWB microwave spectrum," *IEEE Antennas Wireless Propag. Lett.*, vol. 16, pp. 1040–1043, 2017, doi: 10.1109/LAWP.2016.2618946.

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

[15] A. E. Omer, S. Gigoyan, G. Shaker, and S. Safavi-Naeini, "WGM-based sensing of characterized glucose-aqueous solutions at mm-waves," *IEEE Access*, vol. 8, pp. 38809–38825, 2020, doi: 10.1109/ACCESS.2020.2975805.

[16] J. Li, T. Igbe, Y. Liu, Z. Nie, W. Qin, L. Wang, and Y. Hao, "An approach for noninvasive blood glucose monitoring based on bioimpedance difference considering blood volume pulsation," *IEEE Access*, vol. 6, pp. 51119–51129, 2018, doi: 10.1109/ACCESS.2018.2866601.

[17] T.-L. Chen, Y.-L. Lo, C.-C. Liao, and Q.-H. Phan, "Noninvasive measurement of glucose concentration on human fingertip by optical coherence tomography," *J. Biomed. Opt.*, vol. 23, no. 4, p. 1, Apr. 2018, doi: 10.1117/1.JBO.23.4.047001.

[18] K. D. Pathirage, P. Roopasinghe, J. J. Sooriyaarachchi, R. Weththasinghe, and N. D. Nanayakkara, "Removing subject dependencies on non-invasive blood glucose measurement using hybrid techniques," in *Proc. 41st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Berlin, Germany, Jul. 2019, pp. 7197–7200, doi: 10.1109/EMBC.2019.8856391.

[19] T. Tajima, Y. Tanaka, M. Nakamura, and M. Seyama, "Multi-modality analysis of glucose aqueous solution using photoacoustic and dielectric spectroscopy for non-invasive glucose monitoring," in *Proc. Photons Plus Ultrasound, Imag. Sens.*, San Francisco, CA, USA, Mar. 2017, Art. no. 1006445, doi: 10.1117/12.2251049.

[20] M. Shokrekhodaei and S. Quinones, "Review of non-invasive glucose sensing techniques: Optical, electrical and breath acetone," *Sensors*, vol. 20, no. 5, p. 1251, Feb. 2020, doi: 10.3390/s20051251.

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

[22] J. Y. Sim, C.-G. Ahn, E.-J. Jeong, and B. K. Kim, "*In vivo* microscopic photoacoustic spectroscopy for non-invasive glucose monitoring invulnerable to skin secretion products," *Sci. Rep.*, vol. 8, no. 1, p. 1059, Dec. 2018, doi: 10.1038/s41598-018-19340-y.

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

[24] J. Yadav, A. Rani, V. Singh, and B. M. Murari, "Investigations on multisensor-based noninvasive blood glucose measurement system," *J. Med. Devices*, vol. 11, no. 3, Sep. 2017, Art. no. 031006, doi: 10.1115/1.4036580.

[25] X. Xiao, Q. Yu, Q. Li, H. Song, and T. Kikkawa, "Precise noninvasive estimation of glucose using UWB microwave with improved neural networks and hybrid optimization," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–10, 2021, doi: 10.1109/TIM.2020.3010680.

[26] S. Hu, S. Nagae, and A. Hirose, "Millimeter-wave adaptive glucose concentration estimation with complex-valued neural networks," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 7, pp. 2065–2071, Jul. 2019, doi: 10.1109/TBME.2018.2883085.

[27] M. Islam, M. S. Ali, N. J. Shoumy, S. Khatun, M. S. A. Karim, and B. S. Bari, "Non-invasive blood glucose concentration level estimation accuracy using ultra-wide band and artificial intelligence," *Social Netw. Appl. Sci.*, vol. 2, no. 2, p. 278, Feb. 2020, doi: 10.1007/s42452-019-1884-3.

[28] S. Shobitha, P. M. Amita, K. B. Niranjana, and M. A. M. Ali, "Noninvasive blood glucose prediction from photoplethysmogram using relevance vector machine," in *Proc. 3rd Int. Conf. Converg. Technol. (I CT)*, Apr. 2018, pp. 1–4, doi: 10.1109/I2CT.2018.8529481.

[29] K. Rogers, Sr., *Blood: Physiology and Circulation*. New York, NY, USA: Britannica Digital Learning, 2011.

[30] N. C. Dingari, G. L. Horowitz, J. W. Kang, R. R. Dasari, and I. Barman, "Raman spectroscopy provides a powerful diagnostic tool for accurate determination of albumin glycation," *PLoS ONE*, vol. 7, no. 2, Feb. 2012, Art. no. e32406, doi: 10.1371/journal.pone.0032406.

[31] R.-I. Stefan-Van Staden and G. Mitrofan, "Molecular enantiorecognition of L-glucose and D-glucose in whole blood samples," *Chirality*, vol. 30, no. 5, pp. 680–685, May 2018, doi: 10.1002/chir.22843.

[32] M. Kohl, M. Essenpreis, and M. Cope, "The influence of glucose concentration upon the transport of light in tissue-simulating phantoms," *Phys. Med. Biol.*, vol. 40, no. 7, pp. 1267–1287, Jul. 1995, doi: 10.1088/0031-9155/40/7/009.

[33] A. K. Amerov, J. Chen, and M. A. Arnold, "Molar absorptivities of glucose and other biological molecules in aqueous solutions over the first overtone and combination regions of the near-infrared spectrum," *Appl. Spectrosc.*, vol. 58, no. 10, pp. 1195–1204, Oct. 2004, doi: 10.1366/0003702042336136.

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

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

[36] O. I. Abiodun, M. U. Kiru, A. Jantan, A. E. Omolara, K. V. Dada, A. M. Umar, O. U. Linus, H. Arshad, A. A. Kazaure, and U. Gana, "Comprehensive review of artificial neural network applications to pattern recognition," *IEEE Access*, vol. 7, pp. 158820–158846, 2019, doi: 10.1109/ACCESS.2019.2945545.

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

[38] K. G. Sheela and S. N. Deepa, "Review on methods to fix number of hidden neurons in neural networks," *Math. Problems Eng.*, vol. 2013, pp. 1–11, May 2013, doi: 10.1155/2013/425740.

[39] T. K. Gupta and K. Raza, "Optimization of ANN architecture: A review on nature-inspired techniques," in *Machine Learning in Bio-Signal Analysis and Diagnostic Imaging*. Amsterdam, The Netherlands: Elsevier, 2019, pp. 159–182.

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

[41] V. B. Prasath, H. A. A. Alfeilat, A. Hassanat, O. Lasassmeh, A. S. Tarawneh, M. B. Alhasanat, and H. S. E. Salman, "Distance and similarity measures effect on the performance of k-nearest neighbor classifier—A review," *Big Data*, vol. 7, no. 4, pp. 221–248, Dec. 2019, doi: 10.1089/big.2018.0175.

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

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

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

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

[46] T. Hastie, S. Rosset, R. Tibshirani, and J. Zhu, "The entire regularization path for the support vector machine," *J. Mach. Learn. Res.*, vol. 5, pp. 1391–1415, Jan. 2004.

MARYAMSADAT SHOKREKHODAEI (Graduate Student Member, IEEE) received the M.S. degree in electrical engineering from the Sharif University of Technology, Tehran, Iran, in 2016. She is currently pursuing the Ph.D. degree in electrical and computer engineering with The University of Texas at El Paso (UTEP), El Paso, TX, USA.

Since 2017, she has been working in teaching and research with the Department of Electrical and Computer Engineering, UTEP. Her research interests include development of non-invasive biomedical devices and systems, artificial intelligence, and RFIC circuit design.

DAVID P. CISTOLA received the M.D. and Ph.D. degrees in medicine, biochemistry and biophysics from the Boston University School of Medicine, Boston, MA, USA.

He has held a faculty position with the Washington University School of Medicine, St. Louis, from 1989 to 2007, East Carolina University, from 2007 to 2012, the University of North Texas Health Science Center, Fort Worth, from 2012 to 2016, and Texas Tech University Health Sciences Center El Paso, since 2016. He currently serves as Professor and Director of the Center of Emphasis in Diabetes and Metabolism. His research is on the discovery and development of new screening tools for cardiometabolic health and disease risk using portable magnetic resonance technology.

ROBERT C. ROBERTS (Member, IEEE) received the B.S. and M.S. degrees in systems and control engineering and the Ph.D. degree in electrical engineering from Case Western Reserve University, Cleveland, OH, USA.

He worked as a Researcher with The University of Hong Kong, Hong Kong, from 2012 to 2018. He is currently an Assistant Professor with the Department of Electrical and Computer Engineering, The University of Texas at El Paso. His research interests include microelectromechanical systems (MEMS) and additive manufacturing, particularly the use of additive manufacturing for microfabrication in the areas of packaging, energy, and healthcare.

STELLA QUINONES received the M.S. degree in metallurgical and materials engineering and the Ph.D. degree in materials science and engineering from The University of Texas at El Paso (UTEP), El Paso, TX, USA.

She was an Associate Professor with the Department of Electrical and Computer Engineering, UTEP, from 2004 to 2018. She is currently the Chair of the Department Metallurgical, Materials and Biomedical Engineering. Her research interests include non-invasive biomedical sensors, 3D printing, and engineering education.