31st Annual International Conference of the IEEE EMBS
Minneapolis, Minnesota, USA, September 2-6, 2009

Multivariate Regression and Discreminant Calibration Models for a Novel Optical Non-Invasive Blood Glucose Measurement Method Named Pulse Glucometry

Yasuhiro Yamakoshi, Mitsuhiro Ogawa, Takehiro Yamakoshi, Toshiyo Tamura and Ken-ichi Yamakoshi

Abstract—**A novel optical non-invasive *in vivo* blood glucose concentration (BGL) measurement technique, named "Pulse Glucometry", was combined with a kernel method; support vector machines. The total transmitted radiation intensity (I^{λ}) and the cardiac-related pulsatile changes superimposed on I^{λ} in human adult fingertips were measured over the wavelength range from 900 to 1700 nm using a very fast spectrophotometer, obtaining a differential optical density (ΔOD^{λ}) related to the blood component in the finger tissues. Subsequently, a calibration model using paired data of a family of ΔOD^{λ}s and the corresponding known BGLs was constructed with support vector machines (SVMs) regression instead of using calibration by a conventional primary component regression (PCR) and partial least squares regression (PLS). Secondly, SVM method was applied to make a nonlinear discriminant calibration model for "Pulse glucometry." Our results show that the regression calibration model based on the support vector machines can provide a good regression for the 101 paired data, in which the BGLs ranged from 89.0-219 mg/dl (4.94-12.2 mmol/l). The resultant regression was evaluated by the Clarke error grid analysis and all data points fell within the clinically acceptable regions (region A: 93%, region B: 7%). The discriminant calibration model using SVMs also provided a good result for classification (accuracy rate 84% in the best case).**

I. Introduction

Measurement of blood glucose concentration (BGL) has long been considered as important for screening in diabetes, diabetes management, pre-diabetes management and so on. For diabetes management in particular, frequent measurement of BGL is necessary [1], thus, many portable BGL instruments have appeared in the market. However, current instruments suffer from several problems. Almost all BGL monitors are based on withdrawal of blood samples with small needles or lancets. The user must puncture their skin and squeeze the surrounding tissue to draw blood out. Because frequent monitoring is essential, the repeat procedure of skin puncturing becomes painful and troublesome and, furthermore, can cause an infection. Although a non-puncturing type of BGL monitor, the GlucoWatch Biographer, (using iontophoresis to draw glucose molecules via skin), has been approved by the FDA, its measurement procedure can still cause skin irritation after repeated application [2]. Therefore, overall, it has to be said that there is still an important unmet need for a truly non-invasive technique that will allow frequent, convenient and safe BGL monitoring. As the search for such a technique continues it goes without saying that the gold standard for BGL measurement will be based on analysis of a venous blood sample obtained by invasive methods.

In order to obtain physiological variables non-invasively, *in vivo* optical methods using specific parts of the electromagnetic spectrum have been studied and applied up to now and further methods are still being proposed [3]–[5]. We have focused attention on *in vivo* spectrophotometric measurement in living tissues, with analysis to obtain parameters related to blood including blood glucose. Among our developments, we have recently reported a novel art named "Pulse Glucometry", that is based on very high speed near infrared spectroscopy for BGL monitoring without any invasion [6]–[8].

In general, *in vivo* and *in vitro* spectroscopic analysis, including "Pulse glucometry", have utilized multivariate calibration models that are constructed by simple multiple linear regression (MLR) or multiple regression based scheme, such as Partial Least Squares Regression (PLS) and Principal Component Regression (PCR) [9]. MLR, PLS and PCR are generally used for pure linear calibration model or linearly-transformable (nonlinear) calibration. Recently, through developments in the field of multivariate statistical analysis, a kernel-based method has come up in the last decade with the emergence of the Support Vector Machines (SVMs) including the kernel trick [10], [11]. The kernel trick is a method for converting a linear classifier algorithm into a nonlinear one. The SVMs method is currently regarded as one of the strongest methods of supervised learning applied to classification and regression.

In this paper, we describe an attempt to apply three methods for calibration; PCR, PLS and SVMs, to "Pulse glucometry" to obtain a multivariate calibration regression models. Secondly, nonlinear discriminant analysis using

Manuscript received April 22, 2009.

Yasuhiro Yamakoshi is with the yu.sys Corporation, 1-143 Morino-Sato, Kanazawa, Ishikawa, 920-1167, JAPAN (phone: +81-76-234-5771; fax: +81-76-234-5772; e-mail: y@yusys.jp).

Mitsuhiro Ogawa is with the yu.sys Corporation, Ishikawa, JAPAN (e-mail: m_ogawa@y5.dion.ne.jp)

Ken-ichi Yamakoshi, is with The Graduate School of Natural Science & Technology, Kanazawa University, Kanazawa, JAPAN (e-mail: yamakosi@t.kanazawa-u.ac.jp).

Takehiro Yamakoshi, is with The Graduate School of Natural Science & Technology, Kanazawa University, Kanazawa, JAPAN (e-mail: takehiro@kenroku.kanazawa-u.ac.jp).

Toshiyo Tamura is with the Graduate School of Science & Technology., Chiba University, Chiba, JAPAN (e-mail: tamurat@faculty.chiba-u.jp)

SVMs are attempted to "Pulse glucometry."

II. Methods

A. Pulse Glucometry

The previously developed and reported method of "Pulse Glucometry" is based on the application of very fast spectrophotometric analysis in a body tissue segment. Photoplethysmograms exhibiting cardiac-related blood volume pulses are collected for a number of narrow-bands of radiation over a broad spectrum. In our experimental setup, the measurement system consisted of, a light source (halogen lamp: maximum power 150 W), an optical fiber of 10-mm diameter for the incident radiation and a single fiber of 1.2-mm diameter for collecting the transmitted radiation, a spectrometer (polychromator, M25-TP; Bunkoh-Keiki Co. Ltd., Japan), a linear, liquid nitrogen cooled (−50 to −100 °C), InGaAs photodiode-array (multi-photodetector, OMA V: 512-1.7(LN); Princeton Instruments Co., USA), and a conventional personal computer with an appropriate interface. Using this system, optical transmittance spectra in the wavelength range 900 to 1700 nm can be measured with a resolution of 8 nm and 16-bit digitization. The maximum spectrum sampling speed achievable is 125 spectra per second with this instrument, and in this experiment described here we adopted a speed of 100 spectra per second.

In this study, transmittance spectra derived from a fingertip of an index finger were collected from 10 healthy adult volunteers (22 to 59 years old; 8 males and 2 females). Informed consent was obtained from each subject prior to the experiment. Oral glucose tolerance tests (OGTT) were carried out in these subjects in order to create varying BGLs. Immediately after obtaining each Transmittance spectrum, blood samples (about 3 ml) were collected from the cephalic vein of the forearm and analyzed chemically to obtain the actual BGL.

From the time series of transmittance spectra obtained by this procedure optical density change (differential optical density) at wavelength λ (ΔOD^{λ}) can be derived, as:.

$$\Delta OD^{\lambda} = \frac{I^{\lambda}(t_1)}{I^{\lambda}(t_2)} \quad (1)$$

where $I^{\lambda}(t)$ is measured radiation intensity at wavelength λ, time t.

In this experiment, to determine timing point during the cardiac cycle use was made of the pulsatile component superimposed on the transmitted radiation intensity [6], Thus, time t_1 is arranged to correspond with the diastolic phase and t_2 should correspond with the systolic phase. Then, differential spectra over the wavelength 900 to 1700 nm were obtained in each measurement.

B. Multivariate calibration regression models and discriminant models

Three regression models; Primary Component Regression (PCR), Partial Least Squares Regression (PLS), and Support Vector Machines Regression (SVMsR) were attempted to create multivariate calibration models to relate differential optical density spectra to measured BGL employed as the teaching data. For learning of ANN, quasi-Newton method is applied with weight decay. To implement the procedure, the

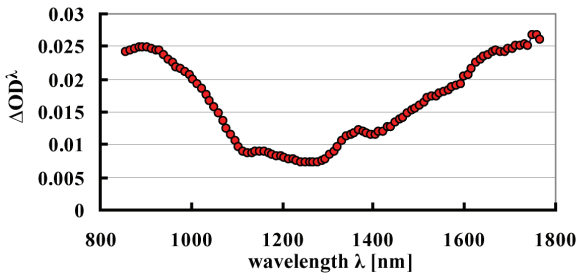

Fig. 1. An example of the differential optical density spectrum that was obtained from a subject with BGL level 116.5 [mg/dl]. The optical density spectrum was used as a part of data set for constructing multi vitiate calibration models.

software "R" version 2.6.0 and the kernlab module version 08-2 for "R" were used on a conventional personal computer (Dimension 9100 with a CPU Pentium D 830 and 2GB memory, DELL Inc.) [12], [13]. Before calculations to obtain a calibration model were performed, the differential optical spectra were filtered and normalized, in a manner already reported [6] by using the software Matlab™ version 7.x. Then spectra with artifact were separated by observation, and rejected from the data set. In order to obtain the resultant calibration model, parameters on calculations were searched repeatedly. For finding the resultant parameters, any sophisticated search algorithm was not applied.

Afterwards, a nonlinear discriminant models using SVM were attempted for the same data sets. Firstly, the measured BGL levels were classified based on the criterion of diabetes screening of fasting blood sugar (case of FBS), and data were classified to "Normal" class, "Impaired Fasting Glycaemia" class and "Diabetes Mellitus" class. Then, a nonlinear discriminant analyses were applied to make discriminant calibration models. Secondly, those were also assumed as 2-hour postprandial blood sugar (case of 2 hour-PC) and data were classified to "Normal", "Impaired Glucose Tolerance" and "Diabetes Mellitus" classes. Then, another nonlinear discriminant analyses were applied.

III. Result and Discussions

101 sets of data for the differential optical density spectra and the measured BGLs over the range of 100.7-246.3 mg/dl (5.59-13.7 mmol/l) were obtained and used as the data set to create a calibration model. Fig 1 shows an example of obtained differential optical density spectra. The calibration models were evaluated by 5-fold cross-validation. Finally, resultant parameters for each methods were obtained as follows. For PCR, the 1st to 20th principal components were applied. For PLS, the number of latent variable is 15. For

SVMsR, the ANOVA RBF (Radial Basis Function) kernel with degree one was used in training and ε in Vapnik's insensitive-loss function is 0.123. The resulting estimated BGLs versus measured BGLs are plotted on a Clarke error grid shown in Fig.2. As an be seen in Fig.2, almost all data points are within clinically acceptable regions: the region A and B in each calibration [14]. Among them, SVMsR calibration provided the best plot distributions. Therefore, it might reasonably be suggested that SVMsR can be used for constructing multivariate calibration models as part of the procedure of implementing "Pulse Glucometry."

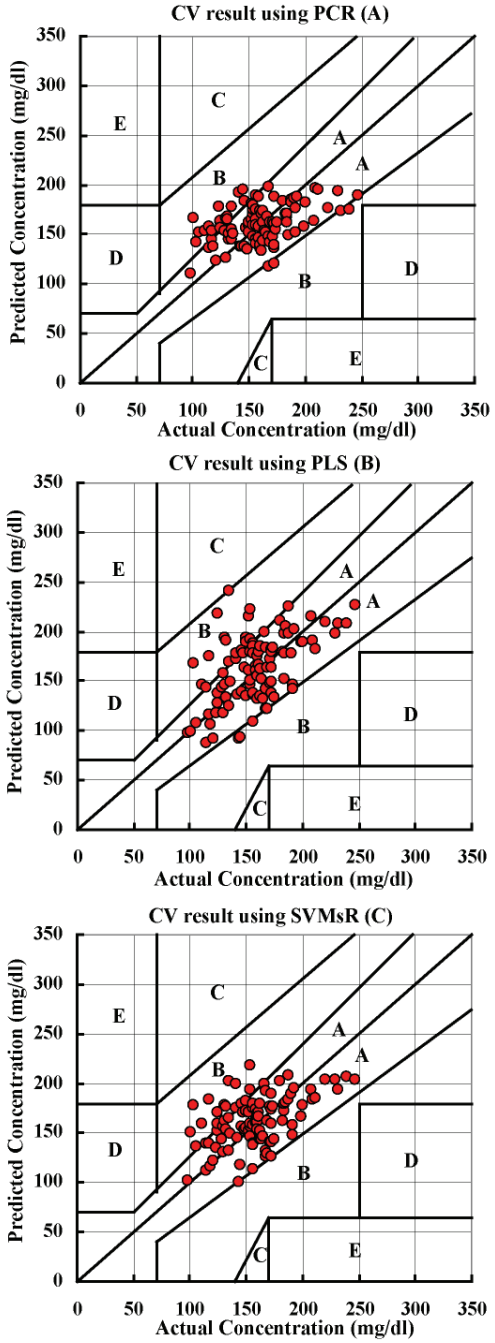

Fig. 2. Result of cross validations plotted on Clarke Error Grid; subplot (A) shows result by using PCR,(B) shows by PLS, and (C) shows by SVMsR

Secondly, nonlinear discriminant models were attempted in same data sets. The calibration models were also evaluated by 5-fold cross-validation. Table 1 shows the result provided by SVMs classification. The results in the case of FBS were considered as good classifications.

By the observation of Fig.2, it can be said that PCR provided the worst regression. However, despite the result, the superiority of regression by SVM over the conventional method PLS is not still clear. However, PLS assumes a linear model. If nonlinear elements are present in a system then PLS based calibrations cannot avoid errors originating from nonlinearity, at least in principle. Meanwhile, in theory, SVMsR can be considered as solvers of the problem with nonlinearity. Thissen et al. attempted to compare the performance of SVMs with conventional PLS for spectral regression applications in the chemometrics field and reported superiority of SVMs over PLS [15].

It might be difficult to expect that non-invasive blood glucose measurement could achieve accuracy comparable to invasive one. From a practical application standpoint, discriminant type calibration (providing qualitative nature) may be applicable instead of regression type calibration (providing quantitative blood glucose level). In this paper, we tried to apply nonlinear discriminant analyses to provide qualitative information about blood glucose level and provide fairly good result. From the results shown on this paper, it is difficult to discuss it result because of not-uniform distribution of BGL of dataset. Additional data will be required to establish the discriminant type calibration.

Furthermore, the system will eventually need to be miniaturized into a portable device like conventional pulse

TABLE I
RESULT OF DISCRIMINANT CALIBRATION USING ANN

(a) In the case of fasting blood glucose (FBG)

		Measured BGL value (mg/dl)		
		～110(NORMAL)	110～126 (Impaired Fasting Glycaemia)	126～(Diabetes Mellitus)
Predicted BGL (mg/dl)	～110(NORMAL)	2	0	0
	110～126 (Impaired Fasting Glycaemia)	0	3	5
	126～(Diabetes Mellitus)	2	9	80

accuracy rate: 0.84 (85/101)
n=101

(b) In the case of 2-hour postprandial blood glucose (PBG)

		Measured BGL value (mg/dl)		
		～140(NORMAL)	110～126 (Impaired Fasting Glycaemia)	200～(Diabetes Mellitus)
Predicted BGL (mg/dl)	～140(NORMAL)	13	12	1
	110～126 (Impaired Fasting Glycaemia)	10	53	3
	200～(Diabetes Mellitus)	0	5	4

accuracy rate: 0.69 (70/101)
n=101

oximeter. Then only a several LEDs should be applied in the device. To find the most effective wavelengths of LED emitting, more data analyses are required. Stepwise model selection by AIC could be used for the purpose.

IV. Conclusion

A recently proposed non-invasive *in vivo* BGL optical measurement technique named "Pulse Glucometry" was combined with three types of regression analyses (PCR, PLS, and SVMsR) and a discriminant analyses using SVMs to construct multivariate calibration models. Good regression and classification were obtained using SVMs methods. These data provide preliminary evidence that Pulse Glucometry with SVMsR can be applied effectively to measure BGLs non-invasively and a discriminant type calibration may be achieved by using SVMs.

Acknowledgment

The authors pay great respect to the R Project for Statistical Computing and its module kernlab.

References

[1] J. Lawton, E. Peel, M. Douglas and O, Parry, " `Urine testing is a waste of time': newly diagnosed Type 2 diabetes patients' perceptions of self-monitoring," *Diabetic Medicine,* vol. 21, no. 9, pp. 1045-8 Sep. 2004 .

[2] M. J. Tierney, J. A. Tamada, R. O. Potts, R. C. Eastman, K. Pitzer, N. R. Ackerman and S. J. Fermi, "The GlucoWatch biographer: a frequent automatic and noninvasive glucose monitor," *Ann. Med*, vol. 32, no. 9, pp. 632-41, Dec. 2000.

[3] R. Marbach, Th. Koschinsky, F. A. Gries, and H. M. Heise, "Noninvasive Blood Glucose Assay by Near-Infrared Diffuse Reflectance Spectroscopy of the Human Inner Lip," *Applied Spectroscopy*, vol. 47, no. 7, pp. 875-81, 1993.

[4] H. Zeller, P. Novak and Landgraf, "Blood glucose measurement by infrared spectroscopy*," Int J Artif Organs*, vol. 12, no. 2, pp. 129-35, Feb. 1989.

[5] A. J. Berger, I. Itzkan and M. S. Feld, "Feasibility of measuring blood glucose concentration by near-infrared Raman spectroscopy," *Spectrochim Acta A Mol Biomol Spectrosc.,* vol. 53A, no. 2, pp. 287-92, Feb. 1997.

[6] K. Yakmakoshi and Y. Yamakoshi, "Pulse glucometry: a new approach for noninvasive blood glucose measurement using instantaneous differential near-infrared spectrophotometry," *Journal of Biomedical Optics,* vol. 11, no. 5, 054028, Sep.-Oct. 2006.

[7] K. Yamakoshi, "A Non-invasive Blood constituent Measuring Instrument and Measuring Method," Intern. Patent No. PCT/JP03/03587.

[8] K. Yamakoshi, Y. Yamakoshi, M. Nogawa, T. Yamakoshi and S. Tanaka, "A New Approach for Noninvasive Blood Glucose Measurement Using Differential Near Infrared Spectrophotometry," *in Proc. of World Congress on Medical Physics and Biomedical Engineering 2006,* Seoul, 2006, 2961.

[9] K. J. Ward, D. M. Haaland, M. R. Robinson and R. P. Eaton, "Post-Prandial Blood Glucose Determination by Quantitative Mid-Infrared Spectroscopy," *Applied Spectroscopy*, vol. 46, no. 6, pp. 959-65, 1992

[10] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and other kernel-based learning methods.* Cambridge: Cambridge University Press, 2000.

[11] T. B. Trafalis and R. C. Gilbert, *Robust classification and regression using support vector machines*. Amsterdam: Elsevier B.V., 2006

[12] http://www.r-project.org/

[13] http://cran.r-project.org/src/contrib/Descriptions/kernlab.html

[14] W. L. Clarke, D. Cox, L. A. Gonder-Frederick, W. Carter, "Evaluating clinical accuracy of systems for self-monitoring of blood glucose," *Diabetes Care.*, vol. 10, no. 5, pp. 622-8, Sep.-Oct. 1987.

[15] U. Thissen, M. Pepers, B. Ustun, W. J. Melssen and L. M. C. Buydens, "Comparing support vector machines to PLS for spectral regression applications," *Chemometrics and Intelligent Laboratory Systems*, vol. 73, pp. 169–79, 2004.