Indian Journal of Pure & Applied Physics
Vol. 63, September 2025, pp. 803-815
DOI: 10.56042/ijpap.v63i9.17462

A Portable AI-Based Non-Invasive Blood Glucose Monitoring System Using Multi-Wavelength NIR Spectroscopy

Alok Kumar* & Mrutyunjay Rout

Electronics and Communication Engineering, National Institute of Technology Jamshedpur, Jharkhand 831 014, India

Received: 26th Feb 2025; accepted: 17th July 2025

Diabetes mellitus (DM) is a metabolic disorder characterized by increased blood glucose levels, which can cause significant health problems and premature death. Patients with diabetes must regularly monitor their blood glucose (BG) levels in order to estimate the insulin intake. Invasive methods are commonly used for blood glucose measurement because they are highly accurate but require skin puncturing, which is uncomfortable and increases the risk of infectious disease transmission. Alternatively, non-invasive techniques need no skin damage, making them simple, painless, and practical for routine monitoring. We present a non-invasive glucose monitoring system that analyses reflectance and transmittance of near-infrared (NIR) signals at three wavelengths (940 nm, 1050 nm, and 1300 nm). LEDs and photo-detector forms an optoelectronics circuit for NIR signals, while an ATmega32 microcontroller processes digital signals and displays the results. A cluster-based piecewise regression model is proposed to develop an optimized model from NIR discrete voltage values and invasive glucometer readings. A five-cluster piecewise linear multivariate regression (PLMR) model is employed on the training data set, which shows an excellent correlation of 92.24% and a low RMSE value of 17.28 mg/dl, indicating that it can be used in practice. The developed device is low-cost, portable and offers a painless alternative for regular monitoring.

Keywords: Blood glucose measurement, Non-invasive, NIR spectroscopy, Clustering, Regression models

1 Introduction

Diabetes is becoming one of the most common and severe diseases that affect people all over the world. It presently influences 415 million peopleand is expected to rise up to 783 million by 2045. The exponential rise in the number of diabetes patients has become a global threat, highlighting the need for a glucose monitoring system. Diabetes mellitus is a disorder characterized by elevated blood sugar levels. which impairs the ability to use the energy found in the food. In general, all types of carbohydrates are turned into glucose, a basic sugar that serves as fuel for body cells. Insulin is a hormone found in the bloodstream that allows blood glucose to be stored as energy. Blood glucose levels rise due to a lack of insulin, insulin resistance, or excessive glucagon production in the body. Insulin imbalance is caused by two factors: the immune system of the body destroys produced insulin, resulting in Type 1 diabetes.Type-2 diabetes, which accounts for nearly 90% of all diabetic cases[1, 2], begins when the body is unable to generate enough insulin to maintain blood sugar levels. An elevated blood sugar level raises the risk of death by about 50% compared to a healthy person since it can lead to major health consequences like cardiovascular disease, depression, cognitive dysfunction, visual impairment, and kidney problems[3].

A diabetic person must regularly check their blood glucose levels and adjust their food and medication schedules as necessary. Blood glucose monitoring methods can be grouped into three main categories: invasive, semi-invasive, and non-invasive, in Fig. 1.

Invasive procedures are traditional blood glucose testing methods that require a patient's blood sample. It delivers accurate readings and is widely used; however, repeated puncturing of the skin produces agony and suffering, as well as the risk of infections[4]. The semi-invasive method measures glucose levels from interstitial or transdermal fluid, which differs from capillary blood glucose levels. Non-invasive methods, on the other hand, are painless, practical, and easy for routine monitoring since they rely on the analysis of physiological signals whose characteristics are correlated with blood glucose levels.

The noninvasive methods of measuring blood glucose have drawn a lot of attention from researchers in the past few decades. A number of non-invasive methods have been proposed for developing an

*Corresponding author: E-mail: researchalok.51289@gmail.com

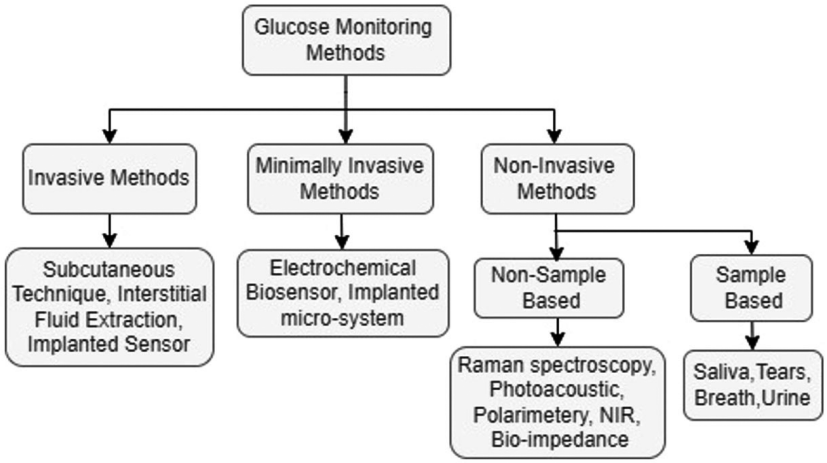

Fig. 1 — Techniques for the glucose monitoring system

accurate blood glucose monitoring system. Raman spectroscopy detects molecules by measuring the inelastic shift in the frequency of monochromatic light (LAESR)[5]. The implementation of the setup is non-portable and requires a lot of space. The photo-acoustic method is based on the detection of the ultrasonic waves that the molecule emits following its absorption of photon energy[6, 7]. The ultrasonic wave is produced when the heated molecules start to vibrate, causing the medium to expand in volume. In the polarization technique, molecules affect the polarity of the light wave[8]. Due to glucose's dextrorotatory characteristics, with the increase in concentration, the linearly polarized wave rotates in a right-handed direction. Dachao *et al*.[9], reported a non-invasive sensor that measures blood sugar levels by using ultrasonic radiation at the skin's surface; however, the sensor has not been downsized in order to be deemed a wearable sensor. Pleitez *et al*.[10] has reported a non-invasive sensor device based on resonance-enhanced pulsed photoacoustic spectroscopy using a windowless resonator cell placed on the fingertip. This sensor system design is complex, and more enhancements to the cell, as well as integration into a portable device, are required. Saliva has also been used as a non-invasive technique for glucose measurement[11, 12]. Impedance spectroscopy through the skin is used for the estimation of glucose level presented in[13,14]. As the electrical properties of the saliva, skin, and sweat vary according to the person, this approach will not be reliable. Glucose measurement utilizes the retina[15], which has good accuracy, although this technique is not appropriate for regular monitoring. The continuous glucose monitoring (CGM) or implanted systems are invasive methods that require replacement due to their short battery life.

IR spectroscopy demonstrates the interaction between optical radiation and matter, and has been gradually applied to different parts of the body for noninvasive blood glucose measurement. In the NIR signal region, blood water molecules becomes reasonably transparent also the penetration depth inside the body tissue of the NIR signals are higher as compared to the mid and far IR signals[16-19]. Additionally, compared to non-invasive laser methods, NIR-based technologies are smaller, less costly, and less harmful to the body. Despite of many advantages, non-invasive devices are not used for diagnosis purpose as they lack in accuracy compared to invasive techniques.

We examine the variation in the intensity of transmitted as well as reflected light waves due to blood glucose at three NIR wavelengths: 940 nm, 1050 nm and 1300 nm, respectively. Photo detectors having high peak sensitivity at specific wavelengths are used to collect transmitted and reflected NIR signals, which are then logged to discrete voltage values. The voltage values along with the glucometer reading are used to develop proposed cluster based piecewise linear regression model. In PLMR, we grouped the data into minimum number of clusters and form regression model between two successive clusters to approximate the nonlinear multivariate regression model as a piecewise linear multivariate regression model. The data samples are gathered from both diabetic as well as healthy individuals while adhering to proper medical safety standards. The

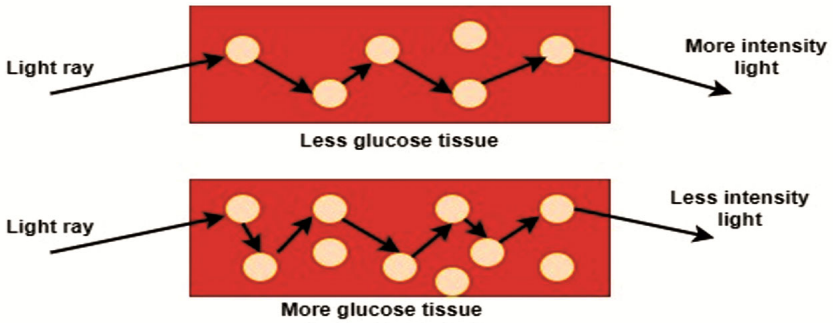

Fig. 2 — A graphical representation of the light attenuation caused by glucose molecules

device's medical accuracy is evaluated using a Clarke Error Grid (CEG) plot. The objective of this research is to develop a highly accurate, low-cost, reliable, non-invasive blood glucose monitoring system.

The remaining section of the paper is organized as follows: Section 2 describes the theory of NIR spectroscopy. Section 3, discusses the details of the prototype design and implementation, while in Section 4, the summary of regression models are discussed. Results and validation of the device are discussed in section 5, and the paper is concluded in section 6.

2 NIR Spectroscopy

2.1 Theory

NIR spectroscopy is based on the vibrational properties of the bonding atom'sat its resonance frequency. LEDs with specific NIR wavelengths are used to generate the required resonance frequency. As the light waves travels through the medium, it gets attenuated, due to scattering and absorption[20,21]. Beer-Lambert Law states that absorbance of light through any solution is proportional to the concentration of the solution and the length of the path travelled by light as illustrated in Fig. 2.

2.1.1 Reflectance Spectroscopy

According to the light transport theory, attenuation of the reflected light can be described Eq. (1).

$$I = I_0 e^{\mu_{eff} L} \quad \text{... (1)}$$

where, I represent the reflected light intensity, I_0represents the incident light intensity, and L is the optical path length of the tissue. The effective attenuation coefficient (μ_{eff}) is the light attenuation within the tissue and is given by Eq. (2).

$$\mu_{eff} = \mu_a \sqrt{3\,\mu(+\mu_s')} \quad \text{… (2)}$$

The absorption coefficient (μ_a) is the photon absorption within tissue per unit path length, and it is given by Eq. (3).

$$\mu_a = 2.303 \in \mathrm{C} \quad \text{… (3)}$$

where, $\in$ is the molar extinction coefficient, C is the tissue chromophore concentration and $\mu_s'$is the reduced scattering coefficient given by Eq. (4).

$$\mu_s' = \mu_s\,(1-g) \quad \text{… (4)}$$

where, g is anisotropy and μ_s is scattering coefficient. The scattering and absorption of the light ray inside the human body is caused due to the mismatch between the refraction index of extracellular fluid (ECF) and the cell membrane. The refractive index of ECF (η_{ecf}) is approximately 1.348–1.352, and that of a blood cell (η_{cell}) is approximately 1.350–1.460[20]. As the concentration of glucose rises, the value of η_{ecf} rises, reducing the difference between the refractive indexes of ECF and cellular membrane. As a result, scattering coefficients and optical paths are reduced. Thus, it follows from Eq. (1-4) that as the concentration of glucose rises, the intensity of reflected light decreases.

2.1.2 Transmittance Spectroscopy

By Beer's law definition, the attenuation in the intensity of the transmitted light wave is given by Eq. (5).

$$\mathrm{T} = \frac{\phi_e^t}{\phi_e^i} = e^{-\tau} = 10^{-A} \quad \text{… (5)}$$

where, T represent the transmitted light intensity ϕ_e^t is the radiant flux transmitted by the surface, ϕ_e^i is the

Table 1 — Fundamental vibration of bonds

Glucose bonds	Vibrational frequency overtones (nm)			
	Fundamental	1st	2nd	3rd
OH stretching	2860-3120	1410-1440	970	738
OH combinations	1920-2080	1100	840	--
CH stretching	3300-3470	1600-1800	1100-1230	910
CH combinations	2100-2352	--	--	--
CH_2 stretching	3460-3500	1720-1765	1215	930
CH_2 combinations	2310-2325	--	--	--

Note: The '–'mark represents no data available

radiant flux received by the surface. The transmittance intensity can also be represented as Eq. (6).

$$T = e^{-\sigma\int_0^t N(z)dz} = 10^{-\varepsilon\int_0^t c(z)dz} \quad \ldots (6)$$

Or equivalently that

$$\tau = \sigma \int_0^t N(z)dz \quad \ldots (7)$$

$$A = \varepsilon \int_0^t c(z)dz \quad \ldots (8)$$

where, l is the path length; N is the number density of the attenuating species in the material; σ is the attenuation cross section; ε is the molar attenuation coefficient; c is the molar concentration of the attenuating species in the material. In case of uniform attenuation these relation becomes as Eq. (9-10).

$$\tau = \sigma N l \quad \ldots (9)$$

$$A = \varepsilon c l \quad \ldots (10)$$

Hence, it is concluded that the intensity of reflected as well as transmitted light, inversely proportional to the molecular concentration. Thus, at resonant frequency of glucose molecules the intensity of reflected and transmitted light wave can be used to estimate its concentration.

2.2 NIR spectroscopy of Glucose Molecule

The glucose molecule ($C_6 H_{12} O_6$) contains single as well as double bond (*C-H, C-OH*, and *C=O*) between its atoms. At specified wavelengths, the bonding atoms starts vibrating through two processes: overtone (stretching) and combination. The fundamental stretching of the glucose molecule is primarily found in the long-NIR region (1300nm-2500nm), whereas its overtone occurs in the short-NIR region (700nm-1300nm) [22-24] in Table 1 .The combination process of vibration occurs due to the sharing of two or more fundamental frequencies in the mid and far-NIR regions.

The absorption of glucose molecules in the short NIR range is very low and highly impacted by water molecules in the bloodstream. By selecting a proper near-infrared wavelength and developing an optimal regression model, the interference caused by water molecules can be minimized.

3 Prototype Design and Implementation

3.1 Wavelength Selection

The spectrum of infrared waves ranges from 600nm to 25,000nm and is divided into three regions: near-IR (700nm–2500nm), mid-IR (2500nm–10,000nm), and far-IR (10,000nm–25,000nm). The near-IR is further divided into the short NIR (600nm-1300nm) and long-NIR (1300nm-2500nm).The absorption of light wave is higher in the long-NIR range but has a lower penetration depth, which further decreases as the wavelength increases. Whereas in the short-NIR region, the absorption is low but it has a good penetration depth of around 5mm inside the human skin[19, 25]. The informative band of NIR signal for glucose detection is observed at 836 nm, 950 nm, 1040 nm, 1150 nm, 1280 nm, and 1395 nm[21-23], with a correlation coefficient larger than 0.7 as indicated in Fig. 3.

In our device, we have used three NIR wavelengths and collected data at four channels: transmittance at 940 nm and reflectance data at 940 nm, 1050 nm and the 1300 nm wavelength respectively. At 940nm, it has high correlation coefficient in spite of having minimal absorption. The NIR wavelength at 1050nm and 1300nm has significant absorption and better correlation coefficient (greater than 0.7) as shown in

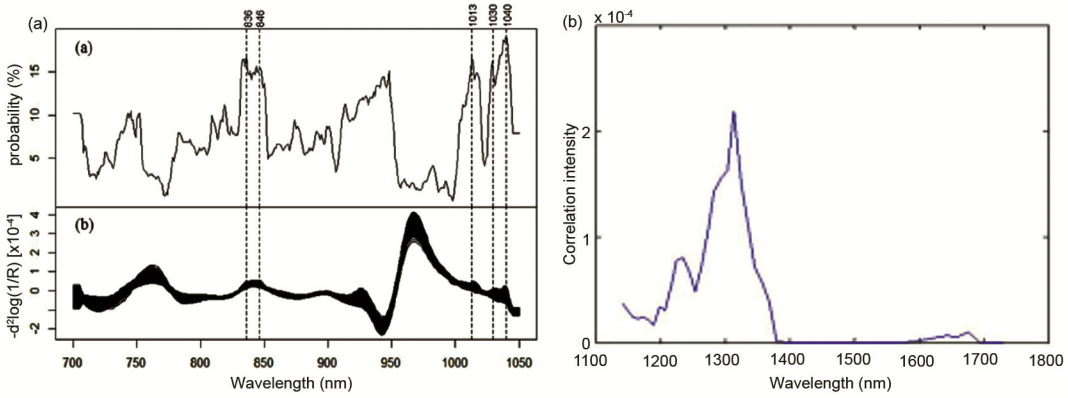

Fig. 3 — Glucose-linked wavelengths in the short NIR region a) Probability of the correlation coefficient [21] b) Correlation spectra of aqueous solutions with a glucose concentration of 100 mg/dl[22]

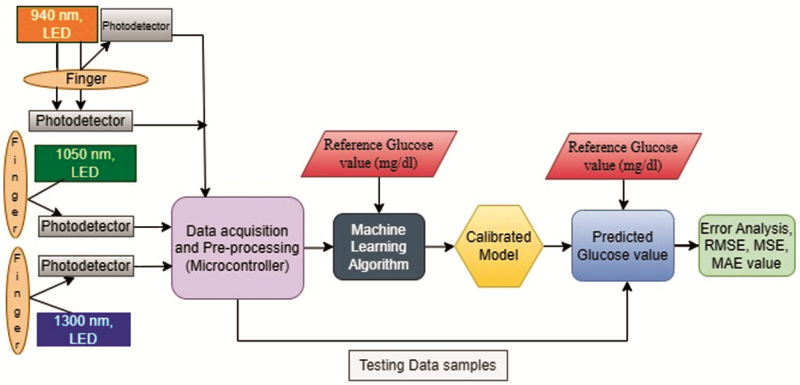

Fig. 4 — Prototype of the proposed glucose measurement device

Fig. 3. But it has considerable attenuation in the transmitted wave, hence the reflected wave is used at these wavelengths. The combination of transmittance and reflectance spectroscopy is used for estimating the concentration of the glucose molecules.

3.2 System Design

The proposed device is a dual NIR spectroscopy-based opto-electric circuit that analyses the variation in the intensity of transmittance and reflectance spectroscopy to estimate the blood glucose level. The device uses LEDs as an NIR source, a microcontroller for processing the data, and an LCD to display the results. All the components are embedded on the PCB board and operated using a 12V DC power supply, the architecture of the device in Fig. 4.

The device uses two probes for which comprise four channels for collecting data. One probe is embedded with 940 nm LED and two photo detectors (sensitivity range 600nm-1750nm), one on the opposite side and the other on the same side of the LED, to collect transmittance as well as reflectance spectroscopy. In another probe, a photo-detector is placed between 1050 nm and 1300 nm at a distance of 4 mm from each to obtain reflectance spectroscopy. The probes are packed in the light-blocking enclosure box in order to minimise the probability of a faulty measurement in Fig. 5.

During data collection, each channel is activated for 10 seconds, while the others are turned off. The two fingertips are inserted into the two probes and held steady for 40 seconds; the timing is shown on the LCD. The photodetector's output current is proportional to the intensity of the received light, which is converted into voltage by a trans-impedance amplifier[25, 26]. The microcontroller (Atmega32) samples the voltage signal at a rate of 128 samples per second, and each sample is represented by 16 bits. The block diagram in Fig. 6, shows the sequence of analog voltage to digital conversion. Thus, for 10 seconds, a total of 1280 voltage samples are collected from each of the four channels and logged to get the decimal voltage value. The external memory module of 32k-bit is used to reduce the quantization error.

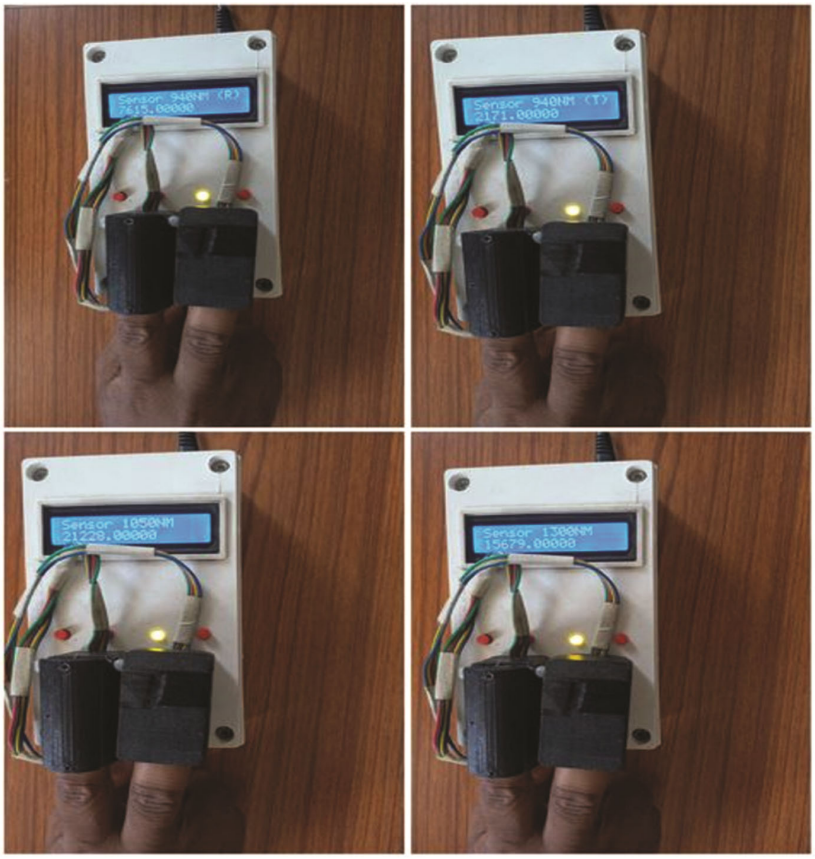

Fig. 5 — Prototype of the proposed device

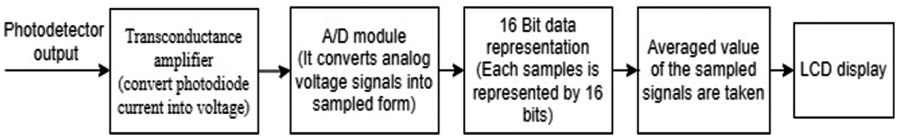

Fig. 6 — Data processing block diagram.

An invasive glucometer (Control D) reading is taken simultaneously and considered as the target value for the corresponding NIR signal reading. The labeled data set included voltage values from four channels as well as non-invasive glucometer readings, which were used to develop an optimized regression model. A cluster-based piecewise regression model is proposed to approximate the non-linear model with piecewise linear models. The model is validated on Clarke Error grid analysis for clinical accuracy. The description of the components used in the device in Table 2.

4 Regression Models

Regression analysis is statistical modeling that describes the relationship between input and target value in a labeled data set in terms of a mathematical equation[27, 28]. Regression models can be linear or non-linear, depending on the correlation between input and target values. There are four major steps to regression modeling.

i. The labelled data set is analyzed using regression.
ii. The scatter plots and correlation matrices are used for evaluating the relationship among the variables.
iii. To obtain an optimized model, a suitable regression algorithm is used according to the characteristic of the data set.
iv. The error analysis measures, accuracy of fitness of the model on the dataset.

The proposed PLMR model as well as the existing regression models are discussed in the following subsections.

Table 2 — Hardware description of the components

Components	Model name	Specifications
NIR LED	MTE5010-525-IR	Peak Wavelength=1050nm Forward voltage=1.5v Radiant Power=4mW
	MTE1300N5	Peak Wavelength=1300nm Forward voltage=1.05v Radiant Power=120 mW
Photo-detector	MTPD1346D-100	Sensitivity range=600nm-1750nm Dark current==2μA Breakdown voltage=3v
Microcontroller	ATmega32	40 pin DIP, 8 channels , 10-bit resolution ADC RAM= 2Kbytes, 32k byte programmable memory
Invasive Glucometer	Control D	Minimum Blood Sample: 0.5 mg/dL Measuring Time: 5 sec

4.1 Linear Regression Model

Linear regression (LR) algorithm fits better on the data set having linear relationship between its input and target value[28, 29]. Eqs. (11-12) provides general expression for the univariate and multivariate linear regression models:

$\hat{y} = \beta_0 + \beta_1 x$, univariate linear regression … (11)

$\hat{y} = \beta_0 + \sum_{i=1}^{N} \beta_i x_i$, multivariate linear regression … (12)

where, x_i, represents the i^{th} feature of N feature vectors, β_i represents the coefficients of the model and the predicted value is given by ŷ. However, in general, datasets of real-world applications are not linear. In those cases, the linear regression algorithm fails to develop an optimized model.

4.2 Support Vector Machine

Support Vector Machine algorithms are used for both regression as well as classification problems. In regression problem, support vector regression (SVR) uses a set of functions called Kernel transform, which maps the input vector X to an N-dimensional space called the feature space[30-32]. SVR locates the hyper-plane in such a way to minimize the difference between predicted and observed values, in Fig. 7. The error function in Eq. (13) is used to minimize the $"w"$ value by customizing the hyper-plane.

$min\|w\|^2 + C\sum_i^n(\varepsilon_i^+ + \varepsilon_i^-)$ … (13)

where, n is the number of input samples, C is a user-defined parameter that represents the degree of optimization between empirical error $\sum_i^n(\varepsilon_i)$ and squared norm of weights. The SVR is applied on the mapped input features to minimize the error by using Eq. (14).

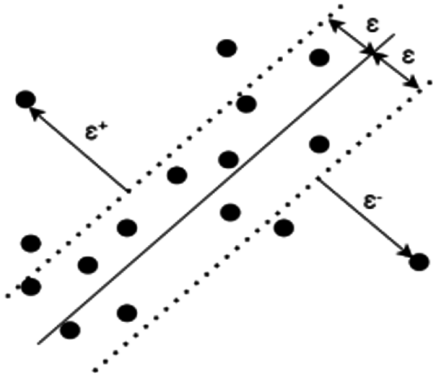

Fig. 7 — Support vector regression model

$$f(x) = \sum_i^n (\alpha_i^* + \alpha_i)K(x, x_i) + B, \forall x_i \in X \quad \ldots (14)$$

where, $(\alpha_i^*, \alpha_i) >= 0$ are Lagrange multiplier, $K(x, x_i)$ is the kernel function and B represent the bias term.

4.3 Regression Tree

A regression tree (RT) is a sort of decision tree that predicts continuous values using regression instead of classification methods. The greedy algorithm is used to implement the regression tree, which iteratively divides the training data into smaller subsets[33]. Implementing a regression tree involves the following steps:

i. A splitting rule that specifies the splitting variable and the breakpoint.

ii. Stopping rule that determines when a node terminal is declared.

iii. Prediction rule for the responses at the terminal node.

An input vector is traversed entirely through the tree by responding to True/False questions until it reaches the leaf node. The average of the dependent variable's value at that particular leaf node is used to determine the final prediction. Even though RTs are very efficient and easy to use, they are also notorious for being unstable[34]. Furthermore due to discontinuities in regression trees, a small alteration to the training set may result in an alternative outcome.

4.4 Proposed Piecewise Linear Multivariate Regression (PLMR) Model

The non linear multivariate regression models are approximated by piecewise linear multivariate regression model. Figure 8, depicts the PLMR algorithm, where we have arbitrarily assumed two-dimensional non-linear data on the normalized x-y plane. The data is divided into three clusters (K = 3)[35, 36], and organized sequentially based on the centroid's distance from the origin. The multivariate regression model is created between two subsequent clusters, and for each cluster model is defined, i.e., Model-1 for cluster one, the intersection of Model-1 and Model-2 for cluster 2, and Model-2 for cluster 3.

The piecewise approximation represent a linear multivariate regression model for the data of two subsequent clusters. At the intersection of the hyper-plane, it is difficult to define a boundary line. We approximate the continuity in the piecewise model by considering the averaged value of the intersecting models.

To estimate an output, the defined regression model should incorporate all of the attributes of the input vector. However, it is possible that some of the attributes of input belongs to a different regression model; in that case, selecting the regression model becomes difficult. We employed cluster-based assignment, where the input is first assigned to its nearest cluster, and then the regression model representing that cluster is used to find the output. The steps of the proposed PLMR algorithm are discussed below.

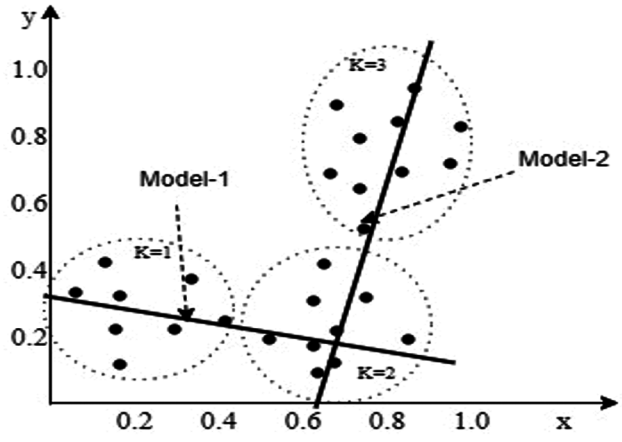

Fig. 8 — Illustration of PLMR using a 3-clusters.

Algorithm: Piecewise linear multivariate regression (PLMR) Model

Input: input data, x_i of N feature vectors, $\forall i \in n$

Output: Predicted value, $\hat{y}$

Step1. The k means clustering algorithm is used to group the data sets into k- clusters (here k=3) by using Eqs. (15-16).

$$\text{Euclidean distance} = \left\|x_i - c_j\right\|^2 \quad \ldots (15)$$

$$\text{Cost function } (J): J = \sum_{j=1}^{k}\sum_{i=1}^{n}\left\|x_I^{(J)} - c_J\right\|^2 \quad \ldots (16)$$

Where, c_J is the j^{th} cluster centroid and n is number of inputs.

Step2. The clusters are serially arranged in increasing order of centroid distance from the origin.

$$\text{Cluster centroid distance:} \quad d_j = \left\|c_j\right\|^2 \quad \ldots (17)$$

If, $d_1 < d_2 < d_3$, then, the sequence of clusters: $C_1 \rightarrow C_2 \rightarrow C_3$.

Step3. The multivariate linear regression model is developed for two successive clusters.

$$\text{Between cluster 1 and 2:} \quad \text{Model } 1 = \alpha_0 + \sum_{i=1}^{N}\alpha_i x_i \quad \ldots (18)$$

$$\text{Between cluster 2 and 3:} \quad \text{Model } 2 = \beta_0 + \sum_{i=1}^{N}\beta_i x_i \quad \ldots (19)$$

where, α and β represents the coefficients of the Model-1 and Model-2 respectively.

Step4. During measurement, the test data sample is assigned to its nearest cluster, and regression model associated to that cluster is used to find the output.

$$\hat{y} = \begin{cases} \text{Model } 1, & \text{If input data lies in the } 1^{st} \text{ cluster} \\ (\text{Model } 1 + \text{Model } 2)/2, & \text{If input data lies in the } 2^{nd} \text{ cluster} \\ \text{Model } 2, & \text{If input data lies in the } 3^{rd} \text{ cluster} \end{cases} \quad \ldots (20)$$

4.5 Evaluation Parameters of the Regression Models

The R-squared (R^2) value is used to assess how accurately the model fits the data set. The higher the R^2 value, the more closely the model fits the data set, Eq. (21).

$$R^2 = 1 - \frac{SSR}{SST} = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y}_i)^2} \quad \ldots (21)$$

where y_i is the i^{th} instance of actual value Y, $\hat{y}$ is the predicted value, and $\bar{y}_i$ is the mean of the actual value. The sum of square errors (SSR) with the best fit line, and the sum of errors (SST) with respect to the average fit of the line. The model's performance is examined in terms of mean absolute error (MAE), mean square error (MSE) and root mean square error (RMSE) values[27,29], calculated using Eq. (22-24).The MAE is the averaged value of the difference between the predicted and target values, whereas RMSE is the standard deviation of the error around the regression plane.

$$MAE = \frac{1}{n}\sum\left|y_i - \hat{y}\right| \quad \ldots (22)$$

$$MSE = \frac{1}{N}\sum_{i=1}^{N}\left|y_i - \hat{y}_i\right|^2 \quad \ldots (23)$$

$$RMSE = \sqrt{\frac{1}{n}\left|y_i - \hat{y}\right|^2} \quad \ldots (24)$$

5 Results and Discussion

5.1 Data Collection

The labelled data set is prepared from the discrete voltage reading of three NIR wavelengths at four channels and the simultaneous value of the invasive glucometer reading. The samples are obtained from 55 individuals in fasting (more than 8 hours after meal) and postprandial (2 hours after meals) states, following medical safety precautions. The details of the individuals are given in Table 3

A total of 204 data samples collected, out of which 142 samples (70%) are used for training the model and 64 samples (30%) are used for testing. In the proposed PLMR model, the data is grouped into clusters and sequentially organized in ascending order of their centroid distance from the origin. The piecewise approximation in PLMR represents a linear multivariate regression model for the data of two sequentially arranged clusters. At the intersection, continuity in the piecewise linear models is approximated by taking the averaged value of the intersecting models. The fitness of the PLMR model is observed for different numbers of clusters and analyzed in terms of the R^2 value shown in Table 4.

Table 3 — Characteristics of individuals

Gender-Age(years)	Normal	Pre-diabetic	Diabetic	Total
Male- 21-65	15	14	8	37
Female-20-58	9	5	4	18

Table 4 — Performance analysis of the PLMR model for different clusters

Cluster (k)	R^2	RMSE(mg/dl)	Execution Time (seconds)
3	0.8799	21.6312	0.0978
4	0.9011	19.4695	0.1181
5	0.9224	17.2833	0.1381
6	0.9257	16.8876	0.1792
7	0.9417	14.9145	0.1863
8	0.9565	12.8827	0.2013

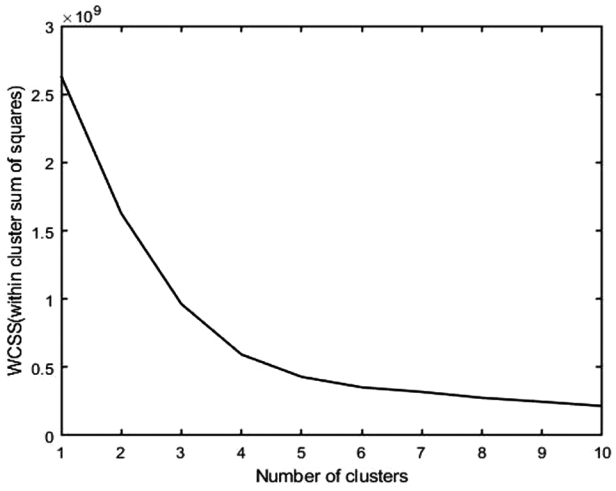

Fig. 9 — WCSS value for different number of clusters on training data set

The accuracy of the PLMR model increases with the increase in the number of clusters. However, as the number of clusters increases, it also increases the intersecting point, increasing complexity and execution time. Thus, to find the optimal number of clusters, we used the elbow technique, which is a plot of the within-clusters sum of squares (WCSS) value for different values of clusters (k) shown in Fig. 9.The WCSS value in Eq. (25) is the sum of the squared distances between data points and the cluster's centroid.

Table 5 — Statistical analysis of calibration of different regression models

Regression Models	R^2	RMSE(mg/dl)	MAE (mg/dl)	MSE(mg²/dl²)
Linear Regression	0.72	32.897	23.417	1082.2
Interaction Linear	0.83	25.422	16.026	646.3
Stepwise Linear	0.84	24.899	15.695	619.94
Fine tree	0.91	18.299	12.067	334.85
Medium tree	0.73	32.3	18.448	1043.3
Quardatic SVM	0.73	32.397	16.91	1048.4
Cubic SVM	0.74	31.852	18.385	1014.5
Medium Gaussian SVM	0.70	34.17	16.87	1167.6
PLMR	0.9224	17.283	4.90	298.70

$$WCSS = \sum_{c_k}^{c_n} \sum_{d_i inc_i}^{d_m} dis\tan ce(d_i, c_k)^2 \quad \ldots (25)$$

where c is the cluster centroid and d is the data point in each cluster.

The optimal value of k for the data is considered to be at the inflection point (elbow point) where the WCSS curve's second derivative becomes negative. In Fig.9, the inflation point is observed at k = 5, also for k=5 is validated via 10-fold cross-validation. Thus we used five- clusters which comprises into four PLMR model equation given by Eq. (26a- 26d):

$\hat{y}_1 = 457.2004 + 0.0064x_1 - 2.2548e^{-5}x_2 - 0.0277x_3 - 0.0043x_4$;
Model-1 for cluster 1 and 2. ... (26a)

$\hat{y}_2 = 343.9150 + 0.0013x_1 - 0.0015x_2 - 0.0059x_3 - 0.0072x_4$; Model-2 for cluster 2 and 3. ... (26b)

$\hat{y}_3 = 294.7678 + 0.0010x_1 - 0.0006x_2 - 0.0052x_3 - 0.0060x_4$;
Model-3 for cluster 3 and 4. ... (26c)

$\hat{y}_4 = 95.5659 - 0.0003x_1 + 0.0053x_2 - 0.0047x_3 - 0.0032x_4$;
Model-4 for cluster 4 and 5. ... (26d)

In the above expression, x_2 represent transmittance spectroscopy at 940 nm, while x_1, x_3 and x_4 represents reflectance spectroscopy at wavelengths 940 nm, 1050nm and 1300nm respectively. The piecewise linear multivariate regression model for pairs of clusters is represented by $\hat{y}_1, \hat{y}_2, \hat{y}_3$ and $\hat{y}_4$. The performance of the model is compared with the other regression model shown in Table 5.

The accuracy of the model's fit is examined in terms of R^2 value. A model with a higher R^2 value fits the data better, resulting in lower RMSE, MAE, and MSE values. In the table.4, the PLMR model with five clusters has a better accuracy (92.22%**)** than the other models, resulting in a low RMSE, MAE, and MSE of 17.283mg/dl, 4.90mg/dl and 298.70mg²/dl² respectively. The correlation plots between predicted and reference glucose concentration on the training data set in Fig. 10.

All simulations and model training were performed using MATLAB. The bar plot in Fig. 11, shows the predicted and reference values on the testing data set.

5.2 Clarke Error Grid

The Clarke error grid (CEG) plot is used to analyze the device's clinical accuracy. The variation between the referenced and predicted glucose values is explored by various zones in the Clarke Error plot. The values that exist in Zone A are preferable because they show a small difference between the predicted and reference blood glucose levels. The Clarke error grid analysis shown in Fig. 12, shows most of the predicted glucose values fall within region A, which is clinically acceptable for glucose measurement devices.

The proposed device has been validated for blood glucose measurements in the range of 79 mg/dl - 390 mg/dl of glucose concentration. The performance of the device is compared with the previous non-invasive works as shown in Table 6.

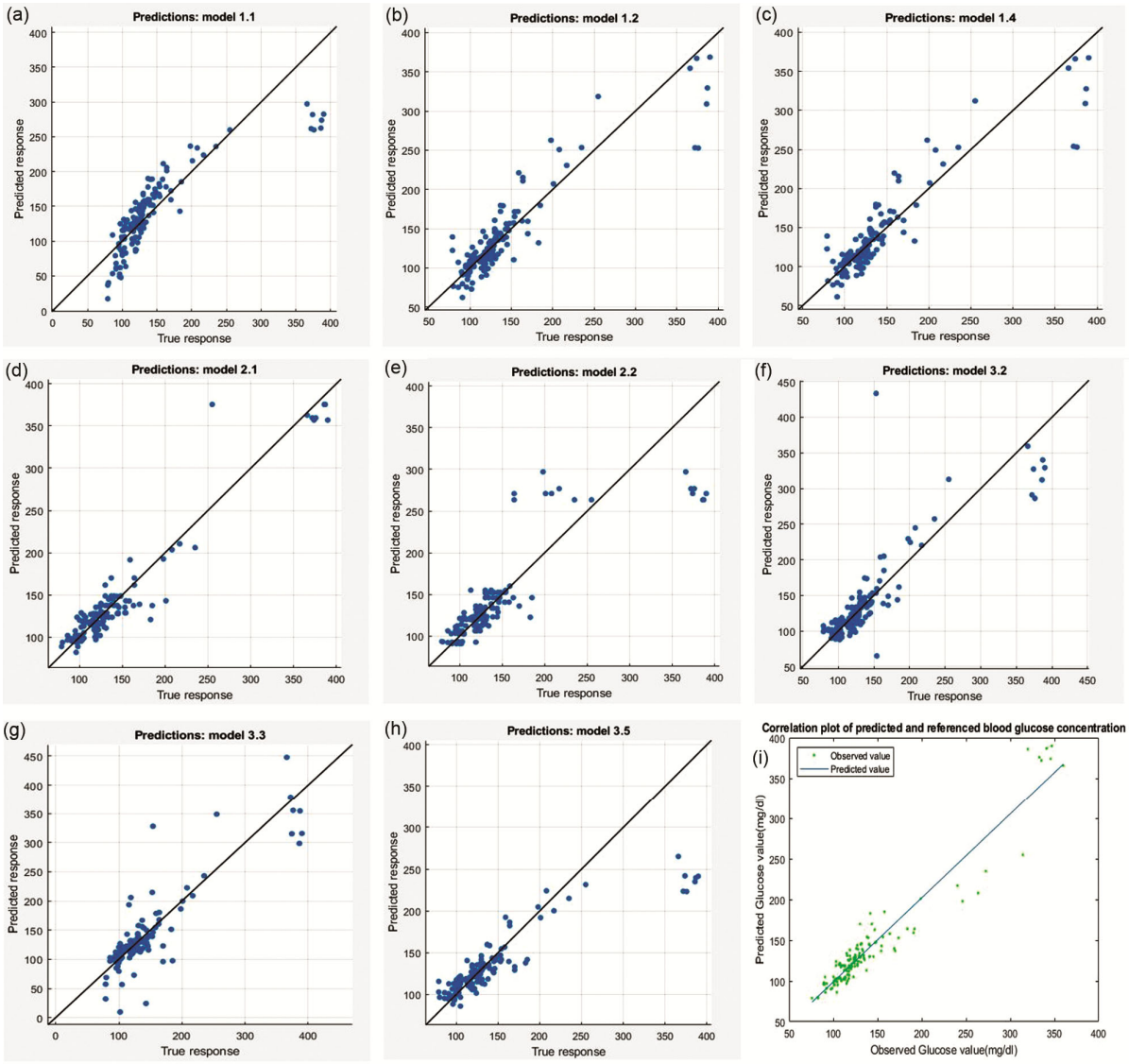

Fig. 10 — Correlation plot between predicted and referenced blood glucose concentrations on training data using various regression models

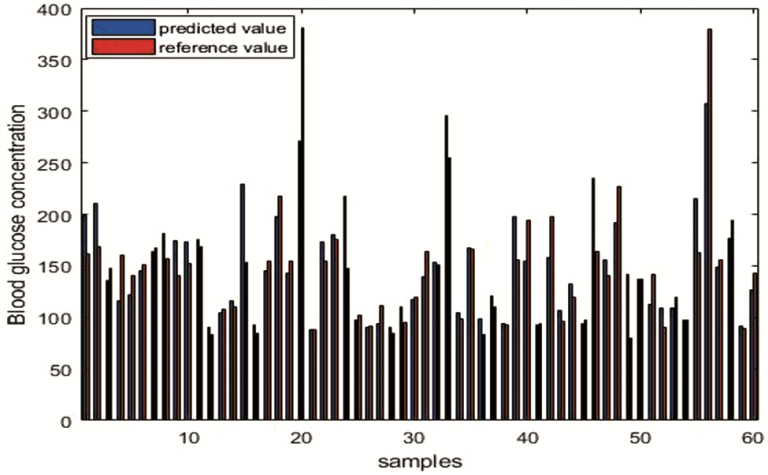

Fig. 11 — Bar plot between predicted and reference glucose value on testing data set

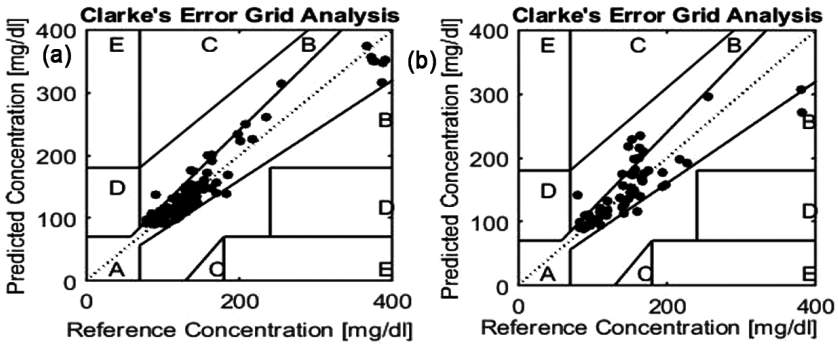

Fig. 12 — Clarke Error grid plot (a) Training data sets (b) Testing data sets

Table 6 — Statistical comparison of previous non-invasive works

Research works	Wavelength (nm)	Spectroscopy	sample	RMSE (mg/dl)	R^2	Device cost
[1]	940,1300	Transmittance and Reflectance	Blood (capillary and serum)	13.57	0.94	high
[4]	940	Reflectance	Blood	18.88	--	low
[14]	940,1300	Transmittance and Reflectance	Blood (capillary)	7.5	0.90	high
[21]	1550	Transmittance	Blood	--	0.74	low
Proposed	940,1050,1300	Transmittance Reflectance	Blood	17.2833	0.9224	low

6 Conclusion

We used the vibrational properties of glucose molecules under NIR light to estimate their concentration. The concept of transmittance and reflectance spectroscopy is used to develop a non-invasive glucose measurement device. This method has several advantages over traditional methods, including painless, convenient, and simple for routine monitoring. In this prototype, the intensity of NIR transmitted light at 940 nm, reflected light at 940 nm, 1050 nm and 1300 nm wavelengths is analyzed using LEDs and photo-detectors. The discrete voltage values are used to develop a proposed PLMR model for non-invasively estimating blood glucose levels. The device is built on a printed circuit board, and all simulations are carried out using MATLAB software. The developed prototype has a good correlation of 92.24% with a low RMSE value of 17.28mg/dl, indicating that it can be used in practice. The device is also validated on Clarke Error Grid Plot and demonstrates the feasibility of using a non-invasive blood glucose measurement technique.

References

1 Joshi A M, Jain P, Mohanty S P & Agrawal N, *IEEE Trans Consum Electron*, 66 (4) (2020) 327.

2 Haxha S & Jhoja J, *IEEE Photon J*, 8 (6) (2016) 1.

3 Dorsaf G, Yacine M & Khaled N, *Curr Res Diabetes Obes J*, 11 (1) (2019) 1.

4 Narkhede P, Dhalwar S & Karthikeyan B, *Ind J sci technol*, 9 (41) (2016) 1.

5 Yang X & Zhang A Y, *IEEE Conference on Lasers and Electro-Optics,* (2012) 1.

6 Tajima T, Okabe Y, Tanaka Y & Seyama M, *IEEE Sensors J*, *17* (16) (2017) 5079.

7 Pai P P, Sanki P K, Sahoo, S K, De A, Bhattacharya S & Banerjee S, *IEEE Trans Circuits Syst I Regul Pap*, 65 (2) (2017) 663.

8 Khan M S, Varshney G & Giri, P, *IEEE Trans Nano Bioscience*, 20 (4) (2021) 488.

9 Li D, Yu H, Huang X, Huang F, Hu X & Xu K, *Opt Diagnos Sensing VII SPIE,* 6445 (2007) 73.

10 Pleitez M A, Lieblein T, Bauer A, Hertzberg O, von Lilienfeld-Toal H & Mäntele W, *Rev sci instrum*, 84 (8) (2013) 084901.

11 Prasad M S, Chen R, Li Y, Rekha D, Li D, Ni H & Sreedhar N Y, *IEEE Sensors J*, 18 (13) (2018) 5203.

12 Singh A K & Jha S K, *IEEE Sensors J*, 19 (18) (2019) 8332.

13 Dai T & Adler A, *IEEE Trans Instrum Meas*, 58 (11) (2009) 3831.

14 Song K, Ha U, Park S, Bae J & Yoo H J, *IEEE J solid-state circuits*, 50 (4), (2015) 1025.

15 De Pretto L R, Yoshimura T M, Ribeiro M S & Zanardi de Freitas A, *J biomedical opt*, 21 (8) (2016) 086007.

16 Maruo K, Tsurugi M, Chin J, Ota T, Arimoto H, Yamada Y & Ozaki Y, *IEEE J sel top quantum electron*, 9 (2) (2003) 322.

17 Guo D X, *J Biosci Med*, 3 (6) (2015) 42.

18 Padalkar M V & Pleshko N *Analyst*, 140 (7) (2015) 2093.

19 Huang M, Kim M S, Chao K, Qin J, Mo C, Esquerre C & Zhu Q, *Sensors*, 16 (4) (2016) 441.

20 Maier J S, Walker S A, Fantini S, Franceschini M A & Gratton E, *Opt lett*, 19 (24) (1994) 2062.

21 Uwadaira Y, Ikehata A, Momose A & Miura M, *Biomed opt express*, 7 (7) (2016) 2729.

22 Zhang W, Liu R, Zhang W, Jia H & Xu K, *Biomed opt express*, 4 (6) (2013) 789.

23 Yang W, Liao N, Cheng H, Li Y, Bai X & Deng C, *Aip Advances*, 8 (3) (2018) 035216.

24 Golic M, Walsh K & Lawson P, *Appl spectrosc*, 57 (2) (2003) 139.

25 Daarani P & Kavithamani A, *Int J Latest Trends Eng Technol,* 1 (2017) 141

26 Saleh G, Alkaabi F, Al-Hajhouj N, Al-Towailib F & Al-Hamza S, *J med eng technol*, 42 (2) (2018)140.

27 Alexopoulos E C, *Hippokratia*, 14 (1) (2010) 23.

28 Sinha P, *Int J Sci Eng Res*, 4 (12) (2013) 962.

29 Uyanık G K & Güler N, *Procedia Soc Behav Sci*, 106 (2013) 234.

30 Rustam F, Reshi A A, Mehmood A, Ullah S, On B W, Aslam W & Choi G S, *IEEE access*, 8 (2020)101489.

31 Cai J, King J, Yu C, Liu J & Sun L, *IEEE Micro Wirel Comp Lett*, 28 (5) (2018) 428.

32 Kavitha S, Varuna S & Ramya R, *online int confer green eng technol (IC-GET) IEEE,* (2016) 1.

33 Bhargava N, Purohit R, Sharma S & Kumar A, *2nd Int Confer Commun Electron Sys (ICCES) IEEE,* (2017) 606.

34 Loh W Y, *Wiley interdisciplinary reviews: data mining and knowledge discovery*, 1 (1) (2011) 14.

35 Li Y & Wu H, *Phys Procedia*, 25 (2012) 1104.

36 Singh A, Kotiyal V, Sharma S, Nagar J & Lee C C, *IEEE Access*, 8 (2020) 208253.