Original Article | *Peer Reviewed* | *Open Access*

Non-Invasive Near-Infrared-Based Optical Glucose Detection System for Accurate Prediction and Multi-Class Classification

Check for updates

M Naresh and Samineni Peddakrishna[*]

School of Electronics Engineering, VIT-AP University, Amaravati-522237, Andhra Pradesh, India

E-mail/Orcid Id:

MN, manda.naresh422@gmail.com, https://orcid.org/0009-0007-5229-1929;
SP, krishna.samineni@gmail.com, https://orcid.org/0000-0002-5925-8124

Abstract: One of the most common diseases around the world is diabetes. Intrusive methods involving blood samples via a finger prick are required to test for diabetes. These treatments are uncomfortable and prone to infection. Non-invasive testing is proposed as a solution to this concerning problem. To test the glucose levels of subjects, a shortwave near-infrared-based optical detection system with a 950 nm wavelength sensor in reflective mode is presented. The system collects the measured signal through voltage, transmittance, absorbance and reflections to estimate glucose. The relation between voltage and predicted glucose is evaluated from the absorbance, reflectance, and voltage for 575 samples. A Multiple linear regression (MLR) expression is used in the proposed method to enhance the accuracy. The proposed method achieves a coefficient of determination (R^2) of 99% and a mean absolute derivative of 3.6 mg/dl in real-time data analysis with the sensor. The root mean square error (RMSE) is also calculated as 3.46 mg/dl. Three additional machine learning classifiers are employed to achieve high accuracy in multi-class classification. Adaboosting and Gaussian Naïve Bayes classifiers achieve an accuracy of 97% each. Furthermore, the system computes performance metrics such as precision, recall, and F1-score, and predicts the class on the test sample.

Introduction

According to statistics from the World Health Organization, the number of diabetic patients has doubled since 2015. In 2019, it was estimated that 9.3% (463 million people) had diabetes, and its prevalence is projected to increase to 10.2% (578 million) and 10.9% (700 million) by 2030 and 2045, respectively (Sun et al., 2022). In recent years, some glucose monitoring methods have been developed. It can be categorized into three main categories: invasive, minimally invasive, and noninvasive. The most used method of checking blood glucose levels is to prick the finger with a conventional blood glucose meter via an invasive method (Gusev et al., 2020). However, no matter how tiny or thin the needle is, it causes the patient pain, making the procedure difficult to incorporate into their daily lives. Additionally, invasive glucometers are not cost-effective (Yeaw et al., 2012) because they come with single-use strips that must be replaced once used. Alternatively, minimally invasive techniques causing little skin damage may be used (Chen et al., 2017). This method requires calibration more frequently than traditional measurement methods. These devices are expensive and have stability and lifespan problems (Smith et al., 2015). Therefore, these devices are unsuitable for regularly monitoring blood glucose levels. Due to these reasons, different researchers have developed painless, accurate, and cost-effective noninvasive methods of measuring blood glucose (Van Enter and Von Hauff, 2018). In this way, regular blood glucose monitoring could become a more relaxed and comfortable experience for millions of people.

Some approaches have been proposed for non-invasive blood glucose detection, including in-vitro and in-vivo techniques (Jain et al., 2019). An in vitro approach involves studies or tests conducted outside a living organism, such as in a laboratory. In the in-vivo

method, the test is conducted on a living organism, which is more suitable for self-monitoring blood glucose levels. In view of the complexity of blood and tissue properties, optical technologies are particularly well suited to the detection of glucose in vivo. Another advantage of optical technologies is that they are less likely to irritate the targeted biological tissue. Various optical technologies, including visible laser light, Raman spectroscopy, mid-infrared (MIR), near-infrared (NIR), etc., have been used to measure from the user's perspective. The other two optical techniques, MIR and NIR, have received more research attention in recent years. NIR signal has wavelengths between 750 nm and 2500 nm, while a MIR signal has wavelengths between 2500 nm and 10000 nm. MIR penetrates only a few micrometers into human tissue, so it can only be used in the reflection mode (Heise, 2021). Therefore, NIR spectrometry is a suitable method for estimating blood glucose levels. In contrast to MIR, NIR light can penetrate through multiple layers of the skin and reach the subcutaneous vessels regardless of the pigmentation of the skin. Among these techniques, NIR spectroscopy has proven to be a useful method for determining glucose levels precisely (Goodarzi et al., 2015). The NIR spectrum is further divided into two methods: the long NIR spectrum and the short NIR spectrum. Compared to the long NIR, the short NIR has a deeper penetration capability beneath the skin, allowing it to detect glucose molecules more accurately. Thus, the proposed work focuses on the short-wave NIR reflectance spectroscopy technique at 950 nm with improved accuracy. The following sections discuss prior research and the novelty of the proposed approach.

In other words, NIR has a deeper penetration into the skin than most other infrared wavelengths. The NIR spectrum analysis can be categorized into two subcategories, which include the analysis of NIR spectrometry and the analysis of NIR Photoplethysmography (PPG) signals. As for NIR PPG signals must be acquired with NIR LEDs, whereas a NIR spectrometry signal must be analyzed by measuring voltage after absorption and reflection. A review is being conducted to summarize these two categories, emphasizing the machine learning analysis necessary to estimate glucose using NIR PPG signals (Hina and Saadeh, 2022). On the other hand, numerous studies in the literature have demonstrated that NIR bandwidths and characteristic spectra vary with blood glucose levels (Jintao et al., 2017; Yang et al., 2018; Lee et al., 2019). Further, NIR spectrometry can be divided into two regions based on their bandwidths: long-wave NIR and Short wave NIR. The NIR waves are partially scattered or

absorbed during penetration through the skin tissue. The scattering and absorption of molecules in a medium occur due to the vibrations of their chemical bonds. This phenomenon makes it possible to determine the glucose concentration bonds that contain the chemical formula $C_6H_{12}O_6$ (Pigman, 2012). The functional bonds of the glucose molecule, which consist of C-H and C-O, can be used to measure the absorption and reflectance of NIR waves to determine glucose concentration in the blood. During light absorption by biological tissue, glucose molecules are easier to detect using long-wave NIR. However, due to its shallow penetration, long-wave NIR will not provide better results for in-vivo tests (Uwadaira et al., 2016). On the other hand, short NIR waves are weakly absorbed by glucose molecules, but they can be used for in vivo testing due to their sharp penetration. A study in (Jain et al., 2019), used both shorter regions to estimate glucose levels. A total of three sensors have been used to operate at 940 nm and 1300 nm, out of which two 940 nm sensors operate in absorbance and reflectance modes, and one 1300 nm sensor is used in absorbance mode. According to these results, short NIR regions are more focused and studied to estimate blood glucose levels.

The NIR absorption peaks for glucose isomers such as fructose, lactose, and galactose are not coincident with glucose absorption at approximately 950 nm (Simeone et al., 2017). Hence, these isomers do not adversely affect the detection of glucose. Also, a significant glucose absorption spike can be seen in the NIR between 930 nm and 960 nm (Yadav et al., 2015). They can be used as reflection and transmission modes depending on the specimen type and human body part selection (Villena et al., 2019). In other words, the reflection configuration is preferred for thick and dense samples, while the transmission configuration is more effective for thin and aqueous samples. Moreover, the reflective configuration has an advantage from a wider selection of human body parts compared to the transmission configuration. In a previous study, NIR photodiodes with wavelengths of 940 nm and 950 nm were used to measure blood glucose concentration levels non-invasively (Abidin et al., 2013). According to this study, 950 nm was the preferred wavelength of light for passing through blood glucose concentrations more effectively than 940 nm. Another method utilizes a 950 nm reflective sensor and a signal conditioning component with a 96% accuracy. This method requires 9 volts of power to measure glucose (Anupongongarch et al., 2019).

On the other hand, multiple linear regression (MLR) is a statistical technique used to analyze the relationship

between multiple independent variables and a dependent variable (Montgomery et al., 2021). When the relationship between the dependent variable and independent variables is not linear but exhibits a nonlinear pattern, such as exponential growth or decay, the logarithmic form is often considered to transform the relationship into a linear form. Therefore, this approach could be combined with the short-wave NIR technique to enhance glucose monitoring accuracy further. Moreover, recently, machine learning (ML) has the potential to revolutionize healthcare by improving disease diagnosis (Chandrasekhar and Peddakrishna, 2023), the machine learning classification of diabetes (Teki et al., 2021) using binary classification performed for PID dataset and personalized treatment (Rajkomar et al., 2019). Various ML techniques have been proposed in six machine learning classifications performed for binary class (Miriyala et al., 2022) to predict glucose. To extract relevant features from measured or predicted data. A real-time emotion identification system using ECG and temperature sensors with machine learning. Random forest (RF) has been employed to analyze continuous glucose monitoring data to predict the occurrence of hypoglycemic events in type 1 diabetes patients (Haak et al., 2017). Moreover, KNN algorithms have been used to classify glucose data based on their similarity to previously observed patterns (Ali et al., 2020). Thus, using machine learning techniques in glucose prediction and classification has yielded encouraging outcomes and this approach holds significant potential for enhancing diabetes management.

This study is to develop a glucose prediction and classification approach by utilizing a combination of ML, shortwave NIR techniques, and MLR. The data collected from the sensor is utilized in MLR, which accurately predicts glucose levels. It explores the correlation between glucose concentration and signal absorbance/transmittance, using MLR to achieve high accuracy. Furthermore, ML algorithms are utilized to categorize glucose levels into multi classes, such as normal, hyperglycemic, and hypoglycemic, using the spectral data obtained from the non-invasive short-wave NIR technique, which measures glucose levels. This combined approach allows for the development of a reliable and accurate glucose monitoring system that can be employed for diabetes management. In order to improve the accuracy of the previous work, a continuous glucose monitoring system using NIR spectroscopy is presented. This system employs a 950 nm reflective sensor that is capable of measuring an accuracy of 99%. This accuracy is achieved by using an MLR. To

determine the accuracy of the proposed method, 184 subject samples are considered. Additionally, the present work discussed the relationship between glucose concentration and signal absorbance and transmittance. The following section discusses the proposed method and its implementation.

Design and Implementation

The reflective glucose sensor at a wavelength of 950 nm is used to examine the variations in the sample's optical properties. Reflective sensor mode measures the quantity of light reflected from a finger with the aid of TCR1000. However, these methods may produce varying baseline values due to differences in the optical properties path. Therefore, it is critical to determine the appropriate baseline values for the sensor and calibrate it accordingly to guarantee accurate and precise measurements. To implement this system, an optical sensor TCRT1000, a precise ADC converter, and a microcontroller computation unit are needed, as shown in Fig. 1. When an IR-emitting LED comes into contact with a finger from TCRT1000. The reflected light can be used to detect glucose-induced energy absorption in TCRT1000. A current-limiting resistor ($R_1 = 340\ \Omega$) must be added in series to protect the IR LED. The circuit design must include a resistor ($R^2 = 47\ k\Omega$) in series with the collector for the light receiver. This will limit the current going through the phototransistor to prevent its destruction. The signal produced from the sensors is fed to an ADS1115 connected to single-ended inputs to A0 of the sensor output reflective sensor respectively. Here it is calibrated with a gain of two third and interfaced with the microcontroller using the I2C bus protocol.

The signal received is converted into millivolts to predict blood glucose. This is computed by the microcontroller using an Algorithm, as in Table 1, which extracts the necessary data, such as voltage (x_1), transmittance (x_2), absorbance (x_3), and reflectance (x_4), to predict glucose concentration for the 950 nm sensor.

The obtained transmittance, absorbance, reflectance, voltage values, and baseline calibrated value with reference device were analyzed. In order to calculate the optical density (OD) or Absorbance of the human finger medium, it is necessary to measure the transmittance of the light (T), which is the ratio of transmittance voltage (Vt) from the medium to the incident voltage (Vi) from the LED source. This can be represented by Equation (1).

$$T = \frac{Vt}{Vi} \quad \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots (1)$$

The optical density (OD), also known as absorbance of human finger medium, can be represented in Equation (2).

$$OD = -\log_{10}(T) \quad \text{.....................(2)}$$

Total infrared signal emitted by the sensor is equal to the sum of absorbance, transmittance and reflections is given by Equation (3),

Absorbance + Transmittance + Reflections=1……(3)

The relationship between voltage, transmittance, absorbance, and reflections of the infrared signals or loss from the finger and sensors are evaluated. Figure 2 and Figure 3 show the relationship between measured voltage and predicted glucose from the reflective sensor with a

Table 1. Algorithm for computation of features

Algorithm
Input: adc0
Output: X1, X2, X3 X4- Input features for Prediction of glucose
Step 1: Read the sensor value from finger to adc0 ADS1115
Step 2: initialize mv ←0, R ←4.08(offset value)
Step 3: convert sensor value into millivolts. $mv \leftarrow (adc0*0.1875)/100$
Step 4: Get 10 sample values from the sensor to smooth the value. $a[i] \leftarrow mv$
Step 5: sort the data from small to large. $b \leftarrow a[i]$ $a[i] \leftarrow c[j]$ $c[j] \leftarrow b$
Step 6: take the average value of 6 center samples. $d+ \leftarrow a[i]$ $e \leftarrow d/6$
Step 7: To get the voltage value in V from the averaged samples. $x_1 \leftarrow e$
Step 8: Calculate the Transmittance of the signal. $x_2 \leftarrow x_1/R$
Step 9: Calculate the Absorbance of the signal. $x_3 \leftarrow -log(x_2)$
Step 10: Calculate the Reflectance of the signal. $x_4 \leftarrow 1-x_2-x_3$
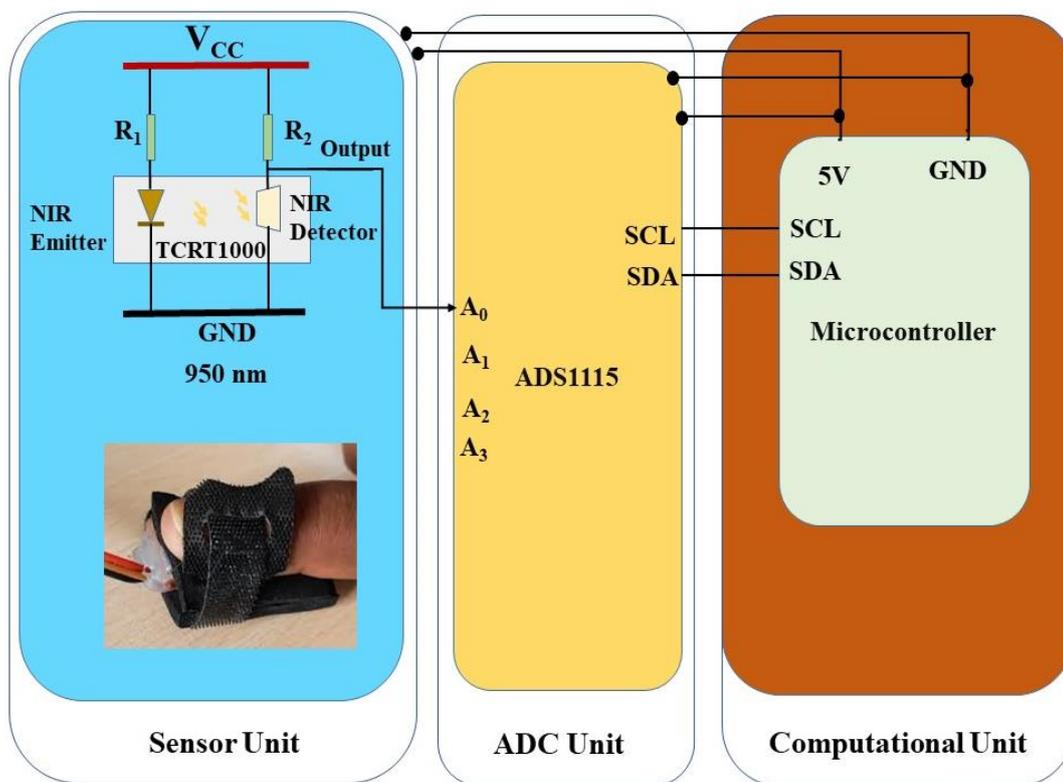
Figure 1. Block diagram of the proposed method
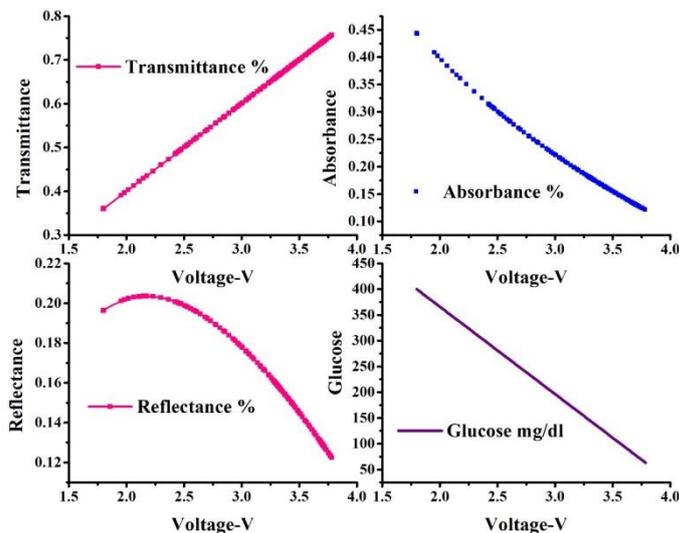
Figure 2. Illustrate the relation between voltage and measured features

950 nm wavelength. In Figure 2, light transmittance increases proportionally as the voltage increases, depending on the glucose concentration in the subject's finger. As the voltage decreases, absorbance increases, depending on the glucose concentration present in the subject's finger, respectively. The minimum and maximum voltage values obtained were 1.79V and 3.78V, respectively. The minimum and maximum absorbance values were 0.443 and 0.120, respectively. Similarly, the minimum and maximum transmittance values were 0.359 and 0.757, respectively, while the minimum and maximum reflection values were 0.196 and 0.126, respectively. Similarly, From Figure 3, the predicted glucose ranges minimum and maximum concerning their absorbance, transmittance, and reflections of the sensor. Here, as the predicted glucose increases, the absorbance of the signal also rises, whereas the transmittance of the signal decreases with decaying behavior. Reflection of the signal shows lossy behavior to the transmitted signal.

Glucose Prediction Using MLR Method

The multiple linear regression (MLR) method is utilized in this study to predict the glucose concentration value by creating a linear combination of input variables such as the measured voltages, transmittance, reflectance, and absorbance for 950 nm (represented by x_1, x_2, x_3 and x_4, respectively). The output value is the predicted glucose concentration (y) based on the reference Dr trust glucometer measurement. The MLR model involves fitting a linear equation to the data, with the NIR measurements at 950nm as independent variables and the glucose value as the dependent variable. The ordinary least squares method is commonly used to find optimal parameter values that minimize the error. The MLR model is considered appropriate for accurate glucose concentration prediction. To minimize the error between the actual glucose concentration values and the predicted values, the MLR model undergoes iterative optimization, searching for the optimal coefficient values. The ordinary least squares (OLS) method is frequently used to determine the optimal parameter values. The iterative optimization process entails performing the MLR modeling process multiple times until the optimal coefficient values are achieved. In this study, MLR is applied to four independent predictors to find the prediction (ln y) in natural logarithmic form because the input predictors demonstrate an exponential increase and decay with respect to voltage, as shown in equation (4).
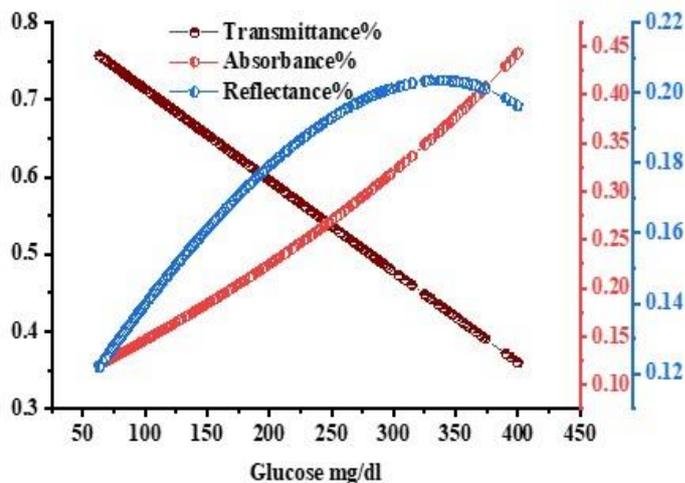
Figure 3. Shows predicted glucose and measured features concerning the reflective sensor

$$\ln(y) = 4.105 + 6.33\ln x_1 + 3.475\ln x_2 + 9.005\ln x_3 - 5.587\ln x_4$$
------ (4)

$$\ln(y) = -1.501 + 9.73\ln x_1 + 9.005\ln x_3 - 5.587\ln x_4$$
------ (5)

The product and power rules with exponential were applied to find y and fit the model, with reference to the ẏ given in equation (6). This further improved the R squared value, which is now 99%.

$$y^{\hat{}} = 0.223x_1^{9.737} x_3^{9.005} x_4^{-5.587} \quad \dots\dots\dots\dots\dots(6)$$

A scatter plot was created for the output versus each input variable, and linearity was assessed with reference to the device, as shown in Fig. 4. The correlation determination coefficient between the input variables and the target variable was calculated at 99% and the root mean square error (RMSE) is 3.4.
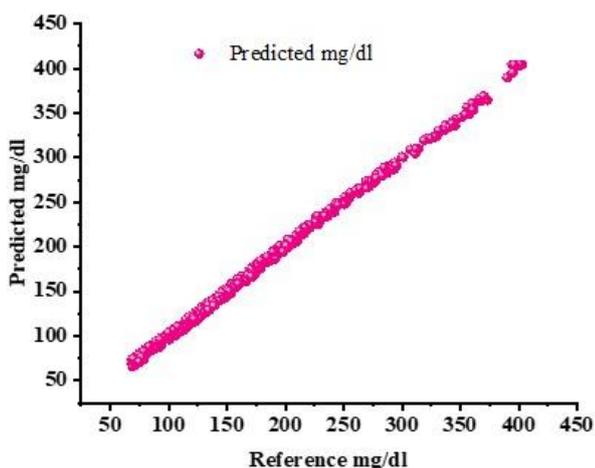
Figure 4. Depicted the reference (Dr. Trust) and predicated values are linear

The technique characterizes the relationship between voltages from the sensor and predicts blood glucose concentration with respect to the reference glucose device (Dr. Trust's glucometer). The detector input features result is an independent variable related to the expected glucose response of the 950 nm sensor. The proposed model was developed with 575 samples, 289 subjects aged 19-69, collected randomly in blood glucose test mode. Precision was evaluated based on the mean absolute relative difference (mARD), mean absolute deviation (MAD), RMSE, and average error. With the proposed method minimizing overall error. Performance is evaluated using equations (7), (8), and (9).

$$MARD = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{BG_{\mathrm{Re}f} - BG_{\mathrm{Pr}e}}{BG_{\mathrm{Re}f}}\right| \times 100 \dots\dots\dots\dots(7)$$

$$MAD = \frac{1}{n}\sum_{i=1}^{n}\left|BG_{\mathrm{Re}f} - BG_{\mathrm{Pr}e}\right| \dots\dots\dots\dots\dots(8)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left|BG_{\mathrm{Re}f} - BG_{\mathrm{Pr}e}\right|^2} \dots\dots\dots(9)$$

BGPre and BGRef represent the predicted and reference blood glucose concentration values, respectively. With n = 575 samples, the MARD is 3.6%, MAD is 2.91 mg/dl, and RMSE is 3.46 mg/dl, indicating high precision.

$$AvgErr = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{V_{md} - V_m}{V_{md}}\right| \times 100 \dots\dots\dots(10)$$

The minimum deviated value Vmd and the measured value Vm is used to calculate the AvgErr using equation (10), which is found to be 3.73 %. The coefficient of determination (R^2) is also calculated and found to be 0.99. To assess the clinical accuracy of the proposed system, a Clarke error grid (CEG) analysis is performed, and the glucose values are shown in Figure 5. The measured glucose values fall within the clinically accepted zone, known as Zone A.
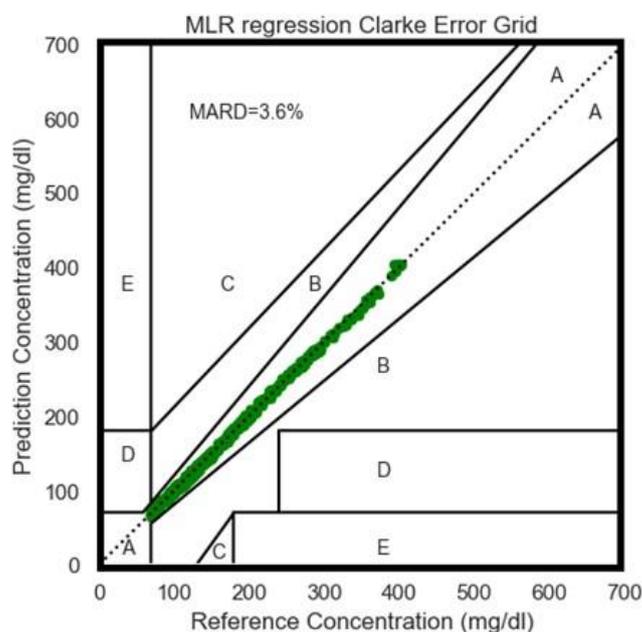
Figure 5. CEG for train data with respect to reference and predicted glucose

Table 2 presents a comparison of the proposed method with existing literature. Jain et al. (2019) use three different NIR sensors, while (Anupongongarch et al., 2019) employs only one sensor. Larin et al. (2002) utilized OCT technology, but most parameters were not measured. Song et al. (2015), despite using two different technologies, the average error rate was still at 19%. Photoacoustic technology was employed with CEGs of 93% and 100%, respectively (Pai et al., 2017a & 2017b), but it is known to be very expensive. Visible laser light technology was used (Ali et al., 2017), and the average reported error rate was between 8% and 10%. In light of these findings, the proposed approach is deemed superior to the above technologies with a high R^2 value.

Table 2. Comparison of performance parameters with previous work

	Proposed Method	Jain et al., 2019	Anupongon garch et al., 2019	Larin et al., 2002	Song et al., 2015	Pai et al., 2017a	Pai et al., 2017b	Ali et al., 2017
R^2 value	0.99	0.908	0.96	0.95	-	-	-	-
mARD (%)	3.60	3.25	-	-	8.30	8.84	7.01	-
AvgErr (%)	3.73	3.77	-	-	19	-	-	8-10
MAD (mg/dl)	2.91	3.87	-	-	-	32.8	5.23	-
RMSE	3.46	5.61	11	-	-	43.64	7.64	-
CEG(A&B %)	100	100	-	-	100	93	100	98
Technology	NIR	NIR	NIR	OCT	Impedance and NIR	Photoacoust ic	Photoacou stic	Visible laser light
System cost	Cheaper	Cheaper	Cheaper	Costly	Cheaper	Costly	Costly	Cheaper
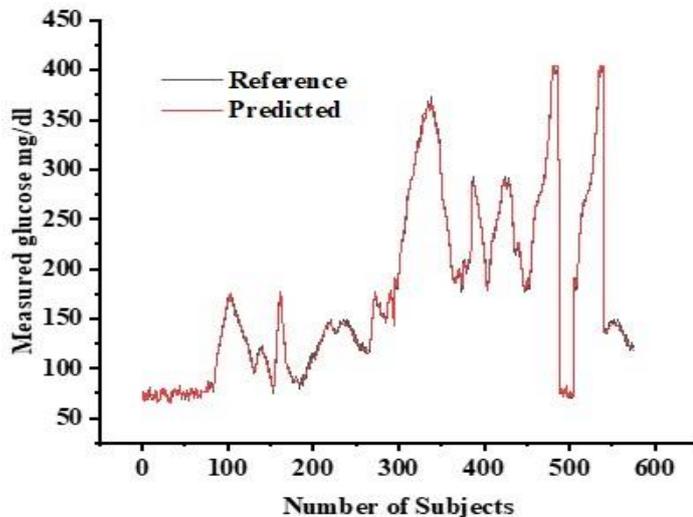
Figure 6. Measured glucose for reference and predicted sensor for the number of samples

Further, Figure 6 compares the measured glucose range of the reference and sensor results for the number of samples which was 575. In addition, Figure 7 demonstrates that both methods exhibit a glucose range of 60mg/dl to 430 mg/dl, indicating the ability of the proposed device to perform accurate measurements.
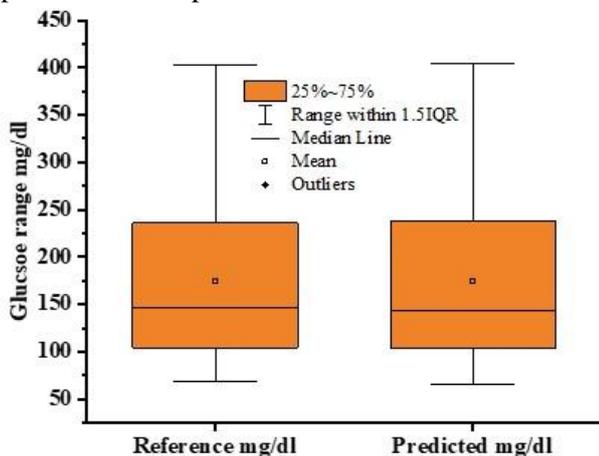
Figure 7. Measured glucose range for reference and predicted sensor

Glucose Multiclass Level Prediction Using Ml Methods

An ML classifier is utilized to categorize glucose levels into different categories based on predicted values. Multiclass classification data obtained from a 950 nm wavelength sensor is categorized into three classes within the glucose concentration range of 60 mg/dL to 430 mg/dL. The three classes are classified as follows: glucose concentrations less than 80 mg/dL (hypoglycemic range) as class 0, greater than 180 mg/dL (hyperglycemic range) as class 2, and concentrations between these values (normal range) as Class 1. The input for the machine learning classification algorithms comprises the data points (x1, x2, x3 and x4), and the classification models are trained to predict the class labels of new cases accurately. The multi-class classification was performed using three classification algorithms: AdaBoosting (AB), Decision Tree (DT), and Gaussian Naive Bayes (GNB). To achieve the best performance of classification algorithms, fine-tuning the model is crucial. The hyperparameter tuning values are set using the gridsearch CV before the training process begins. Repeated looping through predefined hyperparameters helps to fit the model to the training set. Table 3 lists the best-tuning hyperparameter values for the three algorithms.

GNB classifier

The GNB classifier is a probabilistic ML model that is commonly used for classification tasks. The algorithm is based on Bayes' theorem and initially estimates the mean and variance of each feature for each class label based on the training data. The algorithm calculates the conditional probability of each class label given the observed features by utilizing the previously estimated mean and variance values and applying Bayes' theorem when presented with new data. In this study, a 10-fold approach was used to evaluate the algorithm's performance. The obtained

results are presented in Fig. 8. The model's accuracy was evaluated for each fold resulting in the following accuracies: (1, 1, 0.977, 1, 0.977, 1, 1, 1, 0.977, 1), with a mean accuracy of 0.993. The accuracy loss for each fold was (0, 0, 0.02325581, 0, 0.02325581, 0, 0, 0, 0.02325581, 0). The overall accuracy of the Gaussian Naive Bayes test was 96.53%. During testing, one of the data points was predicted to belong to class 2, which matched the original class.

performance on the training data. One of the advantages of AB is that it is less prone to overfitting than a single, more complex classifier. The accuracy, along with the accuracy loss for 10-fold with good cross-validation model performance, is presented in Figure 9. The accuracy of the trained model was evaluated for each fold, resulting in accuracies of (0.955, 1, 1, 1, 0.977, 0.953, 1, 0.977, 0.977, 1), with a mean accuracy of 0.984. The accuracy loss for each fold was (0.04545455, 0, 0, 0,

Table 3. Hypermeter best-tuning parameters values.

Classifier	Gridsearch CV hypermeter tuning values
GNB	var_smoothing=0.0004328761281083057
AB	Learning_rate:0.1, n_estimators:10
DT	max_leaf_nodes': list(range(2, 100)), 'min_samples_split': [2, 3, 4], Cv=10, verbose=1)
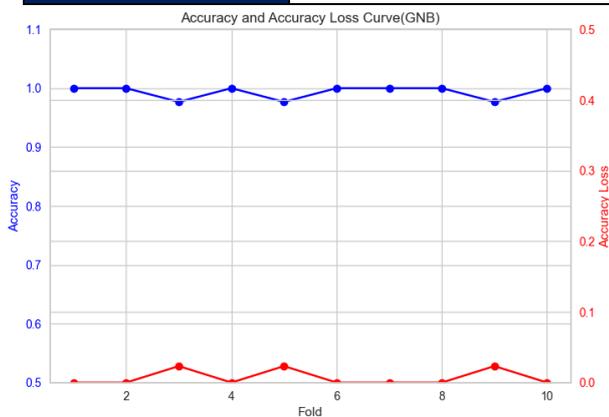
Figure 8. Accuracy, loss for each fold, and model GNB classifier
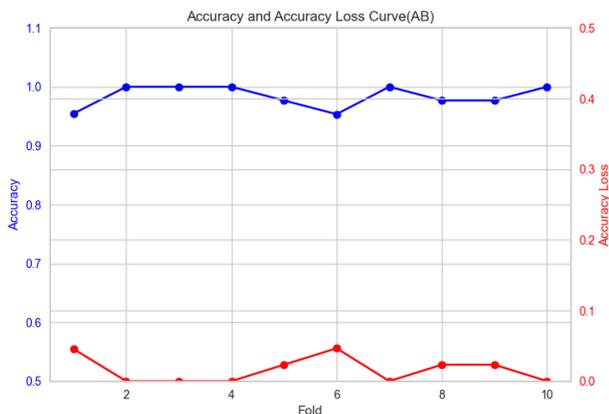
Figure 9. Accuracy and loss for each fold for AB classifier

AB classifier

The AB algorithm is a widely used ensemble learning approach for classification tasks. Its working principle involves iteratively training a series of weak classifiers. The one that performs the best on the weighted data is selected and added to the ensemble. The final classifier is a weighted combination of the weak classifiers, where the weight of each weak classifier is proportional to its

0.02325581, 0.04651163, 0, 0.02325581, 0.02325581, 0). The overall test accuracy for Ada Boosting was 97.22%. During testing, one of the data points was predicted to belong to class 0, which matched the original class.

DT classifier

Thirdly, the DT classifier algorithm is used for classification tasks. It creates a model that predicts the target variable's value by learning simple decision rules inferred from the data features. The model takes the form of a tree-like structure where each internal node represents a test on an attribute, each branch represents the outcome of the test, and each leaf node represents a class label. To make predictions on new data, the algorithm traverses the tree from the root node to a leaf node that corresponds to a class label, and the prediction is based on the majority class of the training that reaches that leaf node. The accuracy, loss, and model performance for 10-fold cross-validation are calculated and presented in Figure 10. The accuracy of the trained model was evaluated for each fold, resulting in accuracies of (0.977, 0.953, 0.93, 0.93, 0.837, 0.93, 1, 0.814, 0.977, 0.953), with a mean accuracy of 0.93. The accuracy loss for each fold was (0.13636364, 0.13953488, 0.02325581, 0.09302326, 0.09302326, 0.11627907, 0, 0.09302326, 0.02325581, 0.02325581). The overall test accuracy for DT was 95.14%. During testing, one of the data points was predicted to belong to class 2, which matched the original class.

To compare the performance of three machine learning classifier models for the multiclass problem with the three classes (0, 1, and 2) of predicting the glucose range in diabetic patients, the performance of the three models was evaluated by measuring precision, recall, and F1-score. The results of these evaluations are presented in Figures 11 to 13.
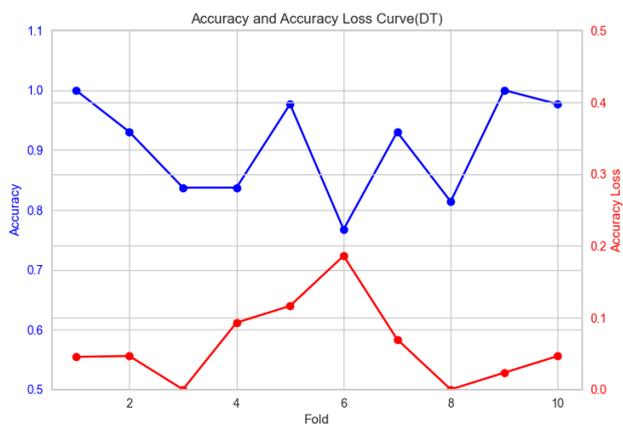
Figure 10. Accuracy and loss for each fold for the DT classifier

After evaluating the performance of the three models, it was found that the AB model achieved an overall accuracy of 97%, the GNB model achieved an overall accuracy of 96.53%, and the DT model achieved an overall accuracy of 95.14%. These accuracy values indicate that all three models could accurately predict the glucose range for diabetes patients across the three classes, with the AB model having the highest overall accuracy. These results demonstrate the potential of ML models in accurately predicting the glucose range for diabetes patients, which can aid in disease management and treatment.
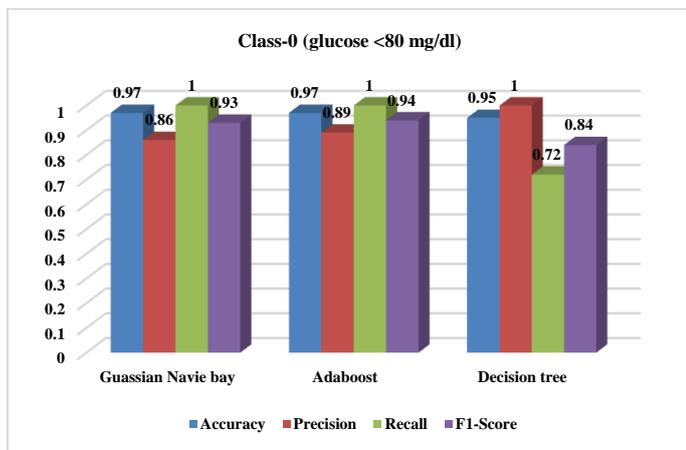
Figure 11. Comparison of three ML classifiers for Class 0
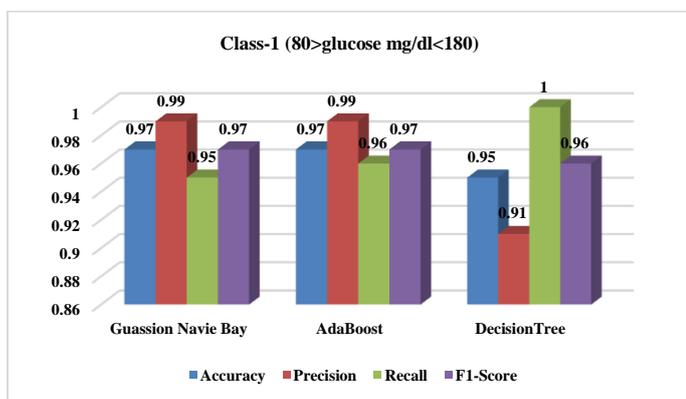
Figure 12. Comparison of three ML classifiers for Class 1
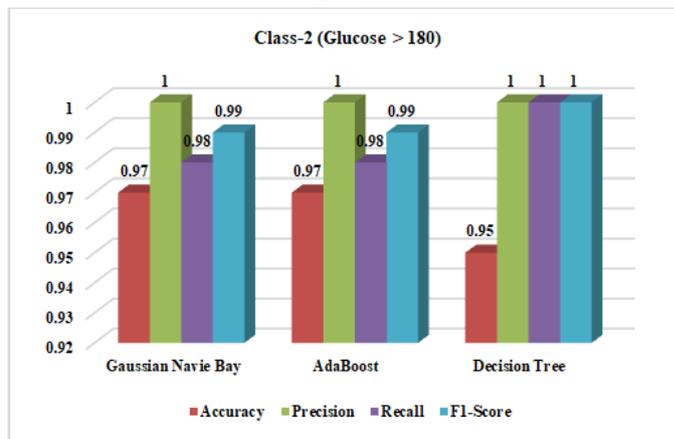
Figure 13. Comparison of three ML classifiers for Class 2

Discussion

The proposed approach for detecting blood glucose levels utilizes wavelength near-infrared (NIR) technology at 950 nm. This approach outperforms other technologies with high accuracy (R^2). This study measured the blood glucose history of 282 T2D and 7 T1D patients under medical supervision at the VIT-AP University health center. All participants provided informed consent under the Helsinki guidelines. Blood glucose levels were measured for 5 minutes using a proposed sensor glucose monitoring system and a reference device, the Dr. Trust fingerpick device. Data from 289 subjects, including males and females aged 19-69 years with hypo, normal, and hyperglycemia, were analyzed, resulting in a total of 575 blood glucose level samples obtained through a glucometer (ranging from 62 to 400 mg/dL). The study identified hypoglycemic (BG level <80 mg/dL), normal (79>BG level <182 mg/dL), and hyperglycemic (BG level >180 mg/dL) levels. The reference device, Dr. Trust, uses glucose dehydrogenase (GDH) flavin adenine dinucleotide (FAD) enzyme (FAD-GAD) with a measuring range of 30-600 mg/dL and requires 0.5 μl blood. This device was validated using HbA1C lab tests, with the results showing 99% accuracy compared to the reference device. The lab test HbA1C values were converted from mmol/L to mg/dL using the validation of the reference device (Dr. Trust).

Conclusions

This proposed system with a 950 nm wavelength was used in this study to determine subjects' glucose levels without invasive methods. Measured non-invasive glucose values are compared with invasive glucose measurements from a gold standard Dr. Trust glucose meter. A total of 575 real-time samples are collected from 289 subjects' random glucose measurements. Regression expression is utilized in the suggested strategy

to increase accuracy based on real-time data analysis. In real-time data analysis with the proposed method for the sensor, the R^2 and MAD increase to 0.99 and 3.6 mg/dl, respectively. Additionally, it is obtained the RMSE is 3.46 mg/dl. The three ML classification methods were used to predict multiclass, the 2-classifiers given 97% and the 1- classifier given 95%. Based on these parameters, the proposed method appears to be more efficient than the existing literature. In the present work, the limitation is in the form of a system that can be further enhanced as a portable device. From the statistical point of view, more subjects should be tested on Type-I diabetes to analyze time series responses in future work.

Acknowledgment

Institutional Review Board Statement

The data collection and measurement method was approved by Institutional Ethical Committee (IEC) VIT-AP-IEC/20220803 and collected samples as per guidelines with the World Medical Association (WMA) & Helsinki.

Informed Consent Statement

Informed consent was obtained from all subjects involved in the study.

Data Availability Statement

We have created our own dataset with help our proposed sensors for this study. The research is still in progress so we cannot publish the dataset right now.

Conflicts of Interest

The authors declare no conflict of interest.

References

Abidin, M. T., Rosli, M. K., Shamsuddin, S. A., Madzhi, N. K., & Abdullah, M. F. (2013). Initial quantitative comparison of 940nm and 950nm infrared sensor performance for measuring glucose non-invasively. *In 2013 IEEE, International Conference on Smart Instrumentation, Measurement and Applications (ICSIMA)*, pp. 1-6. https://doi.org/10.1109/ICSIMA.2013.6717938

Ali, A., Alrubei, M. A. T., Hassan, . L. F. M., Al-Ja'afari, M. A. M., & Abdulwahed, S. H. (2020). Diabetes diagnosis based on knn. *IIUM Engineering Journal*, *21*(1), 175–181. https://doi.org/10.31436/iiumej.v21i1.1206

Anupongongarch, P., Kaewgun, T., O'Reilly, J. A., & Khaomek, P. (2019). Development of a non-invasive blood glucose sensor. *International Journal of Applied Engineering Research*, *12*(1), 1-19.

Chandrasekhar, N., & Peddakrishna, S. (2023). Enhancing heart disease prediction accuracy through machine learning techniques and optimization. *Processes, 11*(4), 1210. https://doi.org/10.3390/pr11041210.

Chen, C., Zhao, X.L., Li, Z.H., Zhu, Z.G., Qian, S.H., & Flewitt, A. J. (2017). Current and emerging technology for continuous glucose monitoring. *Sensors, 17*(1), 182. https://doi.org/10.3390/s17010182.

Goodarzi, M., Sharma, S., Ramon, H., & Saeys, W. (2015). Multivariate calibration of NIR spectroscopic sensors for continuous glucose monitoring. *TrAC Trends in Analytical Chemistry*, *67*, 147-158. https://doi.org/10.1016/j.trac.2014.12.005.

Gusev, M., Poposka, L., Guseva, E., Kostoska, M., Koteska, B., Simjanoska, M., Ackovska, N., & Stojmenski, A. (2020). Trends from minimally invasive to non-invasive glucose measurements. IEEE, *In 2020 43rd International Convention on Information, Communication and Electronic Technology (MIPRO).* pp. 315-320. https://doi.org/10.23919/MIPRO48935.2020.9245402.

Haak, T., Hanaire, H., Ajjan, R., Hermanns, N., Riveline, J.P., & Rayman, G. (2017). Use of flash glucose-sensing technology for 12 months as a replacement for blood glucose monitoring in insulin-treated type 2 diabetes. *Diabetes Therapy*, *8*, 573–586. https://doi.org/10.1007/s13300-017-0255-6

Heise, H.M. (2021). Medical Applications of NIR Spectroscopy. In: Ozaki, Y., Huck, C., Tsuchikawa, S., Engelsen, S.B. (eds) Near-Infrared Spectroscopy. Springer, Singapore. pp. 437-473. https://doi.org/10.1007/978-981-15-8648-4_20

Hina, A., & Saadeh, W. (2022). Noninvasive blood glucose monitoring systems using near-infrared technology—A review. *Sensors*, *22*(13), 4855. https://doi.org/10.3390/s22134855.

Jain, P., Joshi, A.M., & Mohanty, S.P. (2019). iGLU: An intelligent device for accurate noninvasive blood glucose-level monitoring in smart healthcare. *IEEE*

Consumer Electronics Magazine, *9*(1), 35-42. https://doi.org/10.1109/MCE.2019.2940855.

Jain, P., Maddila, R., & Joshi, A.M. (2019). A precise non-invasive blood glucose measurement system using NIR spectroscopy and Huber's regression model. *Optical and Quantum Electronics*, *51*, 1-5. https://doi.org/10.1007/s11082-019-1766-3.

Jintao, X., Liming, Y., Yufei, L., Chunyan, L., & Han, C. (2017). Noninvasive and fast measurement of blood glucose in vivo by near infrared (NIR) spectroscopy. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, *179*, 250-254. https://doi.org/10.1016/j.saa.2017.02.032.

Larin, K. V., Eledrisi, M. S., Motamedi, M., & Esenaliev, R. O. (2002). Noninvasive blood glucose monitoring with optical coherence tomography: a pilot study in human subjects. *Diabetes Care*, *25*, 2263–2267.
https://doi.org/10.2337/diacare.25.12.2263

Lee, S.H., Kim, M.S., Kim, O.K., Baik, H.H., & Kim, J.H. (2019). Near-infrared light emitting diode based non-invasive glucose detection system. *Journal of Nanoscience and Nanotechnology*, *19*(10), 6187-6191.
https://doi.org/10.1166/jnn.2019.17005.

Miriyala, N.P., Kottapalli, R. L., Miriyala, G. P., Lorenzini, G., Ganteda, C., & Bhogapurapu, V. A.R. (2022). Diagnostic analysis of diabetes mellitus using machine learning approach. *Revue d'Intelligence Artif.*, *36*, 347–352. https://doi.org/10.18280/ria.360301

Montgomery, D.C., Peck, E.A., & Vining, G.G. (2021). Introduction to linear regression analysis (6th ed.). John Wiley & Sons.

Pai, P.P., Sanki, P. K., Sahoo, S. K., De, A., Bhattacharya, S., & Banerjee, S. (2017). Cloud computing-based non-invasive glucose monitoring for diabetic care. *IEEE Transactions on Circuits and Systems I: Regular Papers*, *65*, 663–676. https://doi.org/10.1109/TCSI.2017.2724012

Pai, P.P., De, A., & Banerjee, S. (2017). Accuracy enhancement for noninvasive glucose estimation using dual-wavelength photoacoustic measurements and kernel-based calibration. *IEEE, Transactions on Instrumentation and Measurement*, *67*, 126–136.
https://doi.org/10.1109/TIM.2017.2761237

Pigman, W. (2012). The Carbohydrates: Chemistry and Biochemistry Physiology. Elsevier, New York City.

Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine learning in medicine. New England *Journal of Medicine*, *380*, 1347–1358.
https://doi.org/10.1056/NEJMra1814259

Simeone, M. L., Parrella, R. A., Schaffert, R. E., Damasceno, C. M., Leal, M. C., & Pasquini, C. (2017). Near infrared spectroscopy determination of sucrose, glucose and fructose in sweet sorghum juice. *Microchemical Journal*, *134*, 125-130. https://doi.org/10.1016/j.microc.2017.05.020.

Smith, J.L. (2015). The pursuit of noninvasive glucose: Hunting the deceitful turkey (Revised and Expanded).

Song, K., Ha, U., Park, S., Bae, J., & Yoo, H.J. (2015). An impedance and multi-wavelength near-infrared spectroscopy ic for non-invasive blood glucose estimation. *IEEE, Journal of Solid-State Circuits*, *50*, 1025–1037.
https://doi.org/10.1109/JSSC.2014.2384037

Sun, H., Saeedi, P., Karuranga, S., Pinkepank, M., Ogurtsova, K., Duncan, B.B., Stein, C., Basit, A., Chan, J.C., & Mbanya, J.C. (2022). Idf diabetes atlas: Global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045. *Diabetes Research and Clinical Practice*, *183*, 109119.
https://doi.org/10.1016/j.diabres.2021.109119.

Teki, S., Sriharsha, K., & Nandimandalam, M. (2021). A diabetic prediction system based on mean shift clustering. *Nutrition*, *84*, S177–S181. https://doi.org/10.18280/isi.260210

Uwadaira, Y., Ikehata, A., Momose, A., & Miura, M. (2016). Identification of informative bands in the short-wavelength NIR region for non-invasive blood glucose measurement. *Biomedical Optics Express*, *7*(7), 2729.
https://doi.org/10.1364/BOE.7.002729.

Van Enter, B. J., & Von Hauff, E. (2018). Challenges and perspectives in continuous glucose monitoring. *Chemical Communications*, *54*(40), 5032-5045. https://doi.org/10.1039/C8CC01678J.

Villena Gonzales, W., Mobashsher, A. T., & Abbosh, A. (2019). The progress of glucose monitoring-A review of invasive to minimally and non-invasive techniques, devices and sensors. *Sensors*, *19*(4), 800. https://doi.org/10.3390/s19040800.

Yadav, J., Rani, A., Singh, V., & Murari, B. M. (2015). Prospects and limitations of non-invasive blood glucose monitoring using near-infrared spectroscopy. *Biomedical Signal Processing and Control*, *18*, 214-227.

https://doi.org/10.1016/j.bspc.2015.01.005.

Yang, W., Liao, N., Cheng, H., Li, Y., Bai, X., & Deng, C. (2018). Determination of NIR informative wavebands for transmission non-invasive blood glucose measurement using a Fourier transform spectrometer. *AIP Advances*, *8*(3), 035216. https://doi.org/10.1063/1.5017169.

Yeaw, J., Lee, W. C., Aagren, M., & Christensen, T. (2012). Cost of self-monitoring of blood glucose in the United States among patients on an insulin regimen for diabetes. *Journal of Managed Care Pharmacy*, *18*(1), 21-32. https://doi.org/10.18553/jmcp.2012.18.1.21