

Unsupervised calibration for noninvasive glucose-monitoring devices using mid-infrared spectroscopy

Ryosuke Kasahara^{*,†,§}, Saiko Kino[‡], Shunsuke Soyama[†]
and Yuji Matsuura^{†,‡}

**Ricoh Institute of Information and Communication Technology
Research and Development Division, Ricoh Company
2-7-1 Izumi, Ebina 243-0460, Japan*

*†Graduate School of Engineering, Tohoku University
6-6-05 Aoba, Sendai 980-8579, Japan*

*‡Graduate School of Biomedical Engineering
Tohoku University, 6-6-05 Aoba, Sendai 980-8579, Japan
§ryohsuke.kasahara@jp.ricoh.com*

Received 12 April 2018

Accepted 17 September 2018

Published 29 October 2018

Noninvasive, glucose-monitoring technologies using infrared spectroscopy that have been studied typically require a calibration process that involves blood collection, which renders the methods somewhat invasive. We develop a truly noninvasive, glucose-monitoring technique using mid-infrared spectroscopy that does not require blood collection for calibration by applying domain adaptation (DA) using deep neural networks to train a model that associates blood glucose concentration with mid-infrared spectral data without requiring a training dataset labeled with invasive blood sample measurements. For realizing DA, the distribution of unlabeled spectral data for calibration is considered through adversarial update during training networks for regression to blood glucose concentration. This calibration improved the correlation coefficient between the true blood glucose concentrations and predicted blood glucose concentrations from 0.38 to 0.47. The result indicates that this calibration technique improves prediction accuracy for mid-infrared glucose measurements without any invasively acquired data.

Keywords: Noninvasive glucose monitoring; calibration; mid-infrared spectroscopy; domain adaptation; neural network.

§Corresponding author.

This is an Open Access article published by World Scientific Publishing Company. It is distributed under the terms of the Creative Commons Attribution 4.0 (CC-BY) License. Further distribution of this work is permitted, provided the original work is properly cited.

1. Introduction

In recent years, the incidence of diabetes has increased worldwide, thus, increasing the market demand for noninvasive blood glucose-monitoring technologies. Various methods have been proposed for noninvasive blood glucose measurements, including near-infrared sensing,¹⁻³ mid-infrared sensing,⁴ Raman spectroscopy,⁵ and photoacoustics.^{6,7} Of these options, the mid-infrared light spectrum offers good detection accuracy because glucose absorbs light particularly well at these wavelengths. Practical applications of noninvasive blood glucose measurement technologies are limited by the measurement accuracy⁸ and require invasive calibration steps for practical use.

Calibration is vital to ensure that glucose measurements are robust against individual differences and that the noninvasive measurement is maximally correlated with direct glucose concentration readings.⁹⁻¹² In these calibration routines, blood sampling is indispensable to prepare training data; to obtain highly accurate measurements, the blood glucose level must be invasively measured at least once. We propose a novel method to calibrate noninvasive glucose measurements with spectral data alone, which could lead to a truly noninvasive blood glucose-monitoring system.

Machine-learning techniques have been used to analyze a wide range of biomedical data for applications, including those for spectral analysis and disease prediction.¹³⁻¹⁵ Previous studies have focused on creating regression models using supervised learning. Training data that are labeled with reference measurements are required for normal supervised learning. To avoid the need for a labeled training dataset, we applied domain adaptation (DA) using a deep neural network (NN) to learn a model. When we have a model created from original training data, DA can be used to adapt this model to different training data with no label information having a distribution that is different from that of the original training data. DA is used, for example, when adapting a person's spam filter model to the contents of other people's mail with different content distributions.¹⁶

2. Materials and Methods

Previously, a method that measures light spectrum absorbance in the oral mucosa has been proposed.¹⁷

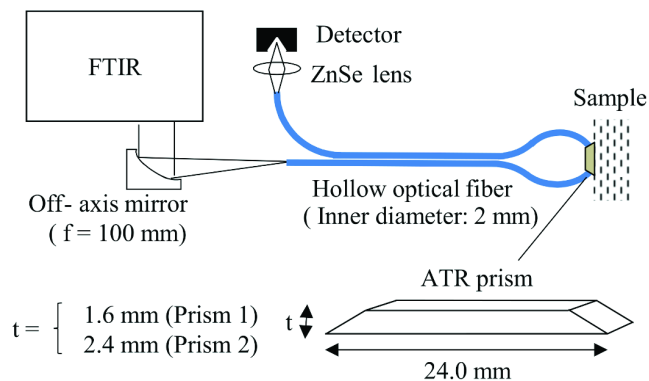


Fig. 1. Experimental setup and dimensions of ATR prism.

We have developed a technique that transmits optical signals using an attenuated total reflection (ATR) prism and a hollow optical fiber¹⁸ that efficiently propagates mid-infrared light.¹⁹ We previously reported the accuracy of noninvasive blood glucose level measurements that were obtained using this device^{20,21}; Fig. 1 shows an outline of the measurement system. We measured the absorbance of the oral mucosa using an ATR prism sandwiched in the patient's inner lip. The ATR prism is made of ZnS. Two Fourier-transform infrared spectroscopy (FTIR) devices (Tensor 27, Bruker, Billerica, Massachusetts, US and Vertex 70, Bruker) are used as mid-infrared spectrometers. Measurements are recorded using ATR prisms (prism 1 and prism 2) of different thicknesses in order to check whether the prediction model of blood glucose level depends on the design of the prism.

Tissues are generally considered as turbid media with high scattering and absorption properties, and in the mid-infrared region, light only penetrates shallow tissue samples. Because the experimental setup used in this study employed an ATR prism as the probe, the light transmitted from the tissue was not used, and only a few microns which the light of the prism that bounced off the reflecting surface were measured. Since the ATR prism exhibited a large number of surface reflections, it was possible to sense the surface conditions with a good signal-to-noise ratio (SNR). The oral mucosa was found to be a suitable tissue for ATR-based sensing because it did not include an epidermis layer. In this study, we investigated the effect of the interstitial fluid on the glucose levels via the vessel wall by applying this method.

To collect reference data, we used two types of self-administering glucose-monitoring devices

Table 1. Dataset properties.

	Labeled dataset	Unlabeled dataset
The number of samples	131	414
The number of subjects	1	5
The number of series	13	18
Kind of meal (various meal or glucose drink)	Various meal (13)	Various meal (16)/Glucose drink (2)
FTIR apparatus	Tensor (13)	Vertex (5)/Tensor (13)
ATR prism	Prism1 (13)	Prism1 (3)/Prism2 (15)
Blood glucose monitor	OneTouch UltraView (13)	OneTouch UltraView (4)/Medisafe Mini (14)
Data acquisition date	10/2016–1/2017	6/2015–11/2016

(Medisafe Mini, Terumo, Tokyo, Japan and OneTouch UltraView, Johnson and Johnson, New Brunswick, New Jersey, US). As the blood glucose levels measured by the two reference devices for the same blood sample differed somewhat, the values measured by the Medisafe Mini were corrected with a linear equation so they matched those from the OneTouch UltraView. Blood sampling and measurements of blood glucose levels were initially performed every 10 min (this was done shortly after the meals when significant changes were still likely occurring), and thereafter, the time interval was increased. The analyses were conducted for approximately 3 h until the blood glucose levels stabilized. In addition, spectral measurements were performed. However, the timing of blood sampling and the number and timing of spectral measurements did not always match, so the blood glucose levels at the time of the spectral measurements were obtained by linear interpolation of the glucose levels obtained from the analysis of the blood samples. We collected blood samples and simultaneously recorded the absorption spectrum of the oral mucosa. Each dataset was collated from a series of measurements taken on a single day.

Two datasets, the characteristics of which are listed in Table 1, were prepared from the measured data. The labeled dataset, containing 131 data points from 13 measurement series, was constructed by requiring the subject to eat a variety of meals before the measurements. The unlabeled dataset contains 414 data points from 18 measurement series. The subject in the labeled dataset was a healthy male in his mid-20s. The subjects in the unlabeled dataset were four healthy men and a healthy woman with their age ranging from 30 to 50. Only one person participated in a series. For this

Table 2. Number of data points.

	Labeled dataset	Unlabeled dataset
Mean \pm SD	10.08 \pm 0.76	23.00 \pm 14.13
Min	9	9
Max	11	55

dataset, the measurements were performed after eating various meals or drinking a glucose solution (75 g of glucose dissolved in 150 mL of water). The unlabeled dataset also includes data acquired from different ATR prisms and different FTIR devices. Parentheses in the table indicate the number of the corresponding measurement series. Our protocol was approved by the ethical committee on the Use of Humans as Experimental Subjects of Tohoku University, and informed consent was obtained from all examinees. Table 2 shows the minimum, maximum, and average values (as well as the variance) obtained from the data points for a series of datasets.

When considering a noninvasive glucose measuring device for practical use, we have assumed the following handling procedures for the two datasets: Device manufacturers recorded the labeled dataset to the device before shipment to create predictive models. The user acquired the unlabeled dataset after shipment and predicted the blood sugar levels from the unlabeled dataset.

3. Calculations

Previously, we found that in the mid-infrared spectrum, even when only three appropriately chosen wavenumbers are used for predicting blood glucose levels, the regression correlations were equal

to or higher than the correlations obtained when using more wavenumbers.²² Partial least squares regression (PLS),²³ support vector machine,²⁴ and NN models^{25,26} have all been proposed as models to predict blood glucose levels from spectrum data. In addition, to avoid the difficulty of labeling a large volume of data, transfer learning is often used to apply learning results from one classification task to other. Of the transfer learning techniques, DA allows machine-learning models to successfully predict test datasets when the training dataset has different distribution from the test dataset. Domain-adversarial NNs (DANNs)²⁷ have been proposed to be an implementation of DA. Adversarial updating in deep NNs improves prediction accuracy. We used DANN to calibrate the algorithm that associates spectral data with blood glucose levels; the ability of DANN for unsupervised learning allows calibrations without a training dataset labeled with invasive blood samples.

3.1. Process flow

Figure 2 shows the process flow for the preprocessing, training, and evaluation of regression results. We applied series cross-validation and multiple linear regression (MLR) models to select an appropriate mid-infrared radiation for noninvasive blood glucose measurements, and only wavenumbers that had a high correlation unconditionally were selected.²² The selected wavenumbers were 1050, 1070, and 1100 cm^{-1} and to reduce disturbances, such as those caused by the contact pressure of the prism, a wavenumber of 1000 cm^{-1} was used, which has low absorption for normalization. Therefore, the

wavenumbers used for regressions were 1050 cm^{-1} , 1070 cm^{-1} , and 1100 cm^{-1} , normalized at 1000 cm^{-1} for all data. Certain experimental conditions, especially the contact pressure and temperature, can affect spectroscopy measurements. Temperature changes are superimposed on the measured value as the blackbody radiation spectrum of the light source of the FTIR device, although this effect can be eliminated by subtracting the background from the FTIR measurement. The contact pressure of the ATR prism can also affect the results, but due to normalization at 1000 cm^{-1} , this effect is mostly canceled out. Because of the time taken by blood to pass from vessels to tissue fluids,²⁸ we used the measured value from the self-measuring devices delayed by 26 min.²² Once the labeled and unlabeled datasets were subjected to this preprocessing, the labeled dataset was used as training data and each series of the unlabeled dataset was used as test data. Using the tests for each series, we tested the accuracy of regression with all the data in the unlabeled dataset. When applying DA, the unlabeled dataset was also used as an unlabeled training dataset.

Figure 3 illustrates the grouping of datasets. Differences in color and shape in the figure indicate differences in series. All series in the labeled dataset are labeled with blood glucose concentration, unlabeled data from one series in the unlabeled dataset are used for training, and the same series of the unlabeled dataset is used for testing. This process is repeated for each series in the unlabeled dataset, and the regression accuracy is calculated. In the training step, labels from the unlabeled dataset are not included. Therefore, even though the same series from the unlabeled dataset are given in both

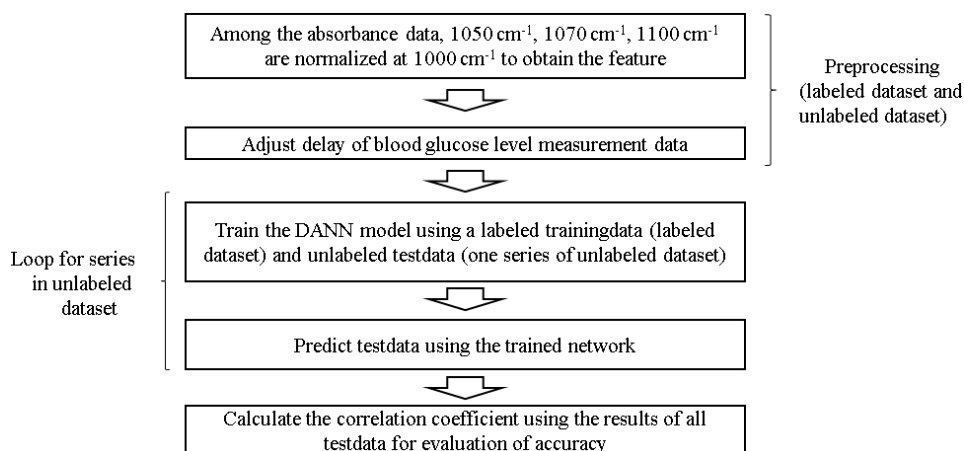


Fig. 2. Process flow of training and evaluation of prediction results.

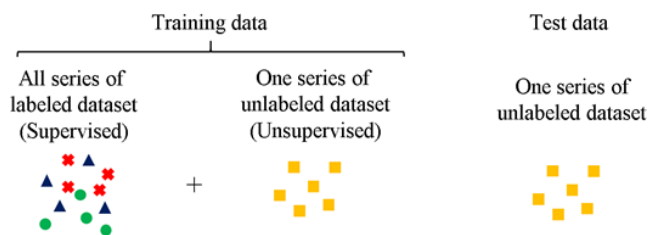


Fig. 3. Treatment of training data and test data.

training and testing, the true value of blood glucose concentration is not provided during training. The number of data points was 131 for supervised training, the average number of data points per data series was 23 for unsupervised training, and 23 for the test. No validation data were used. The unlabeled dataset was used to predict the blood glucose levels obtained by the user of the glucose-monitoring device, and we assumed that the data had no labeled data of the blood glucose level. Therefore, although it was used for DA with no labeled data, it is not used for supervised learning with the labeled data.

3.2. DANN for predicting blood glucose levels

Figure 4 shows the configuration of the network used as a regressor of blood glucose levels. The network input is a set of absorbance values at 1050 cm^{-1} , 1070 cm^{-1} , and 1100 cm^{-1} . L_x and Lc_x are the layers of the network used for regression and classification, respectively, and w_x and wc_x label the weights in the corresponding layers. A leaky rectified linear unit^{29,30} with a gradient $a_i = 0.2$ in the negative region is used for the activation

function. Batch normalization³¹ is also used for each layer. Adam³² is used for optimization.

3.3. Training method

Figure 5 shows the training procedure used. First, in step 1, we train the network to regress the blood glucose level from the absorbance data in the labeled dataset. Then, in step 2, unlabeled absorbance data from one series of unlabeled dataset are added as input data and the network is trained to distinguish data in the labeled dataset from data in the unlabeled dataset. In step 3, the network weights for regression (w_1, w_2) are updated before branching so that the labeled and unlabeled datasets are indistinguishable to the network. With this step, the blood glucose levels can be regressed at the output of L_3 by extracting the features that are common to the labeled and unlabeled datasets. Adversarial updating in steps 2 and 3 increases the regression accuracy by overlaying the distributions of labeled and unlabeled datasets in L_1-L_3 layers. Therefore, by adjusting deviations in the distributions in a series from labeled and unlabeled datasets, the network can estimate the blood glucose level considering the labeled and unlabeled datasets.

Training starts with step 1 using the supervised data of the labeled dataset and trains the weights of w_1-w_4 in 18,000 epochs. After 18,000 epochs, all three steps including step 2 and step 3 were executed simultaneously for 8000 epochs, and this training process also used the unlabeled data in the unlabeled dataset. To balance the regression and DA processes, step 3 was only repeated in the

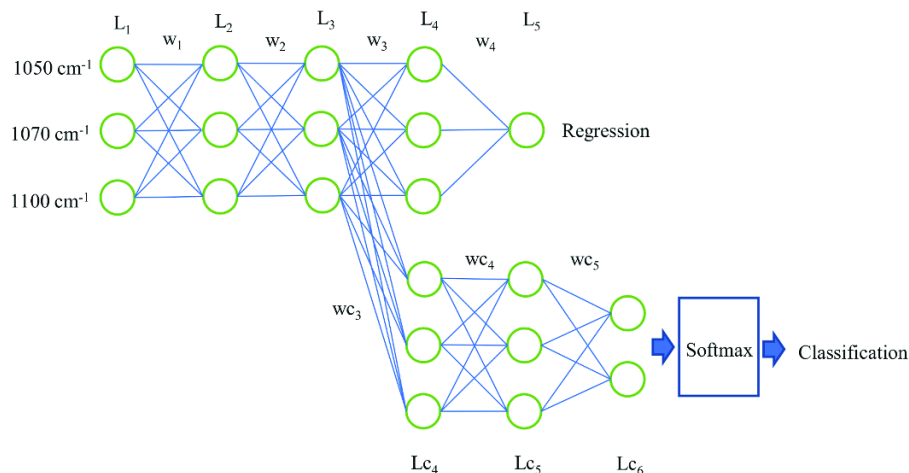


Fig. 4. Network diagram of DANN.

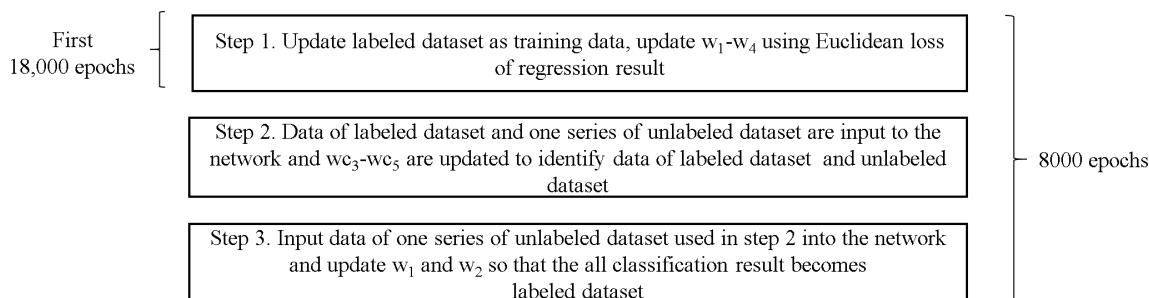


Fig. 5. Algorithm for network training.

iteration in which the loss value of the regression in step 1 was < 320 . The loss value in step 3 is multiplied by 350 to balance it with the loss values in step 1 and step 2. The training process runs for 26,000 epochs in total. All training data were randomized at each epoch. The loss functions used are the Euclidean loss for regression in step 1, and the softmax cross entropy for classification in step 2 and step 3. Note that an epoch in this report refers to a period in which all data are used for training once. The 18,000 and 8000 epochs were chosen to obtain the best correlation coefficient with all the test data series (414 samples). It was necessary to search the parameters using except the test data series to be evaluated fundamentally. However, in this study, only one of the 18 series required removal from the parameter search, and since there was only a small number of parameters that needed to be optimized, the possibility that these would become overfitted was relatively low. Thus, we applied this method due to limitations associated with the calculation time.

4. Results and Discussion

The average values and variances obtained for the three wavelengths of the two datasets are presented in Table 3. Individual differences and variations in the measurement environments (i.e., prism and

Table 3. Absorbance normalized by a wavenumber of 1000 cm^{-1} for each dataset.

Wavenumber (cm^{-1})	Mean \pm SD	
	Labeled dataset	Unlabeled dataset
1050	1.071 ± 0.033	1.153 ± 0.045
1070	1.103 ± 0.039	1.182 ± 0.053
1100	1.081 ± 0.034	1.134 ± 0.040

FTIR device differences) could explain the various values observed.

Figure 6 shows the change of loss values at each epoch of the training process. Each loss decreases as training progresses. A decrease in the loss value of step 1 indicates that the network is training regression for the labeled dataset. If the loss value of step 2 is decreasing, the network is training classifiers for the labeled and unlabeled datasets. If the loss value of step 3 is decreasing, the training is progressing such that the distributions of labeled and unlabeled datasets overlap in the middle of the regression network. If these losses are reduced and well balanced, the network realizes DA.

Figure 7 shows a comparison of data distributions with and without DA in a typical series from the unlabeled dataset. Figure 7(a) shows the data input to the L_1 layer, and Fig. 7(b) shows the output results from the L_3 layer. The three-dimensional spectrum values of datasets are reduced and plotted in two dimensions using principal component analysis. The colors and shapes in the graph indicate whether the data points are from the labeled or unlabeled datasets. At the input stage, the distributions of labeled and unlabeled datasets are

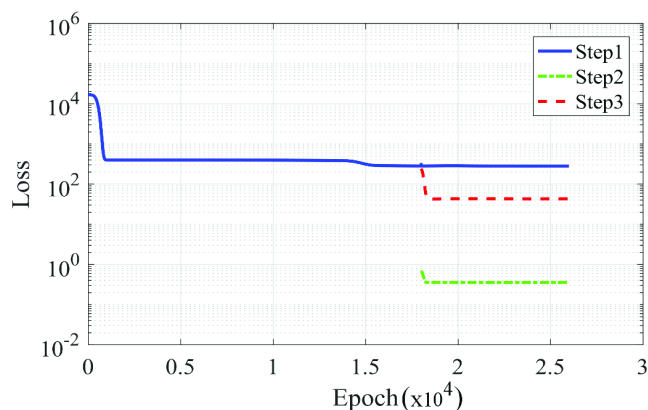


Fig. 6. Loss reduction in learning process.

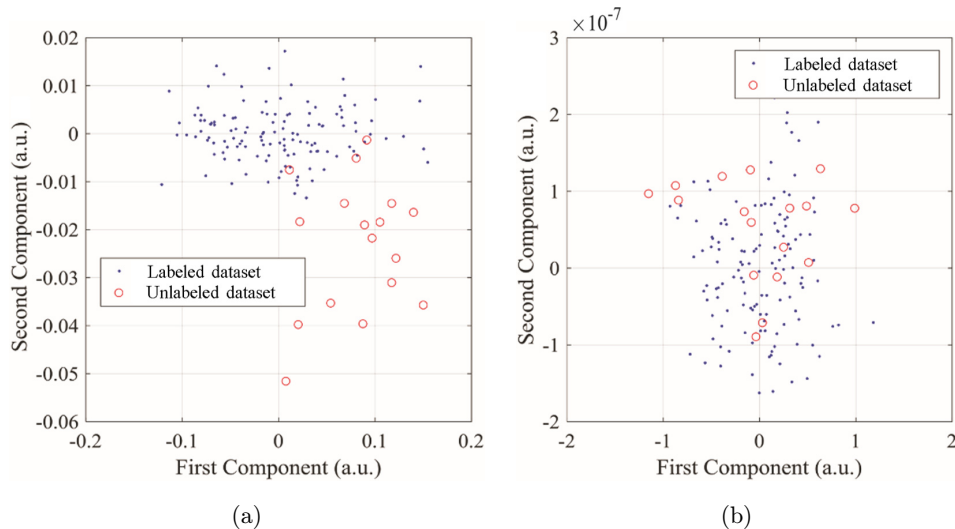


Fig. 7. Comparison of data distributions in typical series: (a) L_1 input layer and (b) L_3 output layer.

clearly shifted. In the output from the L_3 layer, however, the distributions are clearly superimposed, which show that this network adjusts for the differences of subjects between the labeled and unlabeled datasets.

Next, Fig. 8 shows a comparison of prediction accuracy with and without DA application, using the Clark error grid,³³ which is the commonly used scatter diagram for evaluating the accuracy of glucose-measuring devices. Figure 8(a) shows the Clark error grid for the unlabeled dataset in the prediction model trained from the labeled dataset series with only step 1 executed so that DA was not applied. For the result in which DA was applied, Fig. 8(b) shows the Clark error grid for the unlabeled dataset

in the prediction model trained by executing all three steps.

For the prediction model trained without DA, the correlation coefficient was 0.38, and 53.6% of predictions fall into region A in the Clark error grid. For the prediction model trained with DA, the correlation coefficient was 0.47, and 63.8% of predictions fall in region A in the Clark error grid. This indicates that DA improves the correlation coefficient and prediction accuracy. This result also shows that calibration of the prediction model without blood samples is possible with DA. Note that the test data were measured under various conditions such as recent meals, subjects, and the experimental setups. Therefore, this correlation

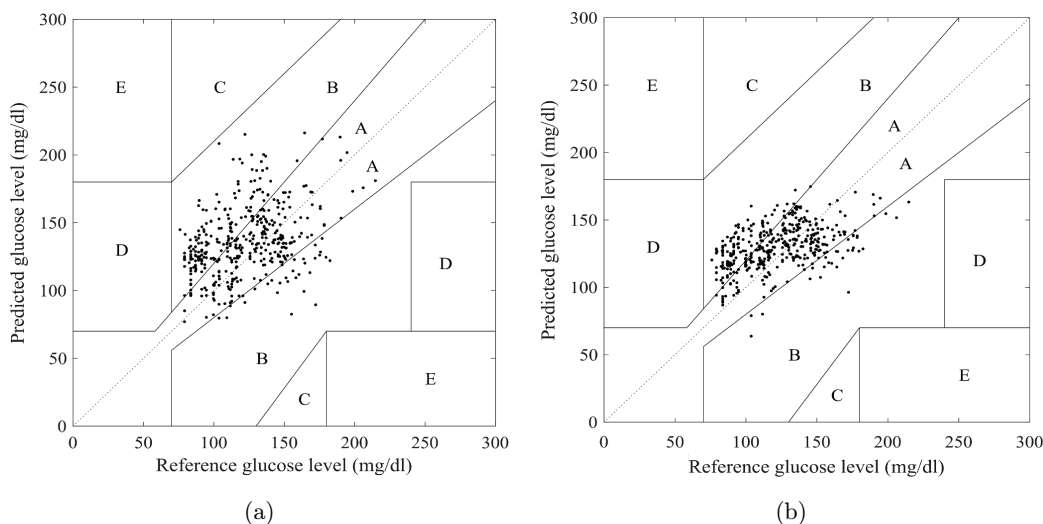


Fig. 8. Clark error grid of the unlabeled dataset for the estimation models: (a) without DA and (b) with DA.

Table 4. Comparison of models.

	MLR ²²	PLS ²²	NN	DANN
Correlation coefficient	0.36	0.25	0.38	0.47
Percentage of data points in region A (%)	51.2	43.0	53.6	63.8
Root mean square error (mg/dl)	32.3	41.7	32.4	25.6
Mean absolute difference (mg/dl)	26.3	33.4	26.6	21.4

result also shows that the model has unconditional regression performance.

Next, we compare the results obtained by comparing the machine-learning models. Table 4 shows the comparison of the correlation coefficient in various models and the percentage of data points in region A of the Clark error grid. It also shows the root mean square error and the mean absolute difference for the obtained data. The results for MLR and PLS are reproduced from previous studies²² in which the data were same as for this study. Table 4 also lists results from an NN without DA and the present test with a DANN using DA. All methods have the same conditions, wherein calibration with blood sampling was not performed. For methods other than DANN, calibration was not performed for each series of the unlabeled dataset. PLS also has a function for wavenumber selection, so its inputs included wide-spectrum absorbance values (every 2 cm^{-1} from 980 to 1200 cm^{-1}). The other models' input wavenumbers were 1050 , 1070 , and 1100 cm^{-1} .

The result shows that PLS, which is generally used for spectrum analysis, does not yield acceptable results. We suspect that this poor performance can be explained as an effect of overfitting because the wavenumber of the input spectrum was larger than the number of data. NN models can handle nonlinear components, which we assume that this explains this approach's slightly better accuracy over MLR. DANN shows the best results of the methods we tested.

The prediction accuracy of the method was also evaluated using a data series for a subject in the situation of applying DA with another series of data for the same person. However, the results varied significantly depending on which data series was applied for the DA. For example, when the DA using different data was applied for the same person, good results were not always obtained, which may be attributed to the differences in the data-acquisition environment and the meal content, even

for the same subject. However, it was possible to apply DA each time a data series was acquired when considering the actual use. Therefore, the user of the glucose-monitoring device could improve the prediction accuracy for the acquired data series by applying the DA method.

We also evaluated the prediction accuracy of the method using series cross-validation in case all the data from the labeled and unlabeled datasets were used for supervised training. The obtained correlation coefficient was estimated to be 0.35 , which was lower than the NN obtained with only the labeled dataset. This is because although measurement conditions of labeled dataset were stable, measurement conditions of the unlabeled dataset varied widely, so the NN could not learn a suitable regression model for blood glucose levels. For this reason, it is better to learn the regression model using data measured under stable conditions and applying DA by DANN for each series.

There are a few potential reasons for the shifting or distortion of the measurement series as follows: First, there are individual differences in the metabolic systems, and since this method does not directly observe glucose, it is affected by metabolism. Second, different meal content has different effects on metabolism. Third, there are individual structural differences in skin-depth direction.³⁴ Fourth, there are structural differences in the horizontal direction, which are also related to the position of the ATR probe.³⁵ For differences related to optical issues (refer to the third and fourth reasons), calibration can be performed by selecting of optimal area with respect to depth direction and horizontal direction. On the contrary, our approach is based on a machine-learning technique. By combining the DA technique with the optical approach, it is possible to calibrate not only optical differences but also metabolism-based differences, which is intended to serve as a foundation for further improvements of the method.

Using the proposed DA method, blood sampling at the time of calibration is rendered unnecessary. This removes the major barrier in using noninvasive glucose-monitoring technology. Without requiring invasive calibration, prediction accuracy can be improved at the time of actual use, which could make noninvasive glucose-monitoring systems available for home use. Additionally, this method can be applied more generally to other medical measurements that usually need to be calibrated for use with each individual using invasively obtained samples. For example, the technique can be applied for noninvasive measurements of other blood components.

5. Conclusion

We developed a method for calibrating mid-infrared spectral blood glucose data without training data derived from blood samples. We showed that a DANN performing DA can be applied to unsupervised calibration with unlabeled spectral data. The training process was successful and improved the correlation coefficient and prediction error over comparable methods. Using the method, blood sampling at the time of calibration becomes unnecessary and the accuracy of prediction of noninvasive glucose-monitoring systems for home use can be improved. In the future, we plan to test the accuracy of this method with data collected from a device that uses a laser light source, which would be a more practical approach for patient use.

Conflict of Interest

The authors declare that there are no conflicts of interest related to this article.

Acknowledgement

The authors are grateful to Daishiro Kobayashi for helpful discussions on DANN.

References

1. J. Yadav, A. Rani, V. Singh, B. M. Murari, "Prospects and limitations of non-invasive blood glucose monitoring using near-infrared spectroscopy," *Biomed. Signal. Process. Control* **18**, 214–227 (2015).
2. K. Yamakoshi, Y. Yamakoshi, "Pulse glucometry: A new approach for noninvasive blood glucose measurement using instantaneous differential near-infrared spectrophotometry," *J. Biomed. Opt.* **11**, 054028 (2006).
3. J. T. Olesberg, M. A. Arnold, C. Mermelstein, J. Schmitz, J. Wagner, "Tunable laser diode system for noninvasive blood glucose measurements," *Appl. Spectrosc.* **59**, 1480–1484 (2005).
4. S. Liakat, K. A. Bors, L. Xu, C. M. Woods, J. Doyle, C. F. Gmachl, "Noninvasive *in vivo* glucose sensing on human subjects using mid-infrared light," *Biomed. Opt. Express.* **5**, 2397–2404 (2014).
5. A. M. Enejder, T. G. Seccina, J. Oh, M. Hunter, W. C. Shih, S. Sasic, G. L. Horowitz, M. S. Feld, "Raman spectroscopy for noninvasive glucose measurements," *J. Biomed. Opt.* **10**, 031114 (2005).
6. H. von Lilienfeld-Toal, M. Weidenmüller, A. Xhelaj, W. Mäntele, "A novel approach to non-invasive glucose measurement by mid-infrared spectroscopy: The combination of quantum cascade lasers (QCL) and photoacoustic detection," *Vib. Spectrosc.* **38**, 209–215 (2005).
7. M. A. Pleitez, T. Lieblein, A. Bauer, O. Hertzberg, H. von Lilienfeld-Toal, W. Mäntele, "In vivo noninvasive monitoring of glucose concentration in human epidermis by mid-infrared pulsed photoacoustic spectroscopy," *Anal. Chem.* **85**, 1013–1020 (2013).
8. S. K. Vashist, "Non-invasive glucose monitoring technology in diabetes management: A review," *Anal. Chim. Acta* **750**, 16–27 (2012).
9. K. Maruo, T. Oota, M. Tsurugi, T. Nakagawa, H. Arimoto, M. Tamura, Y. Ozaki, Y. Yamada, "New methodology to obtain a calibration model for noninvasive near-infrared blood glucose monitoring," *Appl. Spectrosc.* **60**, 441–449 (2006).
10. A. Caduff, M. Zanon, M. Mueller, P. Zakharov, Y. Feldman, O. De Feo, M. Donath, W. A. Stahel, M. S. Talary, "The effect of a global, subject, and device-specific model on a noninvasive glucose monitoring multisensor system," *J. Diabetes Sci. Technol.* **9**, 865–872 (2015).
11. I. Harman-Boehm, A. Gal, A. M. Raykhman, J. D. Zahn, E. Naidis, Y. Mayzel, "Noninvasive glucose monitoring: A novel approach," *J. Diabetes Sci. Technol.* **3**, 253–260 (2009).
12. P. P. Pai, P. K. Sanki, S. Sarangi, S. Banerjee, "Modelling, verification, and calibration of a photo-acoustics based continuous non-invasive blood glucose monitoring system," *Rev. Sci. Instrum.* **86**, 064901 (2015).
13. Y. Wang, Y. Fan, P. Bhatt, C. Davatzikos, "High-dimensional pattern regression using machine learning: From medical images to continuous clinical variables," *NeuroImage* **50**, 1519–1535 (2010).

14. R. C. Deo, "Machine learning in medicine," *Circulation* **132**, 1920–1930 (2015).
15. K. R. Foster, R. Koprowski, J. D. Skufca, "Machine learning, medical diagnosis, and biomedical engineering research — commentary," *Biomed. Eng. Online* **13**, 94 (2014).
16. L. Duan, D. Xu, I. W. H. Tsang, "Domain adaptation from multiple sources: A domain-dependent regularization approach," *IEEE Trans. Neural Netw. Learn. Syst.* **23**, 504–518 (2012).
17. H. M. Heise, R. Marbach, "Human oral mucosa studies with varying blood glucose concentration by non-invasive ATR-FT-IR-spectroscopy," *Cell. Mol. Biol.* **44**, 899–912 (1998).
18. Y. Abe, Y. Matsuura, Y. W. Shi, Y. Wang, H. Uyama, M. Miyagi, "Polymer-coated hollow fiber for CO(2) laser delivery," *Opt. Lett.* **23**, 89–90 (1998).
19. Y. Matsuura, S. Kino, T. Katagiri, "Hollow-fiber-based flexible probe for remote measurement of infrared attenuated total reflection," *Appl. Opt.* **48**, 5396–5400 (2009).
20. S. Kino, Y. Tanaka, Y. Matsuura, "Blood glucose measurement by using hollow optical fiber-based attenuated total reflection probe," *J. Biomed. Opt.* **19**, 057010 (2014).
21. S. Kino, S. Omori, T. Katagiri, Y. Matsuura, "Hollow optical-fiber based infrared spectroscopy for measurement of blood glucose level by using multi-reflection prism," *Biomed. Opt. Express* **7**, 701–708 (2016).
22. R. Kasahara, S. Kino, S. Soyama, Y. Matsuura, "Noninvasive robust glucose monitoring using mid-infrared absorption spectroscopy based on a few wavenumbers," *Biomed. Opt. Express* **9**, 289–302 (2018).
23. M. R. Robinson, R. P. Eaton, D. M. Haaland, G. W. Koepp, E. V. Thomas, B. R. Stallard, P. L. Robinson, "Noninvasive glucose monitoring in diabetic patients: A preliminary evaluation," *Clin. Chem.* **38**, 1618–1622 (1992).
24. I. Barman, C. R. Kong, N. C. Dingari, R. R. Dasari, M. S. Feld, "Development of robust calibration models using support vector machines for spectroscopic monitoring of blood glucose," *Anal. Chem.* **82**, 9719–9726 (2010).
25. S. Ramasahayam, S. H. Koppuravuri, L. Arora, S. R. Chowdhury, "Noninvasive blood glucose sensing using near infrared spectroscopy and artificial neural networks based on inverse delayed function model of neuron," *J. Med. Syst.* **39**, 166 (2015).
26. C. Fischbacher, K.-U. Jagemann, K. Danzer, U. A. Müller, L. Papenkordt, J. Schöler, "Enhancing calibration models for non-invasive near-infrared spectroscopical blood glucose determination," *Fresenius J. Anal. Chem.* **359**, 78–82 (1997).
27. Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, V. Lempitsky, U. Dogan, M. Kloft, F. Orabona, T. Tommasi, "Domain-adversarial training of neural networks," *J. Mach. Learn. Res.* **17**, 1–35 (2016).
28. M. S. Boyne, D. M. Silver, J. Kaplan, C. D. Saudek, "Timing of changes in interstitial and venous blood glucose measured with a continuous subcutaneous glucose sensor," *Diabetes* **52**, 2790–2794 (2003).
29. K. He, X. Zhang, S. Ren, J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," *IEEE Int. Conf. Computer Vision* (IEEE, New York, 2015), pp. 1026–1034.
30. A. L. Maas, A. Y. Hannun, A. Y. Ng, "Rectifier nonlinearities improve neural network acoustic models," in *Proceedings of the 30th International Conference on Machine Learning* (Atlanta, Georgia, USA, 2013), pp. 1–6.
31. S. Ioffe, C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proceedings of the 32nd International Conference on Machine Learning*, Vol. 37 (Lille, France, 2015), pp. 448–456.
32. D. P. Kingma, J. L. Ba, "Adam: A method for stochastic optimization," arXiv:1412.6980.
33. W. L. Clarke, D. Cox, L. A. Gonder-Frederick, W. Carter, S. L. Pohl, "Evaluating clinical accuracy of systems for self-monitoring of blood glucose," *Diabetes Care* **10**, 622–628 (1987).
34. O. Hertzberg, A. Bauer, A. Küderle, M. A. Pleitez, W. Mäntele, "Depth-selective photothermal IR spectroscopy of skin: Potential application for non-invasive glucose measurement," *Analyst* **142**, 495–502 (2017).
35. J. Y. Sim, C. G. Ahn, E. J. Jeong, B. K. Kim, "In vivo microscopic photoacoustic spectroscopy for non-invasive glucose monitoring invulnerable to skin secretion products," *Sci. Rep.* **8**, 1–11 (2018).