Indian Journal of Pure & Applied Physics
Vol. 60, August 2022, pp. 700-706
DOI: 10.56042/ijpap.v60i8.63861

Dual Wavelength based Approach with Partial Least Square Regression for the Prediction of Glucose Concentration

Deepshikha Yadav[a,b], Manjri Singh[a,b], Sahil Sharma[a,b], Surinder P Singh[a,b] & P K Dubey[a,b*]
[a]CSIR- National Physical Laboratory, Dr. K. S. Krishnan Marg, New Delhi 110 012, India
[b]Academy of Scientific and Innovative Research (AcSIR), Ghaziabad 201 002, India

Received 17 June 2022; accepted 14 July 2022

Diabetes mellitus is a group of metabolic disorder characterized by high blood sugar levels. Monitoring of blood glucose levels at regular intervals plays a crucial role in the management of diabetes. The non-invasive real-time monitoring of glucose using near-infrared (NIR), Raman, acoustic and bio-impedance techniques have an edge over available invasive techniques but suffers from low Signal to Noise ratio (S/N) and other interferences. In the present work, we have attempted to improve S/N for the efficient detection of feeble signals corresponding to the physiological glucose concentrations. Investigations were carried out in the NIR region particularly from 800-1400 nm for the identification of the unique absorption features of glucose using UV-Vis NIR spectrophotometer with different ranges of glucose concentrations including 5 g/dl- 45 g/dl, 1400 mg/dl -2500 mg/dl, 35 mg/dl-650 mg/dl . Savitzky Golay (SG) pre-processing filter was applied on the raw data for enhancing the S/N for better prediction of glucose concentrations. The absorption spectra of glucose revealed the presence of a peak at 960 nm. Therefore, considering the absorbance at 960 nm, provided an enhancement in the S/N ratio from 17 dB to 27 dB. Further, partial least square regression (PLSR), has been applied on SG filtered data for a better prediction of glucose concentration with a correlation coefficient (R^2) value of 0.99 and root mean square error of prediction (RMSE) of 2.29 mg/dl. Further, based on the NIR spectral data, we have developed a measurement technique using two LED sources of 950 nm and 860 nm, and a wide detector (700 - 1100 nm) which converts obtained optical signal into voltage. It has been observed that by considering dual wavelength detection points the prediction of glucose concentration is improved. Furthermore, with increase in the test glucose concentrations, the voltage signal decreased corresponding to the 950 nm LED. This is attributed to reduced signal intensities reaching the photodiode as a result of the increase in glucose absorption. Incorporating dual wavelengths for PLSR reduced the RMSE from 8.98 mg/dl to 6.49 mg/dl and also improved the R^2 value from 0.97 to 0.99.

Keywords: Non-invasive; NIR spectroscopy; Savitzky golay; Partial least square regression (PLSR); Multivariate calibration

1 Introduction

The frequent monitoring and control of blood glucose levels is a pre-requisite for the diabetic patients to prevent the health hazards such as coronary heart disease, stroke, retinopathy, nephropathy, convulsions, coma etc associated with hyper/hypo glycemia[1]. According to the World Health Organization (WHO), the number of people with diabetes has risen from 108 million in 1980 to 422 million in 2014 which are estimated to nearly double by the year 2040[2].

The popular glucometers, which involve electrochemical enzyme based detection of glucose levels in blood are invasive in nature. This method may cause traumatic experience to the users particularly in certain special cases like children with diabetes or elderly patients. The expensive one time use needles and risk of infection further limits their frequent usability. Furthermore, minimally invasive techniques like the continuous glucose monitoring (CGM) devices are slowly gaining momentum. However, their low accuracy in comparison to conventional glucometers is a significant obstacle[3]. To this end, non-invasive and real time glucose measurement may provide an edge over the existing invasive glucose monitoring techniques due to its fast and non-destructive nature.

Among the non-invasive techniques, polarimetry, Raman spectroscopy and photo-acoustic spectroscopy are quite popular but suffer from limitations like high component costs, low sensitivity in physiological glucose range and interferences from other components in the body[4-7]. Absorption spectroscopy in the near-infrared (NIR) region is one of the most popular and sought after analytical methods for measuring glucose concentrations as the NIR

*Corresponding authors: (Email: premkdubey@gmail.com)

radiation has a good penetration depth in biological tissues and is absorbed by only a few biological chromophores[8,9]. Thus, NIR absorption spectroscopy is frequently applied in medical science. However, one of the major challenges faced during NIR measurement is the high spectral noise (due to high frequency noise, baseline variations, and overlapping spectral peaks) which masks the relevant features of analyte of interest[10,11].

Apart from the above said, the low glucose specific absorption in NIR region of the spectrum also contributes to a low S/N which affects the sensitivity and thereby, the limit of detection of device[12]. Pre-processing of the collected spectral data is a possible solution for enhancing the poor S/N and improving the subsequent chemometric analysis including techniques like classical least-squares (CLS), inverse least-squares (ILS) regressions, multiple linear regression (MLR), principal component analysis and regression (PCA and PCR), partial least squares regression (PLSR) and net analyte signal (NAS). The different types of pre-processing techniques are multiplicative scatter correction (MSC), standard normal variate (SNV), de-trending, Norris-Williams (NW) derivatives and Savitzky-Golay (SG) polynomial derivative filters etc[13].

Studies have been carried out using pre-processing filters for improving the chemometric analysis. A combination of two pre-processing techniques including the bandpass and SG filters have been employed to improve the prediction performance of PCR and PLSR tools in NIR spectroscopy for the quantitative analysis of aqueous glucose solutions. The results show that the approach improved the prediction accuracy of PCR and PLSR models with an RMSE value of 28.9 mg/dl and R^2 value of 0.99[14]. In addition, exponential moving average (EMA) filter has been used for the quantitative analysis of the solution containing glucose, urea and triacetin. The filter length was optimized for improving the S/N. An RMSE of 22.4 mg/dl was obtained by using the EMA filtered data for PLSR based modelling[15]. Furthermore, a new pre-processing approach involving the combined application of Chebyshev filter with the baseline correction technique, asymmetric least squares (ALS) and SG filter, has been adopted for the prediction of glucose concentrations by using PLSR and PCR from NIR spectra. An RMSE value of 12.76 mg/dl has been achieved through the combined application of these filters for aiding the prediction[16]. However, in none of these studies, the improvement in the S/N ratio has been quantified.

Considering the necessities and impediments in the existing glucose detection techniques, in the present work, we have proposed an NIR based technique that utilizes dual wavelength (950 nm and 860 nm) that shows potential to be used for the detection of blood glucose concentration. In the experimentation, NIR spectroscopy has been utilized for identifying glucose specific peaks. To resolve the issue of high spectral noise and to have a better S/N ratio, SG filter has been applied to the raw absorption spectra and a significant improvement in the S/N ratio has been achieved from 17dB to 27 dB[17]. The pre-processed data with the high S/N ratio was used for the further predictive analysis using PLSR technique. For developing NIR based glucose measurement technique two LED sources of 950 nm and 860 nm along with the broad range LED (700 - 1100 nm) for the detection have been used. The consideration of LED of 860 nm is for the background cancellation. The effectiveness of the technique was finally verified by applying PLSR on the voltage data generated by the LEDs.

2 Materials and Methodology

Standard glucose solution (45%) was procured from Sigma Aldrich. The NIR spectra of different glucose concentrations were acquired by Agilent Cary 5000 UV-Vis-NIR spectrophotometer having the full spectral bandwidth from 175 nm to 3300 nm. The measurements were carried out using a quartz cuvette as a sample holder having 10 mm path length. The studies were focused particularly in the second/third overtone band (800-1400 nm) of NIR region at ambient temperature. In addition, LEDs of 860 nm (spectral FWHM bandwidth of 30 nm), 950 nm (spectral FWHM bandwidth of 42 nm) and a photodiode having a wide spectral bandwidth of 700-1100 nm, with peak sensitivity at 940 nm were used for the development of NIR based glucose measurement technique.

SG pre-processing filtering technique has been used to improve S/N of the acquired raw spectra. The chemometric multivariate calibration technique PLSR has been utilized for the prediction of glucose concentrations.

3 Results and Discussion

3.1. NIR absorbance spectra of glucose powder

To identify the dominant glucose signatures, spectra of glucose powder in the NIR region was acquired. For recording the spectra, a pellet of glucose powder was prepared in KBr. Fig. 1, shows the acquired spectra of the glucose powder, the increase in the absorbance can be observed due to the glucose specific absorption at about 920 nm which may be attributed to the third overtone stretching of CH_2 groups[18].

3.2. NIR absorbance of glucose solutions

Examining the effect of glucose on the absorption of water in the NIR region is a critical parameter for non-invasive glucose sensing as water is the dominant background signal in the human body and it exhibits strong absorption throughout the NIR region[19]. Therefore, to analyse the variation in the absorption of glucose solutions, we prepared solutions of very high concentrations (1400-2500 mg/dl) due to its low intrinsic absorption. The NIR spectra of glucose concentrations of 1400 mg/dl, 1600 mg/dl, 2000 mg/dl and 2500 mg/dl along with water is shown in Fig. 2. The peak corresponding to water specific absorption at ~ 975 nm may be attributed to the second overtone OH stretching[18]. The decrease in the absorption of water with increasing glucose concentration was observed as shown in the inset of Fig. 2. Extremely small variations in the absorption, particularly due to glucose is a critical problem. Therefore, non-invasive detection of glucose concentrations becomes challenging.

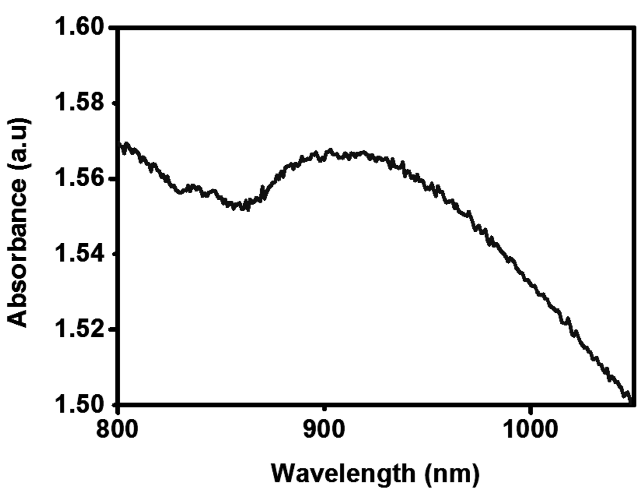

Fig. 1 — NIR spectra of glucose powder showing the dominant absorption peak of glucose at ~920 nm.

Further, in order to study, the absorption of glucose in the normoglycemic, hypoglycemic and hyperglycemic levels of the human body, different glucose solutions having concentrations of 35 mg/ml, 100 mg/dl, 200 mg/dl, 325 mg/dl, 480 mg/dl and 650 mg/dl were prepared and their absorbance spectra was recorded. Lowering down the glucose concentrations upto the physiological levels become even more challenging because of the small variation in the glucose absorption. Fig. 3 shows the typical acquired glucose concentration spectra with water baseline correction. As the raw data is having weak glucose absorption signal, therefore it suffers from low S/N. Hence, further pre-processing is required in order to improve the S/N for extracting meaningful glucose specific information from the NIR spectra.

The significant noise reduction in the NIR data was accomplished by applying SG filter on the raw data with an improvement in the S/N from 17 dB to 27 dB

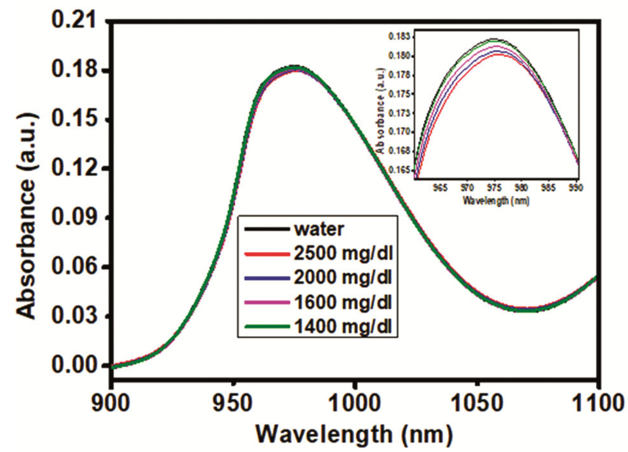

Fig. 2 — NIR spectra of glucose solution showing the small variation in absorption due to glucose concentration at ~ 975 nm.

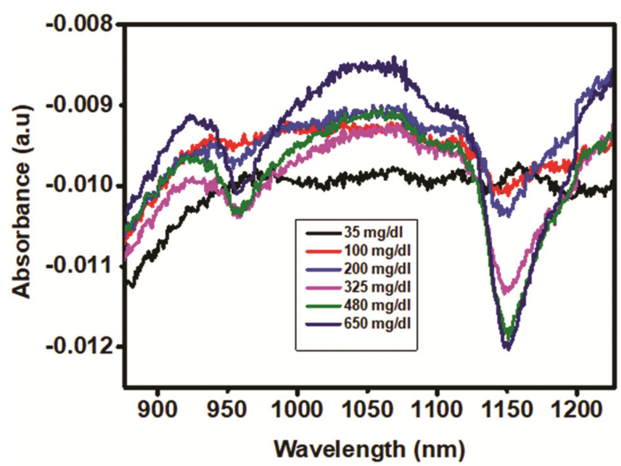

Fig. 3 — Raw NIR spectra of glucose solutions showing low S/N.

as shown in the Fig. 4. The peak at 960 nm corresponds to the second overtone stretching of OH and the peak at 1150 nm corresponds to the second overtone CH stretch[18,20].

There is shift observed in the pure water peak from 975 nm to ~960 nm after the addition of glucose, which corresponds to the second overtone OH stretch. This could be attributed to the fact that the OH functional groups present in water are involved in hydrogen bonding with solute molecules. H bonding will affect the overall spectra and hence the stretching and bending frequencies of the OH bonds reflecting a shift in the absorption wavelength to 960 nm due to glucose[18]. Further, an increase in glucose molecules results in fewer free water molecules resulting in a reduction in the water specific absorbance[19].

To study the variation in absorption at 960 nm, prominently, almost 100 fold concentrations of glucose compared to the physiological range were chosen. Studies were carried out on the glucose concentrations of 5 g/dl, 15 g/dl, 20 g/dl, 25 g/dl, 40 g/dl and 45 g/dl. It is clearly observed that with rising glucose concentration the absorption at 960 nm is becoming more negative as shown in Fig. 5. Further, this change may be explained by the fact that with increase in glucose concentration, the effective absorption of glucose solution at 960 nm decreases as a result of the hydrogen bonding between glucose and water leading to the reduction in the free water molecules, thereby minimizing the effect of the water specific absorption[21]. This occurs because of the displacement of a finite volume of water as glucose molecules take up space in the solution[22]. It is imperative that we take into account the effect of displacement of water due to glucose, because for the detection of glucose real time, the effects from water specific absorption cannot be neglected, as a result of the dominating background of water in the human body.

3.3. Glucose prediction in the normoglycemic, hypoglycemic and hyperglycemic levels using PLSR

Owing to the complex molecular structure of most biomolecules, the NIR spectra consist of overlapping spectral features that may correspond to molecules other than glucose[10]. Therefore, for non-invasive blood glucose sensing multivariate calibration, techniques may be preferred for accurate glucose prediction in blood because they help in building prediction models for fishing out analyte specific information from a complex matrix with substantial S/N[23,24]. In the present study PLSR was used to predict the glucose concentration using the acquired spectroscopic data. Further, the window points of the SG filter were optimized based on the best prediction model obtained using PLSR with the lowest RMSE values and highest R^2 value. The SG filter optimization at 7 points is shown in Table 1.

Different glucose concentrations in the range from 45-200 mg/dl (covering, hypoglycemic-normoglycemic-hyperglycemic conditions) were prepared and their NIR spectroscopic data was recorded. The concentrations used for the studies were 45 mg/dl, 65mg/dl, 90 mg/dl, 110 mg/dl, 135 mg/dl, 155 mg/dl, 180 mg/dl and 200 mg/dl. 5 spectral scans for each concentration were acquired. In our study the specific absorption points at 924 nm, 940 nm and 960 nm were used for model building. The objective of choosing the specific wavelengths was to include the peak which was highly sensitive to glucose concentration changes at 960nm[25,26]

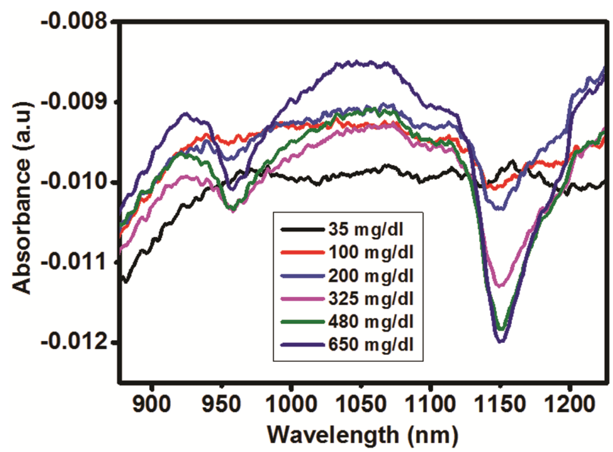

Fig. 4 — NIR spectra of glucose solutions after applying SG filter showing significant improvement in the S/N from 17 dB to 27 dB.

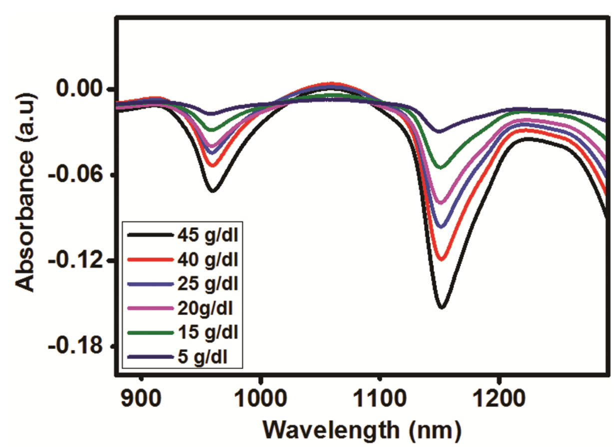

Fig. 5 — Water displacement effect in glucose solutions at high concentrations.

and the peaks which hardly showed any variation in response to glucose concentration at 924 nm and 940 nm. We applied the PLSR technique to generate two models. In the first case SG filter was not applied to the raw spectra and the data was used as it is as shown in Fig. 6. In the second case data was pre-processed using SG filter, before applying PLSR as shown in Fig. 7.

The results reveal that there was a significant improvement in the prediction capability of the model after the application of the SG filer to the raw data. The R^2 value was greatly increased from 0.76 without application of SG filter to 0.99 after application of SG filter, and the RMSE was also significantly reduced from 35.22 mg/dl without application of SG filter to 2.29 mg/dl after application of SG filter.

Table 1 — Optimization of window points for SG filter.

Window points	RMSE (mg/dl)	Correlation Coefficient (R^2)
3	35.77	0.75
4	13.77	0.96
5	13.77	0.96
6	2.29	0.99
7	2.29	0.99

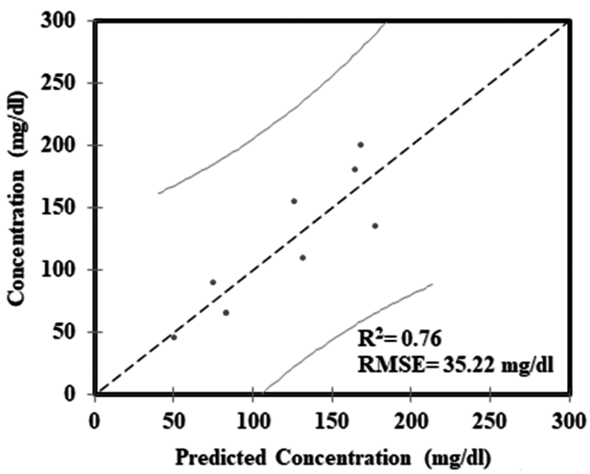

Fig. 6 — Prediction of glucose concentrations using the raw NIR spectroscopic data without the application of pre-processing SG filter.

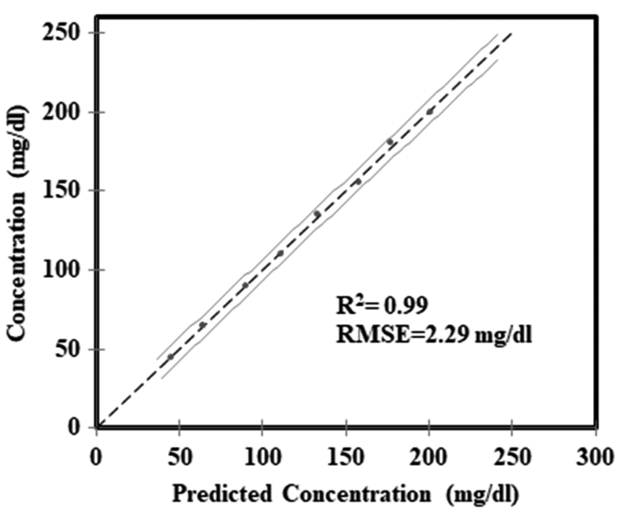

Fig.7 — Prediction of glucose concentrations using the raw NIR spectroscopic data after the application of pre-processing SG filter.

3.4. Prediction of glucose concentration using developed NIR based glucose measurement technique

In order to explore the possibility for development of a prototype device for glucose prediction a simple dual wavelength based circuit was developed as shown in Fig. 8. It consists of two LED sources of 950 nm (OSRAM Opto Semiconductors GmbH, model: SFH 4141) and 860 nm (OSRAM Opto Semiconductors GmbH, model: SFH 4258-Z) along with a broadband photo detector having spectral range from 700 to 1100 nm (Wurth Elektronik, model: 1541201EEA400)

Measurements were carried out on the same concentration of glucose solutions, i.e., 45 mg/dl, 65mg/dl, 90 mg/dl, 110 mg/dl, 135 mg/dl, 155 mg/dl, 180 mg/dl and 200 mg/dl in a cuvette of 10 mm path length. The cuvette was placed in a cuvette holder equipped with source LEDs and wide band photodetector as shown in Fig. 9.

A linear decrease in voltage with increase in glucose concentration was observed which may be attributed to the increased absorption of glucose and decreased absorption of water at 950 nm as shown in Fig. 10 due to an increase in the number of glucose molecules displacing the water molecules. In other words, the increased absorption due to glucose leads to the reduced intensity of NIR light reaching the photo detector, which results in decrease in voltage response. The voltage response at 860 nm remains almost constant until 125 mg/dl and seems it does not represent absorption feature of glucose. The slight decrease in voltage response beyond this concentration level is due to the spectral bandwidth of 860 nm LED (FWHM of 30 nm). Therefore, relatively lower sensitivity for glucose has also been observed in the 860 nm LED compared to the 950 nm LED.

The data was used to build two prediction models using PLSR. The first model was generated based on the voltage data corresponding to the LED of 950 nm and the second model was generated based on the voltage data corresponding to LEDs of both 950 nm and 860 nm. The voltage response was recorded 10 times for each concentration of glucose. The first

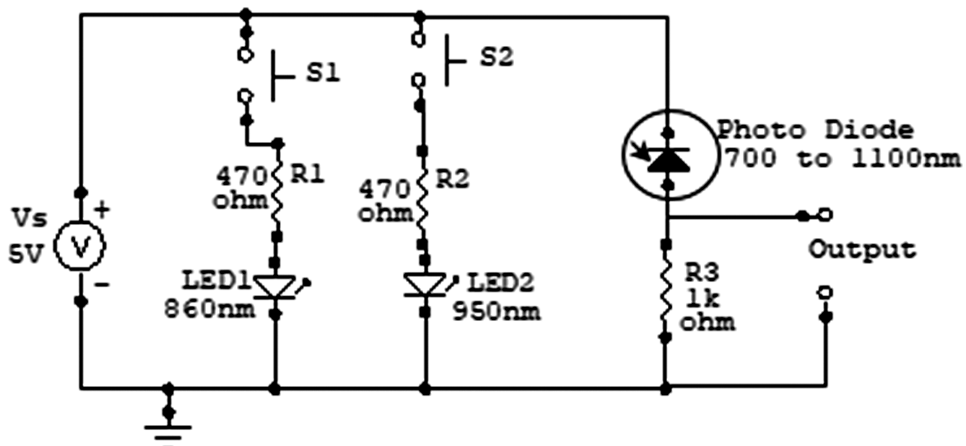

Fig. 8 — Circuit depicting the developed glucose measurement technique using dual wavelength approach.

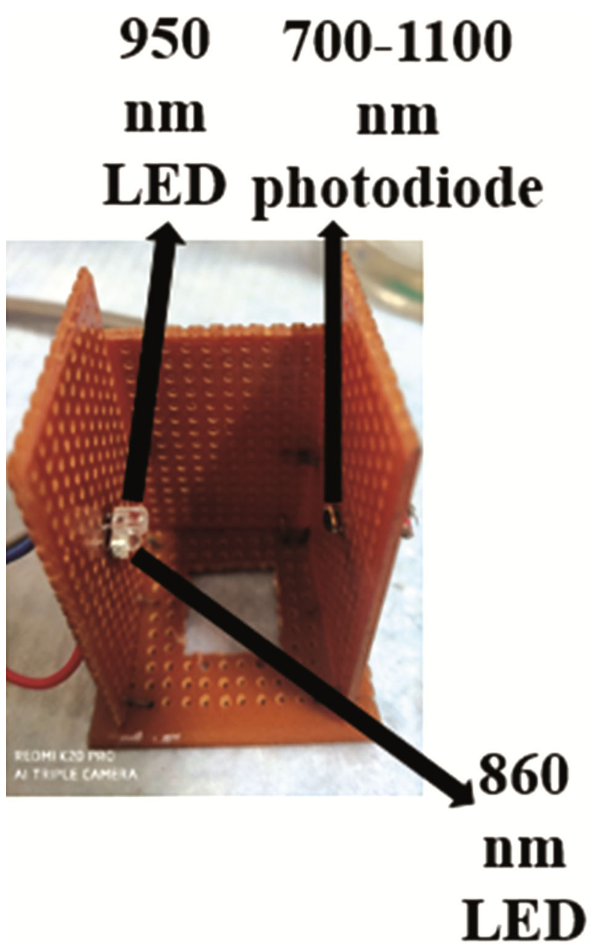

Fig. 9 — A typical arrangement for the cuvette holder equipped with source LEDs and wide band photodetector.

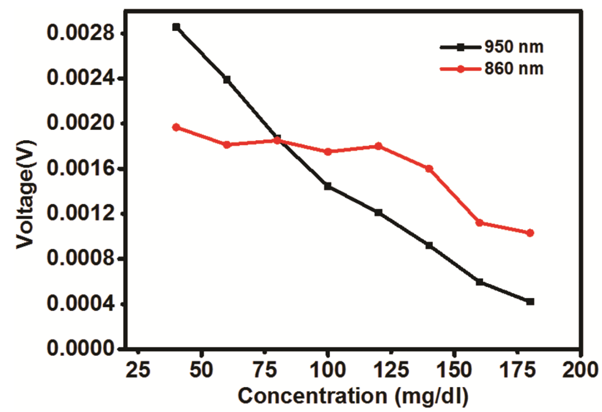

Fig. 10 — Voltage response of the developed NIR based glucose measurement technique.

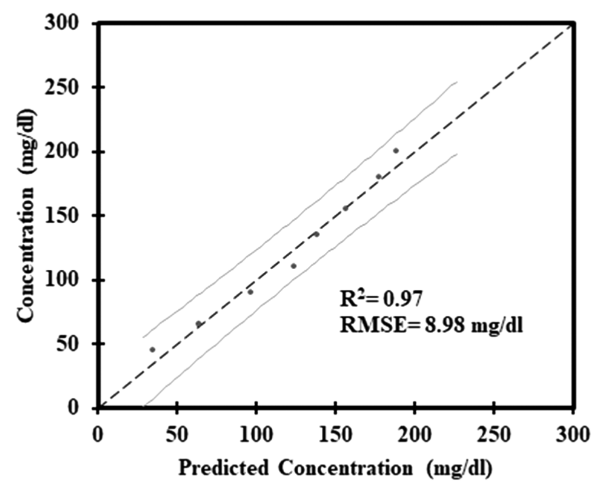

Fig. 11 — Prediction of glucose concentrations based on the voltage data corresponding to the LED of 950 nm.

model predicted the glucose concentrations with an R^2 value of 0.97 and RMSE of 8.98 mg/dl as shown in Fig. 11.

However, the second model predicted the glucose concentration with a higher R^2 value of 0.99 and the RMSE value was also lower at 6.49 mg/dl as shown in Fig. 12. From the result it is evident that the dual wavelength approach improves the prediction of glucose concentration.

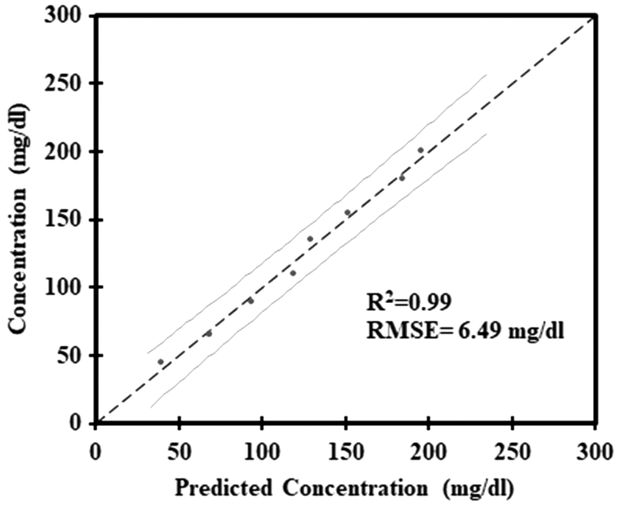

Fig. 12 — Prediction of glucose concentrations based on voltage response corresponding to both 950 nm and 860 nm LED.

4 Conclusion

In the present work, the NIR region from 800-1400 nm has been investigated for unique absorption features of glucose and glucose specific peak was observed at 960 nm. SG filter was applied on raw spectra for enhancing S/N. Significant enhancement in the S/N ratio was observed from 17 dB to 27 dB, resulting in an overall improvement of 10 dB. Although, PLSR model applied to raw spectra provides RMSE of glucose concentration of 35.22 mg/dl but by using SG pre-processing to raw data and applying PLSR onto it the RMSE becomes significantly reduced to 2.29 mg/dl. Further, considering dual wavelength measurements along with PLSR the RMSE reduced to 6.49 mg/dl from 8.98 mg/dl in comparison to measurements at 960 nm alone indicating the potential of the dual wavelength approach in improving the prediction of glucose concentration.

Acknowledgments

The authors thank the Director, NPL for providing necessary research facilities. Deepshikha Yadav gratefully acknowledges the Senior Research Fellowship (SRF) from Indian Council of Medical Research (ICMR), grant number: ISRM/11(60)/2017, dated: 17/4/2018.

Conflicts of Interest

The authors report no conflicts of interest in the above research work.

References

1 Nathan D M, Zinman B, Cleary P A, Backlund J C, Genuth S, Miller R & Orchard T J, *Arch Intern Med*, 169 (2009) 1307.
2 https://www.who.int/news-room/fact-sheets/detail/diabetes. Accessed on 24/10/2021
3 Gonzales W V, Mobashsher A T & Abbosh A, *Sensors*, 19 (2019) 800.
4 Cameron B D & Cote G L, *IEEE Trans Biomed Eng*, 44 (1997) 1221.
5 Wood M F G, Cote D & Vitkin I A, *J Biomed Opt*, 13 (2008) 044037.
6 Kang J W, Park Y S, Chang H, Lee W, Singh S P, Choi W, Galindo L H, Dasari R R, Nam S H, Park J & So P T C, *Sci Adv*, 6 (2020) 5206.
7 De Pretto L R, Yoshimura T M, Ribeiro M S & de Freitas A Z, *J Biomed Opt*, 21 (2016) 086007.
8 Zhang H, Salo D, Kim D M, Komarov S, Tai Y C & Berezin M Y, *J Biomed Opt*, 21 (2016) 126006.
9 Tang J Y, Chen N Y, Chen M K, Wang M H & Jang L S, *Jpn J Appl Phys*, 55 (2016) 106601.
10 Ozaki Y, Ikehata A & Shinzawa H, Near-Infrared Spectroscopy in Biological Molecules and Tissues', in Encyclopedia of Biophysics, Edited by Roberts G C K, Springer, Heidelberg, Berlin, (2018) 1695.
11 Ham F M, Kostanic I N, Cohen G M & Gooch B R, *IEEE Trans Biomed Eng*, 44 (1997) 475.
12 Uwadaira Y, Adachi N & Ikehata A S K, *J Near Infrared Spectrosc*, 18 (2010) 291.
13 Rinnan A, Berg F V D & Engelsen S B, *Trends Anal Chem*, 28 (2009) 1201.
14 Patchava K C, Alrezj O, Benaissa M & Behairy H, 38th *An Int Conf IEEE Eng Med Biol Soc*, Orlondo, Florida, USA, (2016) 6210.
15 Al-Mbaideen A A, *J Anal Chem*, 74 (2019) 686.
16 Mishal M R, Islam T T, Antor S H & Rahman T, *Proceedings*, 2 (2018) 1010.
17 Savitzky A & Golay M J E, *Anal Chem*, 36 (1964) 1627.
18 Golic M, Walsh K & Lawson P, *Appl Spectrosc*, 57 (2003) 139.
19 Goodarzi M, Sharma S, Ramon H & Saeys W, *Trends Anal Chem*, 67 (2018) 147.
20 Shenk J S, Workman J J & Westerhaus M O, Application of NIR spectroscopy to agricultural products, in Handbook of Near-Infrared Analysis, Edited by Dekker M, CRC Press, New York, (2001) 419.
21 Tarumi M, Shimada M, Murakami T, Tamura M, Shimada M, Arimoto H & Yamada Y, *Phys Med Biol*, 48 (2003) 2373.
22 Jung Y & Hwang J, *Appl Spectrosc*, 67 (2013) 171.
23 Olivieri A C, Faber N M, Ferre J, Boque R, Kalivas J H & Mark H, *Pure Appl Chem*, 78 (2006) 633.
24 Heise H M, Delbeck S & Marbach R, *Biosensors*, 11 (2021) 64.
25 Liu J, Zhu C, Jiang J & Xu K, *Biomed Opt Express*, 9 (2018) 5903.
26 Yang W, Liao N, Cheng H, Li Y, Bai X & Deng C, *AIP Adv*, 8 (2018) 035216.