Support vector machines as multivariate calibration model for prediction of blood glucose concentration using a new non-invasive optical method named pulse glucometry

Support Vector Machines as Multivariate Calibration Model for Prediction of Blood Glucose Concentration Using a New Non-invasive Optical Method Named Pulse Glucometry

Mitsuhiro Ogawa, Yasuhiro Yamakoshi, Makoto Satoh, Masamichi Nogawa, Takehiro Yamakoshi, Shinobu Tanaka, Peter Rolfe, 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^λs and the corresponding known BGLs was constructed with support vector machines regression instead of using calibration by a conventional partial least squares regression (PLS). Our results show that the calibration model based on the support vector machines can provide a good regression for the 183 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%).

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* 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]. However, 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 non-linear 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 the SVM method to "Pulse glucometry" to obtain a multivariate calibration model. This may then be compared with our prior use of the PLS method reported previously [6]–[8]

II. METHODS

A. Pulse Glucometry

The 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 27 healthy adult volunteers (20 to 43 years old; 24 males and 3 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:.

$$\{ \text{EMBED Equation.3} \} \tag{1}$$

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. Construction of calibration model using support vector machines

Regression by support vector machines was attempted to create a multivariate calibration model to relate differential optical density spectra to measured BGL employed as the teaching data. To implement the procedure, the software "R" version 2.3.1 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 normalized, in a manner already reported [6]. 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.

III. RESULT AND DISCUSSIONS

183 sets of data for the differential optical density spectra and the measured BGLs over the range of 89.0-219 mg/dl (4.94-12.2 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 10-fold cross-validation. Finally, resultant parameters were obtained as follows: the kernel function applied in training and predicting is ANOVA RBF (Radial Basis Function) kernel and ε in the insensitive-loss function is 0.01. The resulting estimated BGLs versus measured BGLs are plotted on a Clarke error grid shown in Fig.1. As an be seen in Fig.2, all data points are within clinically acceptable regions: the region A and B (region A: 93%, region B: 7%: A+B=100%) [14]. Therefore, it might reasonably be suggested that SVMR can be used for constructing multivariate calibration models as part of the procedure of implementing "Pulse Glucometry."

Despite this result, the superiority of regression by SVM over conventional methods (MLR, PLS and PCR) is still unclear. A significant problem in MLR is multi-colinearity, but PLS and PCR are generally felt to be methods to solve problems arising from multi-colinearity. Therefore, these methods (especially PLS) are widely employed in chemometrics that can have a data set with a large dimension.
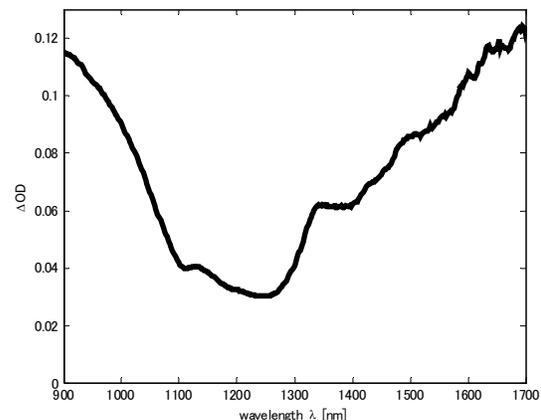

Fig. 1. An example of the differential optical density spectrum that was obtained from a subject with BGL level 116.5 [mg/dl]. The optical density spectrum was used as a part of data set for constructing multi vitiate calibration models.
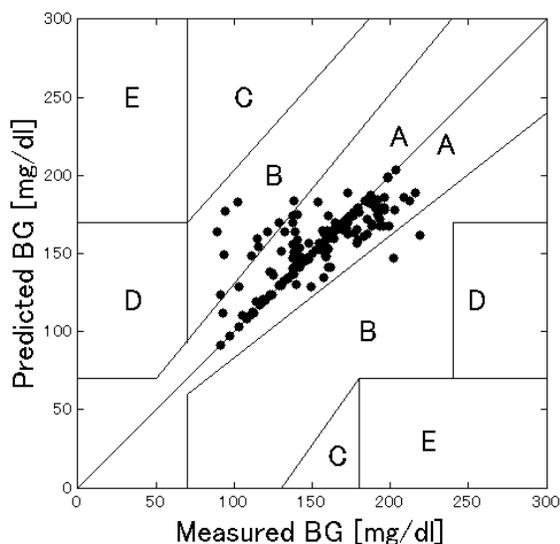
Fig. 2. A scatter diagram showing comparison of the predicted and measured BGL values by Clarke error-grid analysis. Only values in regions A and B are clinically acceptable. As a result, region A: 93%, region B: 7%: (A+B=100%), region C, D and E: 0%. Regression formula is Predicted BG =0.56*Measured BG + 69 [mg/dl] (r=0.76).

However, these methods assume a linear model. If nonlinear elements are present in a system then even PLS and PCR based calibrations cannot avoid errors originating from nonlinearity, at least in principle. Meanwhile, in theory, the kernel trick included in the SVM method provides a linear classification in the new space equivalent to non-linear classification in the original space. Such a mapping can solve problems of multi-colinearity and also non-linear problems. 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 SVM over PLS [15]. Therefore, it might be expected that the regression performance of SVM would be superior to conventional methods even in the implementation of "Pulse glucometry".

IV. CONCLUSION

A recently proposed non-invasive *in vivo* BGL optical measurement technique named "Pulse Glucometry" was combined with support vector machines regression to construct a multivariate calibration model instead of a conventional PLS-based calibration. Using this SVMR approach with 183 measured differential optical density spectra from fingertips of adult humans whose blood glucose levels were in the range of 89.0-219 mg/dl (4.94-12.2 mmol/l) yielded good results. When subjected to evaluation by a Clarke error grid analysis all estimated BGL data points were within clinically acceptable regions (region A: 93%, region B: 7%). The result obtained suggests that support vector machines regression can provide a suitable calibration model for "Pulse Glucometry."

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