Influence of PCA Components on Glucose Prediction using Non-invasive Technique

J.S.Parab, R.S.Gad, G.M.Naik
Department of Electronics
Goa University
Taleigao Plateau, India
{jsparab,rsgad,gmnaik}@unigoa.ac.in

***Abstract*—A truly noninvasive blood glucose-sensing device could revolutionalize diabetes treatment and coupled with advances in microelectronics can improve compliance with recommended glucose levels and greatly influence the cost of diabetes monitoring. The paper describes a Multivariate PLSR system for blood glucose prediction by considering 5 major variants in human blood i.e Glucose, Alanine, Ascorbate, Lactate and Urea. The typical biological system that resembles human blood tissue consisting of five major constituents has been tested on the Partial Least Square Regression (PLSR) model reported by the authors elsewhere. Multivariate PLSR model is experimentally validated for 12 templates recorded using Schimatzu FTIR 8400S in the range 400cm^{-1} to 5000 cm^{-1}. The model was validated using 2 approaches, namely, Root Mean Square Error (RMSE) and Clark Error Grid Analysis (CEGA) plot. The dependence of variate prediction on Principal Component Analysis(PCA) factors is also described and analyzed in detail.**

Keywords— Non Invasive, PLSR, PCA, Diabetes, Multivariate.

I. INTRODUCTION

Measurement of glucose level in blood has long been considered as an essential need for self-monitoring in diabetes and for screening in pre-diabetes. For diabetes management in particular, frequent measurements of Blood glucose level is inevitabl[1], and thus many kinds of portable devices for self-monitoring of blood glucose have been commercialized worldwide. However, many of the current self monitoring devices are based on the patient puncturing their skin with a small needle or lancet and squeezing the surrounding tissue to remove a blood drop. The diabetes case where frequent monitoring of glucose is necessary, the repeated procedure of skin puncturing becomes painful and troublesome and, furthermore, can cause an infection. Although a non-puncturing type device, such as GlucoWatch Biographer, which is based on reverse iontophoresis to draw glucose molecules via inner skin[2] has been approved by the FDA, its measurement procedure can still cause skin irritation after repeated applications[3].Extensive research has focused on the development of continuous (or nearly continuous) blood glucose monitors that can provide valuable information related to changes in glucose concentration. Both the rate and the direction of changes in glucose concentration are important pieces of diagnosis.

Noninvasive glucose sensing is yet another measurement strategy that promises pain-free operation without the complications of an adverse biological response[4],[5],[6] In the most common implementation, a noninvasive measurement involves the passage of a selected band of electromagnetic radiation through a vascular region of the body i.e earlobe, lips, finger etc. As this radiation propagates through the human tissue, it interacts with the various components of the tissue, including glucose. The spectrum of the light exiting the body is then collected and analyzed to estimate concentration of glucose within the sampled tissue volume. This approach is noninvasive in the sense that nothing is extracted from the body. Noninvasive measurements offer the promise of being painless, reagentless, fast, convenient, continuous, and yet they are biocompatible. In addition, information from multiple analytes is possible depending on the wavelengths used in the analysis.

Even after considerable efforts in these developments for more than four decades reliable and clinically acceptable measurement methods have not yet emerged. The major obstacle for accurate estimation of glucose by in -vivo optical technique is that small signature of blood glucose and often gets buried in the noise. Furthermore, certain optical characteristic features of biological tissues create significant interference in the estimation, for example: other absorbing variants in blood; multiple scatter in skin, muscle, and bone; and also the strong absorption bands of water[7],[8],[9].

Here we have developed a calibration model for the continuous spectral data in the range 400-5000cm-1 using PLSR regression technique. PLSR technique which most commonly used for analysis in econometrics and social sciences has been found to be excellent method for the determination of concentration of blood analytes such as glucose, ascorbate, lactate, cholesterol etc.

For applying this techniques to predict glucose concentration a set of calibration training data is formulated from the collected absorption spectra of five major constituents data which contains absorption bands associated

978-1-5090-2889-4/16/$31.00 ©2016 IEEE

with glucose along with an absorption bands of other constituents[10].

In the PLS method, both the absorption data and the concentration data are used at the outset to formulate a calibration model[11].

II. MULTIVARIATE MODEL GLUCOSE ESTIMATION

Twelve continuous spectra's in the range 400-5000cm-1 were collected with a Schimatzu FTIR8400S spectrophotometer equipped with an external 50-watt tungsten halogen lamp, Germanium-coated KBr plate beam splitter, and a Temperature controlled high sensitivity detector (DLATGS detector).

A. Multivariate model for Human Whole Blood

Following reagents were used to prepare the phantom of blood tissue : Glucose, Sodium lactate, Sodium Ascorbate, Alanine, Urea (from Loba Chemical Co., Inc.) .

Procedure: 12 calibration samples were prepared by carefully weighing the above mentioned reagents and mixing them in their normal proportion as shown in Table I with the KBr acting as base. The sample: KBr ratio maintained is 1:100 so as to make the thickness (path length) of the pellet same for all the 12 sample. Error propagation indicates a relative concentration uncertainty of nearly 0.5% for these mixtures, which Corresponds to a maximum uncertainty of 0.07 mM (i.e 2 mg/dl) for glucose. This level of uncertainty ultimately limits the analytical performance of all PLSR calibration models.

All these samples were scanned using Schimatzu FTIR 8400S in the range 400-5000cm-1 comprising total of 2387 points. The recorded Absorption spectra of all these 12 samples are shown in Fig.1 which are used for calibrating the PLSR model.

TABLE I. NORMAL RANGE OF VARIANTS IN HUMAN BLOOD.

Sample	Glucose mg/dl	Urea mg/dl	Alanine mg/dl	Ascorbate mg/dl	Lactate mg/dl
1	70	10	10	1	10
2	120	20	30	3	20
3	95	15	20	2	15
4	120	15	20	2	15
5	70	15	20	2	15
6	95	20	20	2	15
7	95	10	20	2	15
8	95	15	30	2	15
9	95	15	10	2	15
10	95	15	20	2	20
11	120	10	20	2	15
12	70	20	20	2	15

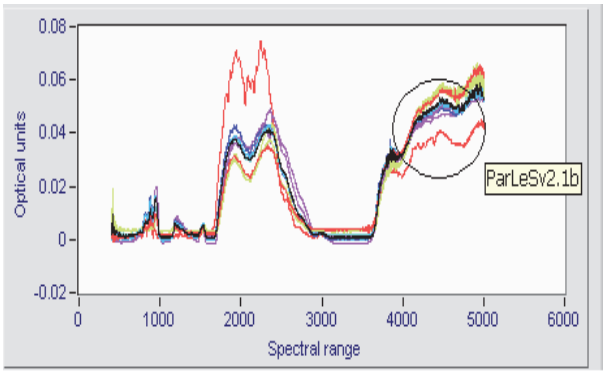

Fig.1. Absorption spectra for all 12 samples

The whole blood spectrum model for various concentrations of the five variants has been generated. The model developed as above is based on concentration for the human blood chemo metrics system well within the pathological range for 5 variants.

B. Lorentzian model for Blood Spectra

The Lorentz model for spectra generation is so flexible that just by varying the strength, line width, and natural frequency any practical spectrum can be generated with highly nonlinear behavior.

Just for the testing purpose we have developed a model based on Lorentz oscillator for 5 variants individually in the range 4000-5000cm-1 as shown in Fig.2.

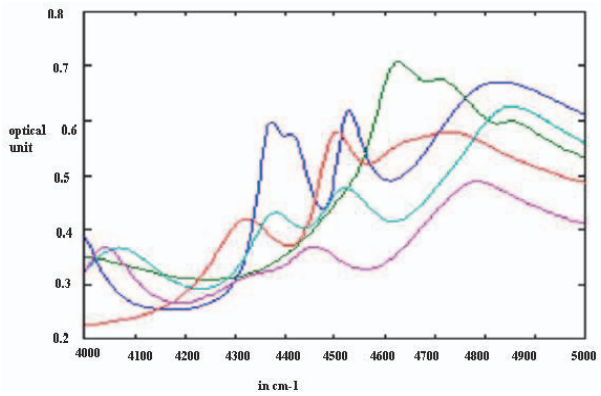

Fig.2. Signature of five major components simulated using Lorentz oscillator

We have also added the individually responses of blood variants during the simulation and the resultant spectra is shown in Fig.3. The resultant spectra generated using above technique resembles the actual spectra recorded using FTIR 8400 in 4000-5000 cm-1 range as shown inside the black color circle in Fig.1. The prediction result of glucose in human blood for the simulated spectra using Lorentz oscillator are validated by RMSE and Clarke error grid analysis[12].

We have considered the continuous spectral range from 400-5000cm- 1 so that we have more calibration points (2387 points).

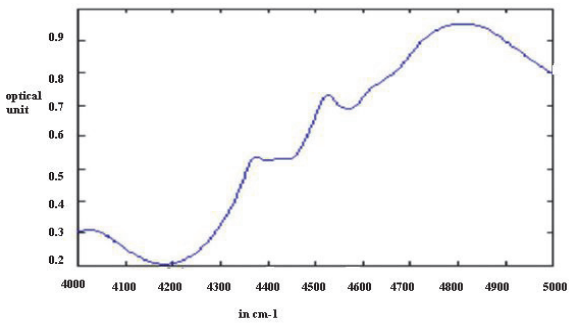

Fig.3: Resultant normalized signature of various components simulated using Lorentz oscillator.

III.PARTIAL REGRESSION ANALYSIS FOR CHEMOMETRIC SYSTEM

PLSR is an extension of the multiple linear regression models. In its simplest form, a linear model specifies the relationship between a dependent variable "*Y*" and a set of predictor variables, the "*X*".

PLSR has been used in various disciplines such as chemistry, economics, medicine, psychology, pharmaceutical, and medical science where predictive linear modeling, especially with a large number of predictors is necessary. PLSR has become a standard tool for modeling linear relations between multivariate measurements in chemometrics[13].
There are basically 2 PLSR algorithms i.e SIMPLS and NIPALS .we have developed the PLSR algorithm based on SIMPLS.

A. PLSR model Validation

Calibrating the non-invasive Glucometer is the biggest challenge faced by researcher due to weak detected signal in the transmission (and reflectance) mode due to the interferences from other blood constituent, varied skin complexion over the globe, temperature compensation, physiology dependent calibration, dependence of wavelength and number of PCA factors.

Here we have passed 3 test spectra having unknown glucose concentration is passed through PLSR model for prediction .Out of these 3 spectra, one spectra was taken from the calibration set of PLSR model and others two spectra had glucose concentration outside the calibration set. The predicted results for all 3 test samples for individual variate is shown in Table II.

The Table II also shows impact of number of PCA factors on model performance. As is typical of PLSR calibration models, the standard error decreases to a minimum as more PCA factors are used in the analysis. For these data, the minimum is reached at 5 factors while, the optimum number of factors is 10 since we have five variate.

TABLE II.PREDICTED *RESULT*

Unknown	Variants concentr-ation.	PCA components	Glucose	Urea	Alanine	Ascorbate	lactate
Test sample1	**Actual**		**70**	**15**	**20**	**2**	**15**
	Predicted	**5**	77.245	15.358	19.081	2.172	16.256
		10	70.164	15.00	19.982	1.998	14.999
		12	70.00	15	20.00	2.00	15.00
		13	70.00	15	20.00	2.00	15.00
		15	70.00	15.0	20	2.00	15.00
Test sample2	**Actual**		125	18	9	2	8
	Predicted	**5**	99.292	11.791	24.367	1.620	14.619
		10	121.075	11.821	20.663	1.617	13.481
		12	121.031	11.792	20.633	1.618	13.454
		13	121.035	11.785	20.684	1.619	13.459
		15	121.039	11.83	20.343	1.618	13.355
Test sample3	**Actual**		**104**	**14**	**24**	**2**	**17**
	Predicted	**5**	94.26	14.376	21.864	1.767	14.616
		10	106.879	14.617	20.146	1.805	15.383
		12	106.808	14.583	20.136	1.803	15.347
		13	106.500	14.581	20.096	1.806	15.349
		15	106.275	14.646	20.116	1.810	15.408

Most of the models of non-invasive glucose instrumentation have difficulty in satisfying the Clarke Error Grid (CEGA) [14] shown in Fig.4. This is due to poor multivariate model. The fine tuning of the same multivariate model is required by incorporating the various parameters influencing glucose[15],[16].

- Region A are those values within 20 % of the reference.
- Region B contains points that are outside of 20 % and are not so accurate but would not lead to inappropriate treatment,
- Region C are those points leading to chances of wrong treatment,
- Region D are those points indicating a potentially dangerous failure to detect hypoglycemia or hyperglycemia, and
- Region E are those points that would confuse treatment of hypoglycemia for hyperglycemia and vice-versa.

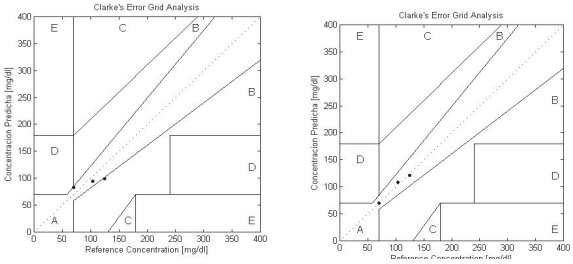

Fig.4. CEGA Plot for glucose with a)5 PCA b)10 PCA

We have developed Matlab code to generate the grid analysis plot to test the acceptability of PLSR model for glucose. From CEGA analysis for glucose prediction with PLSR model with 5 and 10 PCA components and we have found that the prediction results show 100% of predicted values lie around the reference line in 'A' region for 10 PCA components (Fig. 4(b). For 5 PCA prediction, the points lie

on the border line of region A and B but these results are still acceptable and doesn't lead to wrong treatment.

The PCA score plots and glucose prediction plot for all 3 test samples with 10 PCA components are shown in Fig.5 and Fig.6 respectively.

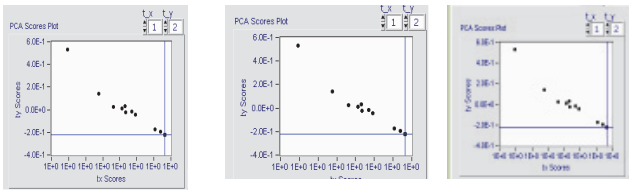

Fig.5. PCA scores Plot

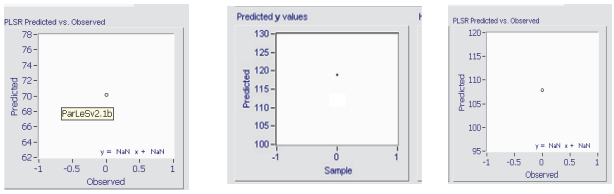

Fig..6: Glucose prediction plot for all 3 unknown

IV .RESULTS ,ANALYSIS AND DISCUSSION

In all the 3 test samples the prediction error is more for 5 PCA. In case of glucose prediction in test sample 2, error is 20% which is high for 5 PCA components while it is 3.14% for 10 PCA components and 3.16% for 15 PCA. For better estimation, the number of PCA factors should be twice the number of Variants (i.e. PCA should be at least 10 in present case).Prediction Error does not improve further by increasing the PCA component beyond 10.

Test sample 2 having a variate concentration above (or below) the normal ranges selected to build the calibration model shows high prediction error except glucose which is 3.14%. A magnitude of glucose outside the normal range i.e. in test sample2 gives higher error for 5 PCA which is 20 % as compared to 10.35% for test sample 1 and 9.36 % for test sample3. The analysis for same is true for Alanine & Lactate shows similar behavior as observed in glucose. Variate concentration prediction for test sample1 taken from calibration set shows very accurate prediction results with 10 PCA.

V.CONCLUSION

The multivariate analysis performed on 5 blood constituents based on a multivariate matrix developed with 12 different concentration reveals that 10 PCA components gives the best estimation of constituents. Further increase in PCA does not improve the prediction error much and also requires lot of computation resources. The technique so developed has best estimation of 3.14 % for glucose even if the concentration is outside the range. A similar behavior is also observed for other constituents like Alanine, Urea, Lactate and Ascorbate.

ACKNOWLEDGMENT

Financial support by Indian Council of Medical Research (ICMR, New Delhi), University Grant Commission (UGC, New Delhi) and ALTERA Inc. is acknowledged.

REFERENCES

[1] Lawton J, Peel E, Douglas M, Parry O. `Urine testing is a waste of time': newly diagnosed Type 2 diabetes patients' perceptions of self-monitoring. Diabetes Med 2004; 21(9): 1045-8.

[2] Tierney MJ, Tamada JA, Potts RO, et al. The GlucoWatch biographer: a frequent automatic and noninvasive glucose monitor. Ann Med 2000; 32(9): 632-41.

[3] Eastman RC, Chase HP, Buckingham B. Use of the GlucoWatch biographer in children and adolescents with diabetes. Pediatric Diabetes 2002; 3(3): 127-34.

[4] Arnold MA, Small GW. Noninvasive glucose sensing. Anal Chem. 2005 Sep 1;77(17):5429-39.

[5] Khalil OS. Non-invasive glucose measurements at the dawn of the new millennium: an update. Diabetes Technol Ther. 2004 Oct;6(5):660-97.

[6] Khalil OS. Spectroscopic and clinical aspects of noninvasive glucose measurements. Clin Chem. 1999 Feb;45(2):165-77.

[7] Heise HM. In: Siesler HW, Ozaki Y, Kawata S, Heise HM, Eds. Near-Infrared Spectroscopy. Wiley-VCH : Weinheim 2002.

[8] Cote GL, Fox MD, Northrop RB. Noninvasive optical polarimetric glucose sensing using a true phase measurement technique. IEEE Trans Biomed Eng 1992; 39(7): 752-6.

[9] Guclu B, Engbretson GA, Bolanowski SJ. Constrained optimization of Drude's equations eliminates effects of confounding molecules for the polarimetric measurement of glucose. J Biomed Opt 2004; 9(5): 967-77.

[10] F. M. Hamy, G. M. Cohenz, I. Kostanicy, and B. R. Goochx, *Physiological Measurement* _IOP, London, 1995_, Vol. 37, pp. 1–20.

[11] M. Cope, P. van der Zee, M. Essenpreis, S. R. Arridge, and D. T. Delpy, Proc. SPIE **1431**, 251 _1991.

[12] J. S. Parab, R. S. Gad, and G. M. Naik, J. Appl. Phys. **107**, 104701 ,2010.

[13] P. Geladi and B. R. Kowalski, Anal. Chim. Acta **185**,1 ,1986.

[14] Chung, H., Arnold, M. A. , Rhil, M. and Mruhammer, D. W. , Simultanious measurement of glucose , glutamine, ammonia, lactate , and glutamate in aqueous solutions by near-infrared spectroscopy, Appl. Spectroscopy, 50, 270, 1996.

[15] A. Maran "Continuous Subcutaneous Glucose Monitoring in Diabetic Patients" Diabetes Care, 2002,vol.25(2).

[16] B.P. Kovatchev "Evaluating the Accuracy of Continuous Glucose-Monitoring Sensors "Diabetes Journal ,2004.