Multivariate Regression and Classification Models for Estimation of Blood Glucose Levels Using a New Non-invasive Optical Measurement Technique Named "Pulse-Glucometry"

Yasuhiro Yamakoshi¹, Mitsuhiro Ogawa*,²,³ and Toshiyo Tamura¹

¹Graduate School of Engineering, Chiba University, Chiba, Japan
²Graduate School of Natural Science and Technology, Kanazawa University, Kanazawa, Japan
³Yu.sys Corporation, Kanazawa, Japan

Abstract: An optical method recently proposed for non-invasive in vivo blood glucose concentration (BGL) measurement, named "pulse glucometry", was implemented with three multivariate regression methods, Principal Component Regression (PCR), Partial Least Squares Regression (PLS) and Support Vector Machines Regression (SVMsR), as well as with a classification method, Support Vector Machines Classification (SVMsC), for carrying out calibration. A very fast spectrophotometer provided instantaneously and simultaneously the total transmitted radiation spectrum (I') and the cardiac-related pulsatile component (ΔI') superimposed on I' in human fingertips over a wavelength range from 900 to 1700 nm with resolution of 8 nm in 100 Hz sampling. From a family of I's including information relating to various BGL values, differential optical densities (ΔOD's (−log(1+ΔI'/I'))) were calculated and normalized by the ΔOD values at 1100 nm. Finally, the 2nd derivatives of the normalized ΔOD's served as regressors. Subsequently, calibration models from regressors and regressands (the corresponding measured BGL or classified BGL values) were constructed with PCR, PLS, SVMsR and SVMsC. Each regression model showed a relatively good result by evaluating a 5-fold cross validation using total 100 data-sets: Clarke error grid analysis indicated a good correlation in each model compared with the measured BGL values, and the SVMsR calibration provided the best plot distributions. Good regression models were obtained by these three methods. This study suggests that "pulse glucometry" can produce clinically acceptable results when implemented with regression and classification type calibrations, and, through rapid BGL assessment, promises to offer a more practical, easier and more convenient way for diabetes screening and health care in normal subjects than is possible with existing methods.

INTRODUCTION

Measurement of blood glucose concentration (BGL) has long been considered as an essential need for self-care in diabetes and for screening in pre-diabetes, as well as for health care in normal subjects. For diabetes management in particular, frequent measurements of BGL is inevitable [1], and thus many kinds of portable devices for self-monitoring of blood glucose (SMBG) have been commercialized worldwide. However, many of the current SMBG devices are based on the user puncturing their skin with a small needle or lancet and squeezing the surrounding tissue to withdraw a blood sample, and this is a severe limitation; frequent monitoring is necessary and the repeated procedure of skin puncturing becomes painful and troublesome and, furthermore, can cause an infection.

Although a non-puncturing type of BGL device, the 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].

Truly non-invasive BGL measurement has therefore been most desired and has been explored worldwide particularly using optical methods such as surface plasmon resonance [4], photo-acoustic technique [5], optical coherence tomography [6], diffuse reflectance spectrometry [7], and so on [8, 9]. However, despite considerable efforts in these developments for more than four decades reliable and clinically acceptable measurement methods have not yet emerged. The major obstacle to success with in vivo optical BGL measurement is the very small in vivo optical signature of glucose. Furthermore, certain optical characteristic features of biological tissues create significant interference, for example: other absorbing species; multiple scatter in skin, muscle, and bone; and the strong absorption bands of water [10-12]. Consequently, a method to observe only blood has long been sought.

We have recently focused attention on in vivo spectrophotometric measurement in living tissues, with analysis to obtain parameters related to blood including blood glucose, and have subsequently proposed a novel technique named "pulse glucometry". This is a non-invasive approach for BGL monitoring based on very high speed near infrared spectroscopy, having a sampling rate significantly faster than the cardiac cycle [13-15].

In the main, methods of in vivo and in vitro optical spectroscopic analysis, including "pulse glucometry", have
be implementing with standard chemometric procedures. These have utilized multivariate calibration models constructed either by simple multiple linear regression (MLR) or multiple regression based schemes, such as partial least squares regression (PLS) and principal component regression (PCR) [16]. MLR, PLS and PCR are generally used for pure linear calibration models or linearly-transformable (nonlinear) calibration. Just recently, through developments in the field of multivariate statistical analysis, a kernel-based method has appeared with the emergence of the support vector machines (SVMs) including the kernel trick [17, 18]. This kernel trick is one of several useful methods for converting a linear classifier algorithm into a nonlinear one. The SVMs method is also currently regarded as one of the strongest methods of supervised learning applied to classification and regression.

This paper deals, firstly, with an attempt to apply the nonlinear method, SVMs, to "pulse glucometry" to obtain a multivariate calibration regression model. This is then compared with our prior use of the PLS method reported previously [13, 15] and the conventional PCR. Secondly, the classification of BGL levels using SVMs is attempted by application to "pulse glucometry."

MATERIALS AND METHODS

A. Basic Principle of Pulse Glucometry

Details of the previously proposed method of "pulse glucometry" are reported elsewhere [13, 15]. Briefly, we assume a simplified optical model of biological tissue, comprising of three compartments, as shown in Fig. (1): arterial; venous; bloodless tissue. The essence of this new method is to utilize a blood volume change in a tissue segment under optical interrogation, such as a finger, and then, with a subtraction process, to remove the influences of basal interfering elements such as bone, skin, muscle, water and non-pulsatile blood [13]. We perform very rapid spectrophotometry in a finger such that the changes in optical absorption and scattering properties of the tissue produced by cardiac related blood volume pulses can be determined accurately.

If we use incident radiation at various wavelengths (λ) with intensity (I₀(λ)) to interrogate the tissue, the transmitted radiation intensity, Iᵣ(λ), can be obtained. This transmitted intensity Iᵣ(λ) contains very small pulsatile changes in intensity (ΔIᵣ(λ)). These observed pulsatile optical changes are due to cardiac related arterial volume changes. The overall optical signal obtained is the photoplethysmogram (PPG). If the transmitted intensities at time ti and ti+1 (Iᵣ(ti), Iᵣ(ti+1)) are detected during the same cardiac cycle, we can obtain the optical densities at ti (ODᵣ(ti)=log(I₀(λ)/Iᵣ(ti))) and ti+1 (ODᵣ(ti+1)=log(I₀(λ)/Iᵣ(ti+1))). Hence we can calculate the difference between ODᵣ(ti) and ODᵣ(ti+1), (ΔODᵣ(ti)=log(Iᵣ(ti+1)/Iᵣ(ti)))=log(1+ΔIᵣ(ti)/Iᵣ(ti)), where Iᵣ(ti)=Iᵣ(ti)+ΔIᵣ(ti), as shown in the upper-right part of Fig. (2). This has the effect of removing the interfering elements to acquire only the change in intensity due to the pulsating blood volume.

When we choose ΔIᵣ(ti) as a pulsatile component ΔIᵣ for different wavelengths, λ₁, λ₂, ..., λₙ, ΔODᵣ can be calculated as a function of λ. If such spectra are derived at different BGL levels (BGLᵢ), we can obtain a family of spectra, as schematically shown in the right part of Fig. (2), which may then be used for multivariate analyses. Calibration models constructed by several multivariate analyses are applied for spectrophotometric BGL measurements. Measured spectra of ΔODᵣ vs λ for the unknown BGL derived from in vivo measurements (see the left lower part of Fig. (2)) on a finger can be compared with the derived models and predicted BGL values thereby calculated.

![Image](https://via.placeholder.com/150)

Fig. (1). Simplified optical model of living tissue.

B. Experimental Setup

For the present study 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 low-OH optical 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, the transmitted radiation intensity (Iᵣ(λ)) over the wavelength (λ) range from 900 to 1700 nm was measured with a resolution of better than 8 nm and with 16-bit digitization. The maximum spectrum sampling speed achievable with this system is 1800 spectra/s, but in the experiment described here we adopted a speed of 100 spectra/s, which was sufficient to obtain accurately the cardiac-related transmitted intensity change superimposed on the intensity (Iᵣ(λ)) (ΔIᵣ: i.e., photoplethysmographic pulsation) over this wavelength range.

C. Subjects and Experimental Procedures

In order to carry out in vivo measurements, oral glucose tolerance tests (OGTT) were carried out to create varying BGL levels in 10 healthy volunteers (22 to 59 years old; 8 males and 2 females). Prior to the experiments, the local ethics committee approved the experimental procedures and informed consent was obtained from each subject. The
subjects were asked to abstain from food and alcohol from 9 p.m. on the previous day until the end of the experiment on the next morning. The experiment began at 9 a.m., in a room maintained at a temperature of 25 °C. During the experiment, the subjects were also requested to sit quietly in a chair to undergo the test.

In this study, radiation intensity measurements were made in the left index fingertip at 10 min intervals before and after oral uptake of glucose solution (75 g/225 ml, Trelan-G75, Shimizu Seiyaku, Co. Ltd., Japan) for a study period of 100 or 120 min. During the optical measurement, the subject’s left hand was held horizontally at heart level. Immediately after each optical measurement blood samples (about 3 ml) were collected from the cephalic vein of the left forearm and analyzed by an automatic blood analyzer (DRI-CHEM 7000, Fujifilm Medical Co. Ltd., Japan) to obtain the reference (measured) BGL values (regressand for calibration models). The BGL values obtained in the present study ranged from 100.7 to 246.3 mg/dl (5.59-13.7 mmol/l).

For the optical intensity measurements, the fingertip was placed carefully in an adjustable space between the ends of the transmitting and detecting fibers so as to ensure gentle contact with the skin, as shown in the left upper part of Fig. (2). The intensity measurements were made for about 20 to 30 s (corresponding 20 to 30 cardiac pulses), and stable and almost constant intensity signals (I^t and ΔI^t) for 3 to 5 successive cardiac beats were used as analytical data.

D. Signal Pre-Processing

To improve further the signal-to-noise ratio (S/N) of the pulsatile intensity change (ΔI^t) superimposed on the I^t signal over the wavelength range from 900 to 1700 nm, the Savitzky-Golay filtering method, which is used generally for signal pre-processing, was applied for both wavelength (λ) and time (t). Parameters of the Savitzky-Golay filters were empirically selected; the polynomial order was 2 and the frame size was 3 for the wavelength direction and the polynomial order 2 and the frame size 15 for the time-sequential direction. With these parameters an S/N of more than 25 dB was obtained.

The mean OD spectrum calculated from the 3 to 5 successive cardiac beats was used for multivariate analyses, and a family of these spectra for different BGL levels (known BGL values) served as the data-set. The OD spectra thus obtained were then normalized (OD^norm) using the AOD^t value at 1100 nm as 1.0 together with its minimum value as 0 in each spectrum [13]. Finally, the second order difference spectra (ΔOD^norm/Δλ; Δλ=8 nm) were derived to reject baseline fluctuations of spectra. This second order difference spectral method is usually applied in the field of chemometrics, in order to remove low frequency sources of variations that are not related to solution concerned [19]. In the calculation of the spectra δλ was equivalent to the wavelength resolution of the measurement system (8 nm).

The processing of these data was carried out using MATLAB™ version 7.x software and a conventional
personal computer (Dimension 9100 with a CPU Pentium D 830 and 2GB memory, DELL Inc., Japan).

E. Multivariate Calibration Regression and Classification Models

Using the data-sets of the difference spectra (regressor) and the corresponding known (measured) BGL values (regressand), two linear regression methods, PCR and PLS, and a non-linear regression method, SVMsR, were applied to create multivariate calibration models. Table 1 shows features of each multivariate regression method. 5-fold cross-validation was introduced to evaluate the calibration models. In order to obtain the resultant calibration model, parameters on each regression were searched repeatedly.

Then, the classification model constructed by SVMsC was attempted for the data-sets described above. Firstly, the known BGL levels were classified into three classes on the basis of criteria used for diabetes screening [20]. In this study, optical data $\delta^2 \Delta OD_{norm}/\delta \lambda^2$ were classified by SVMsC in the case of fasting blood glucose (FBG) as "Normal class (BGL < 110 mg/dl)", "Impaired Fasting Glycaemia (IFG) class (110 ≤ BGL ≤ 126)" or "Diabetes Mellitus (DM) class (BGL > 126)". On the other hand, data were classified in the case of 2-hour postprandial blood glucose (PBG) as "Normal class (BGL < 140)", "Impaired Glucose Tolerance (IGT) class (140 ≤ BGL ≤ 200)" or "Diabetes Mellitus class (BGL > 200)". Then, 5-fold cross-validation was also applied to evaluate the calibration models. Parameters on each classification were also searched repeatedly.

To implement these calibration procedures, the software "R" version 2.3.1 and the kernlab module version 08-2 for the software "R" [21-23] were used on a conventional personal computer.

RESULTS

Obtained optical data containing noise and/or artifact were rejected manually and then 100 data-sets were selected. Fig. (3) shows an example of the differential optical density ((a): $\Delta OD$), normalized $\Delta OD$ ((b): $\Delta OD_{norm}$), and the 2nd derivatives $\delta^2 \Delta OD_{norm}/\delta \lambda^2$ obtained in the fingertip of a healthy female subject. Using a family of these spectra determined in all subjects studied, multivariate calibration models were constructed and evaluated. Resultant parameters for each method were obtained as follows. For PCR, the 1st to 20th principal components were introduced into the regression model. For PLS, the number of latent variable was reduced to 15 against a data dimension (dimension of regressor) of 100. For PCR the parameter determination was done by evaluating models for the first n principle components with the parameter n from 1 to 100. Then, the best parameter n = 20 was selected by counting the number of plots falling within Region A of Clarke error grid. In the parameter determination of PLS, the models with latency variables from 1 to 100 were evaluated, and the variable of 15 was selected in the same way. For SVMsR, the ANOVA RBF (radial basis function) kernel with a degree of 0.01 and sigma of 1 was used in training and $\epsilon$ in Vapnik's insensitive-loss function was 0.123. For SVMsC, the ANOVA RBF kernel with degree of 0.01 and sigma of 1 was used in training and $\epsilon$ in Vapnik's insensitive-loss function was 0.15. Table 1 summarizes these multivariate analysis methods and the obtained parameters.

Table 1  Summary of Multivariate Regression and Classification Methods and Parameters Used in the Present Study

<table>
<thead>
<tr>
<th>Methods</th>
<th>Regression/Classification</th>
<th>Linear/Non Linear</th>
<th>Calculation Time</th>
<th>Parameters</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCR</td>
<td>Regression</td>
<td>Linear</td>
<td>Short</td>
<td>Applied Principal Components</td>
<td>1st to 20th</td>
</tr>
<tr>
<td>PLS</td>
<td>Regression</td>
<td>Linear</td>
<td>Short</td>
<td>Number of Latency Variable</td>
<td>15</td>
</tr>
<tr>
<td>SVMsR</td>
<td>Regression</td>
<td>Non Linear</td>
<td>Intermediate</td>
<td>Kernel Parameter Degree</td>
<td>ANOVA RBF</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Kernel Parameter Sigma</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$\epsilon$ in Vapnik's Insensitive Loss Function</td>
<td>0.01</td>
</tr>
<tr>
<td>SVMsC</td>
<td>Classification</td>
<td>Non Linear</td>
<td>Intermediate</td>
<td>Kernel Parameter Degree</td>
<td>ANOVA RBF</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Kernel Parameter Sigma</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$\epsilon$ in Vapnik's Insensitive Loss Function</td>
<td>0.35</td>
</tr>
</tbody>
</table>

As shown in this diagram, almost all data points were within clinically acceptable regions [24], that is, the regions A and B in each calibration. Among them, the SVMsR calibration model provided the best plot distributions (Region A: 84/100, Region B: 16/100 and Region C-E: 0/100) compared with the PCR model (Region A: 78/100, Region B: 22/100 and Region C-E: 0/100) and the PLS model (Region A: 80/100, Region B: 19/100 and Region C-E: 1/100). Therefore, it might reasonably be suggested that SVMsR can be used for constructing a multivariate calibration model as part of the procedure of implementing "pulse glucometry". Table 2 shows the result using the nonlinear classification model (SVMsC). In the case of FBG (Table 2a), the accuracy rate (and the probability of misclassification) was 2/4 (2/4) for Normal, 3/11 (8/11) for IFG, and 80/85 (5/85) for DM. In the case of...
Multivariate Regression & Classification Models for Pulse Glucometry

The Open Optics Journal, 2009, Volume 3

In the present study, in order to predict blood glucose (BGL) levels using “pulse glucometry”, we introduced two linear (PCR and PLS) and one nonlinear (SVMsR) calibration methods. Our results indicated that there were no significant differences in the performance of these three calibration methods, as shown in the Clarke error plots (see Fig. 4), although the SVMsR model appears to produce a better regression than the others. However, it is not to say that the regression by SVMsR is superior to the other methods, since each has strengths and weaknesses. It is the case that the PLS approach is one of the most widely used regression techniques in the chemometrics field [16]. From a

DISCUSSION

PBG (Table 2b), these were 13/23 (10/23) for Normal, 53/69 (16/69) for IGT and, 4/8 (4/8) for DM. Additionally, the mean square errors of each calibration model were as follows: PCR; 28.15 [mg/dl], PLS; 31.54 [mg/dl] and SVMsR; 28.27 [mg/dl].

Fig. (3). Optical data obtained over the wavelength range from 900 nm to 1700 nm by the very-fast NIR spectrophotometric system obtained in the fingertip of a healthy female subject: (a) differential optical density ($\Delta OD$), (b) normalized differential optical density ($\Delta OD_{\text{norm}}$), and (c) the 2nd derivatives of normalized optical density spectrum ($\Delta^2 OD_{\text{norm}}/\Delta\lambda^2$).

Fig. (4). Scatter diagrams evaluated by 5-fold cross validation showing comparison of measured and predicted BGL values by Clarke error grid analysis: (a) PCR, (b) PLS, and (c) SVMsR.
It is perhaps unrealistic at present to expect that non-invasive BGL measurement could be a more accurate alternative to invasive methods. From the standpoint of practical use, the classification type calibration (qualitative in nature), attempted in the present study, might be valuable to consider in favour of the regression type calibration (providing quantitative BGL levels). At present, diabetes sufferers seem compelled to rely on frequent and long-term finger-pricking for the control of BGL levels using an SMBG device. The classification approach using a non-invasive method like “pulse glucometry” could therefore be a viable alternative for those with established diabetes, for pre-diabetes screening, as well as for health care in normal subjects, offering an easier and more convenient way of achieving rapid self BGL monitoring during normal daily life. It is also notable that, in the field of urine glucose test strips, rough quantification of urine glucose levels has been successful and is widely utilized. Needless to say, regression type calibration is necessary for accurate management of diabetes, such as self-adjustment of insulin dosage. The present study therefore deals with both possible approaches of the quantitative and qualitative BGL assessment using multivariate calibration and classification models. The present classification result seems to be good for groups with a large number of subjects; DM group in the case of FBG and IGT group in the case of PBG. Since the numbers of data sets in this study were not sufficient for the groups of Normal and IFG in FBG and for those of Normal and DM in PBG, it is at present difficult to lead make firm conclusions on the adequacy of classification. Further investigations will be needed to obtain an appropriate number of data sets so as to allow a statistical analysis.

Since the present study is a first stage to validate the applicability of multivariate calibration and classification methods to the newly proposed “pulse glucometry”, further work is still required to obtain comprehensive optical data in a large group of diabetic subjects and also to assess the full potential of this method for clinical and research purposes. Currently, OGTT is recommended for evaluation of patients in whom diabetes is still strongly suspected but who have normal FPG (fasting plasma glucose) or IFG (impaired fasting glucose) [28]. Therefore, diabetes mellitus patients were not selected as subjects in this preliminary phase study. In the near future, however, we will plan a safe experimental procedure for diabetic patients. In addition, the present experimental setup is laboratory-based and thus further development is needed to
design a small portable device using several LEDs as light sources which will be suitable for use by patients and normal subjects to achieve self-monitoring of BGL.

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 classification analysis (SVMsC) to construct multivariate calibration models. The present method provides clinically promising BGL measurements with multivariate regressions. Also, a proper classification model was obtained using the nonlinear method, SVMsC. These results provide preliminary evidence that "pulse glucometry" with SVMsR can be applied effectively to predict BGL levels non-invasively and a classification type calibration could be applicable for diabetes screening and health care in normal subjects in a more practical, easier and more convenient way through rapid BGL assessment. Further comprehensive studies will need to be carried out in order to investigate an optimal construction of multivariate regression and classification models including data from diabetic subjects as well as to design a compact BGL device based on "pulse glucometry" for practical use.

ACKNOWLEDGMENT

The authors wish to give their sincere thanks to Professor K. Yamakoshi, Kanazawa University, and Prof. P. Rolfe, Harbin Institute of Technology, China, and OBH Ltd., UK, for their valuable suggestions and comments on this work and for their assistance in preparing the manuscript. This work was partly supported by Grant-in-Aid for Scientific Research (A) of the Ministry of Education, Culture, Sports, Science and Technology, Japan.

REFERENCES