

## Quantitative analysis of glucose in whole blood using FT-Raman spectroscopy and artificial neural network

Qiaoyun Wang<sup>\*</sup>, Nianzu Zheng, Zhigang Li, and Zhenhe Ma

College of Information Science and Engineering, Northeastern University, Shenyang, 110819, China

<sup>\*</sup>wangqy@neuq.edu.cn

**Keywords:** FT-Raman spectroscopy, Glucose, Artificial neural network, Whole blood

**Abstract.** In this paper, the FT-Raman spectroscopy and artificial neural network were used to quantify the glucose concentration in the whole blood. The quantitative analysis was performed through the artificial neural network (ANN) associated to training set selection strategy method. All analysis was carried out by whole spectrum. The ANN model with the highest  $r^2$ , the lowest values of root mean square error (RMSE) for both training and validation data was used in our work. The results indicated that the FT-Raman spectroscopy and ANN is a rapid, simple, and efficient method to quantitative analysis of glucose in the whole blood.

### Introduction

Raman spectroscopy is unique among spectroscopy techniques because liquid, gas or solid samples can be analyzed directly and rapidly without any sample preparation[1]. Raman spectroscopy offers several advantages over chromatography and Infrared (IR) spectroscopy in the analysis of water-rich and multi-component samples. For example, sample preparation is not required in Raman spectroscopy, which is therefore suitable for on-line quantitative analysis. The band intensities caused by the OH stretching vibration are also weak in Raman spectra, such that a water-rich sample can be directly analyzed[2].

Although Raman spectra can be collected quite easily, the procedure for quantifying the spectra is rather cumbersome because the intensity of the Raman spectra depends on the sample concentration as well as other factors, such as the laser power and the instrumental effects. Therefore, multivariate data analysis plays an important role in the quantitative analysis of Raman spectroscopy. So the statistical methods for the quantitative for Raman spectra has been proposed, such as partial least squares (PLS)[3], principal component analysis (PCA)[4] or artificial neural networks[5-7] (ANN) approaches. In these years, the ANN, which is a mathematic technique based in human brain, has been widely used in chemometrics to replace traditional multivariate calibration methods based on models of multiple linear regression because they are able to efficiently map and extract nonlinear, noisy or incomplete relationships from the study data [4, 8].

Neural networks have been widely used in chemometrics because they are able to efficiently map and extract nonlinear, noisy, or incomplete relationships from the study data [9-10]. The neurons in the model are interconnected in layers and have weights that mediate this interconnection. The operation in the biologic neurons is based on input information (independent variables), in its processing and obtained output information (dependent variables).

This work proposes the use of FT-Raman spectroscopy and ANN to obtain an efficient predictor model that can quantify glucose in the whole blood. The results show that this method is a potential analytical tool for glucose monitoring in the whole blood.

### Materials and Methods

#### *Raman Spectroscopy*

The Raman spectra of samples were obtained with a Bruker MultiRAM (Bremen, Germany) Fourier Transform (FT) Raman spectrometer equipped with a germanium detector using liquid nitrogen as the coolant. The excitation light was generated by a near infrared Nd:YAG laser at 1064nm. And the setup of the whole blood measurement employed in this study has been shown in

Figure 1. A quartz cuvette with optically flat faces, a mirrored backside and near infrared (IR) transparency was used as the cell for the sample. The laser light with a power of 150mw was introduced and focused on the sample, the low power laser was used to ensure that there is no thermal degradation of the sample during analysis, nor will be a problem to be used for the analysis of biological tissue. And the scattered radiation was collected by the detector. All of the Raman spectra recorded in the  $4000 - 400\text{cm}^{-1}$  range using a spectral resolution of  $6\text{cm}^{-1}$  and a total of 128 scans were averaged for each spectrum. A full factorial experimental design was used to determine the most appropriate acquisition parameters, such as the number of accumulations, the acquisition time and the laser power. The OPUS 7.0 (Bruker Optics, Germany) software program was used for Raman spectral data acquisition.

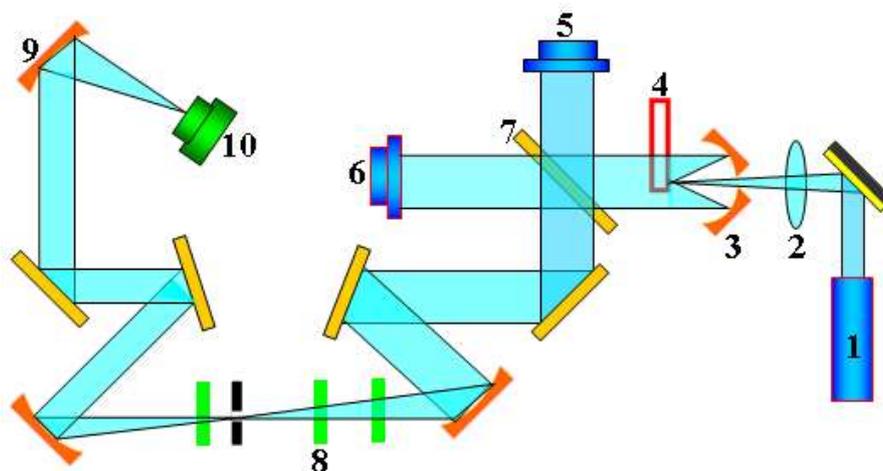


Figure 1. The schematic of the Raman spectroscopy. 1.1064nm near infrared Nd:YAG laser, 2.Lens, 3.Paraboloid converging lens, 4.Sample, 5.Fixed mirror, 6.Moving mirror, 7.Beamsplitter, 8.Light filter, 9.Condensing lens, 10.Ge detector cooled by liquid nitrogen

### **Blood Sample Preparation**

64female (29-78 years old), 52 male (29-96 years old), the 5 mL blood was collected by venipuncture from healthy donors into vacutainer blood collection tubes containing Ethylenediaminetetraacetic acid (EDTA) as an anticoagulant. Blood samples were stored in a  $4^{\circ}\text{C}$  refrigerator immediately after collection until required for measurements or additional separation steps. The range of glucose is 4.12-16.32 mmol/l.

### **Calibration with ANN**

The baseline correction, normalization, smoothing or filtration were used for Raman spectra. After pre-processing, the Raman spectral data and corresponding concentration of nutrients, cells and metabolites were defined for each sample from experiments performed in bioreactors. From this data set, the pattern was established, by the input set, which is described by the Raman spectra in a scan from  $4000$  to  $400\text{cm}^{-1}$ , and the output set was given by glucose concentrations, as shown in Figure 2. The multi-layer perceptron (MLP) and the Levenberg-Marquardt (LM) back-propagation training algorithm are used in the neural network. In the MLP there are three kinds of layers: the input layer which receives the input signals, one or more hidden layers where the processing takes place and the output layer which provides the output. The specific transfer function was used to characterize the neurons in each layer. In our paper, the hidden neurons with non-linear transfer function were used, while the output neurons have a linear or non-linear transfer function. In order to minimize the average error across all training patterns, the number of hidden neurons is chosen experimentally.

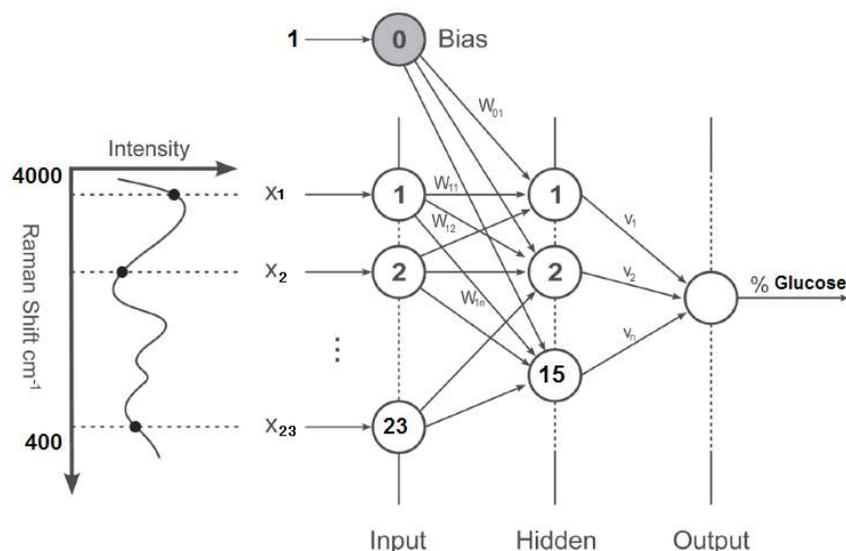


Figure 2. Diagram of the discretization of the Raman spectra and neural network topology used in the calibration of the percentage of glucose concentration. The input number that feeds the input layer is 1 to 23, the number of the hidden layer is 1 to 15. The  $w_{in}$  was used to connect the input layer to the hidden layer, and the  $v_n$  is used to connect the hidden layer to the output layer.

## Results and Discussion

The Raman spectra of whole blood with different glucose concentration were shown in Figure 3. The difference of the spectral profile was caused by the glucose concentration. The Raman spectra were acquired in the range from 4000 to 400  $\text{cm}^{-1}$ . However, only the smaller range (3100-2900  $\text{cm}^{-1}$  and 1800-800  $\text{cm}^{-1}$ ) was used according to the PCA analysis. This approach can eliminate the need of a very large number of nodes in the input layer of the neural network, which will make training time too long. Thus, the information of Raman spectra extracted by the ANN was mapped in the input layer nodes to corresponding to the concentration of the glucose in the output layer.

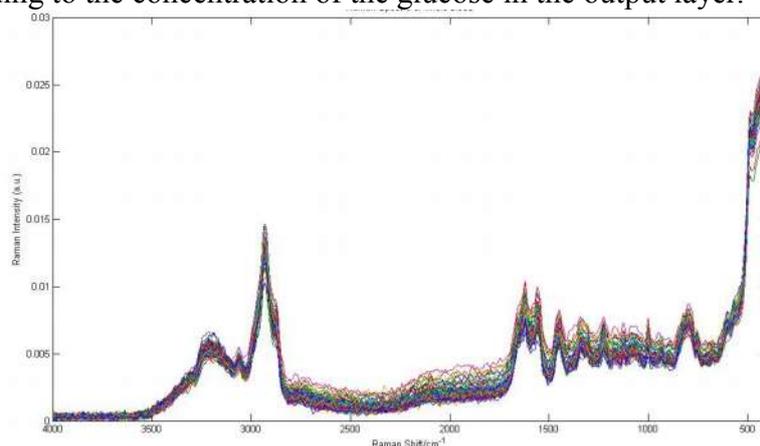


Figure 3. The Raman spectroscopy of whole blood.

The optimal ANN model obtained by the Matlab software after the Raman spectra performed with baseline correction, normalization, smoothing or filtration. The results of all ANN configurations were shown in Table 1. The ANN model with high values of  $R^2$  and low values of RMSE for both training and validation date has good prediction capacity. The ANN configuration with 18 nodes has the highest score and was considered the best predictor. The values of the glucose concentration estimated by the neural network (horizontal axis) and the actual values (vertical axis) were shown in Figure 4. Figure 4a shows the data used to train the network, whereas Figure 4b refers to the data used

for validation. In our experiments, the neural network was able to predict all the values used for the training and validation.

Table 1. Coefficient of determination ( $R^2$ ) values and root mean square error (RMSE) between actual values (known) and predicted for the glucose.

Hidden nodes (no.)	Training		Validation		Score
	$R^2$	RMSE	$R^2$	RMSE	
5	0.9915	0.0579	0.9910	0.0195	2538.958
6	0.9829	0.5367	0.9935	0.386	211.676
7	0.9876	0.275	0.9834	0.0185	661.8158
8	0.9922	0.2905	0.9812	0.0381	592.5789
9	0.9948	0.8841	0.9844	0.3477	159.0088
10	<b>0.9955</b>	<b>0.0323</b>	<b>0.9953</b>	<b>0.0270</b>	<b>3308.318</b>
11	0.9855	0.1639	0.9941	0.1025	735.5274
12	0.9810	0.2809	0.9974	0.18	424.6404
13	0.9822	0.0429	0.9984	0.0475	2169.822
14	0.9842	0.2586	0.8423	0.103	464.0816
15	0.9848	0.0742	0.9921	0.0012	2591.636
16	0.9827	0.3477	0.9365	0.3048	282.411
17	0.9984	0.382	0.9801	0.1718	353.4486
18	0.9917	0.1012	0.9723	0.0406	1360.251
19	0.9866	0.1350	0.9932	0.0621	994.3307
20	0.9843	0.0633	0.9999	0.0058	2848.982

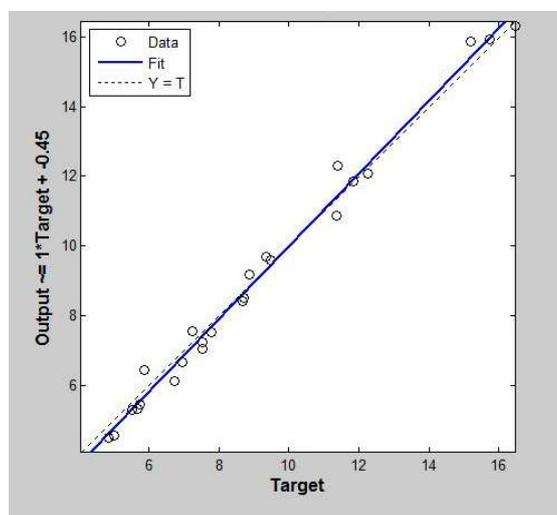


Figure 4. Comparison between the actual glucose concentration and the predicted by a model obtained by artificial neural networks.  $N=23$ , hidden nodes =15, and  $R^2=0.995$ .

## Conclusion

In this paper, the FT-Raman spectroscopy in combination with ANN model was able to estimate critical components of the glucose concentration in the whole blood because of its high speed, simplicity, and efficiency. The pre-processing, such as filtration, baseline correction, normalization, or smoothing procedure, was applied to the spectra to remove the noise. The PCA has been adopted to reduce the number of input variables to the neural network to the minimum number of variables, representing the features selected as the most significant variables for approximating the functional relationship. A high-capacity prediction model with  $R^2=0.9955$  for training and  $R^2=0.9953$  for validation was obtained using ANN. As the result of this experiment, the acceptable values of

trueness and precision were able to achieve. The prediction error of ANN model is low. The results show that the ANN model is an excellent method in the whole blood detection.

## Acknowledgement

This work was partly supported by the National Natural Science Foundation of China (11404054, 31170956), the Natural Science Foundation of Hebei Province (F2014501137, H2015501133), the Fundamental Research Funds for the Central Universities (N120323004, N120223001, N142304003), Science and technology research & development project funds of Shenzhen (JCYJ20120618142137681), the Educational Commission of Hebei Province (Z2013040) and the science and technology planning project of Hebei province (13273303D).

## References

- [1] J. Omar, A. Sarmiento, M. Olivares, I. Alonso, N. Etxebarria, Quantitative analysis of essential oils from rosemary in virgin olive oil using Raman spectroscopy and chemometrics, *J. Raman Spectrosc.*, 43 (2012) 1151-1156.
- [2] Y. Numata, Y. Iida, H. Tanaka, Quantitative analysis of alcohol–water binary solutions using Raman spectroscopy, *J. Quant. Spectrosc. Radiat. Transfer*, 112 (2011) 1043-1049.
- [3] B. Kuang, Y. Tekin, A.M. Mouazen, Comparison between artificial neural network and partial least squares for on-line visible and near infrared spectroscopy measurement of soil organic carbon, pH and clay content, *Soil and Tillage Research*, 146, Part B (2015) 243-252.
- [4] M. Takahashi, J. Leme, C. Caricati, A. Tonso, E. Fernández Núñez, J. Rocha, Artificial neural network associated to UV/Vis spectroscopy for monitoring bioreactions in biopharmaceutical processes, *Bioprocess Biosyst Eng*, 38 (2015) 1045-1054.
- [5] Y. Huan, G. Feng, B. Wang, Y. Ren, Q. Fei, Quantitative analysis of cefalexin based on artificial neural networks combined with modified genetic algorithm using short near-infrared spectroscopy, *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 109 (2013) 308-312.
- [6] B. Özbalci, İ.H. Boyaci, A. Topcu, C. Kadılar, U. Tamer, Rapid analysis of sugars in honey by processing Raman spectrum using chemometric methods and artificial neural networks, *Food Chem.*, 136 1444-1452.
- [7] M.C. Soto-Barajas, M.I. González-Martín, J. Salvador-Esteban, J.M. Hernández-Hierro, V. Moreno-Rodilla, A.M. Vivar-Quintana, I. Revilla, I.L. Ortega, R. Morón-Sancho, B. Curto-Diego, Prediction of the type of milk and degree of ripening in cheeses by means of artificial neural networks with data concerning fatty acids and near infrared spectroscopy, *Talanta*, 116 (2013) 50-55.
- [8] R. Alves da Rocha, I.M. Paiva, V. Anjos, M.A.M. Furtado, M.J.V. Bell, Quantification of whey in fluid milk using confocal Raman microscopy and artificial neural network, *J. Dairy Sci.*, 98 (2015) 3559-3567.
- [9] S. Ramji, G. Latha, S. Ramakrishnan, Estimation and interpolation of underwater low frequency ambient noise spectrum using artificial neural networks, *Applied Acoustics*, 70 (2009) 1111-1115.
- [10] R. Chakraborty, H. Sahu, Intensification of biodiesel production from waste goat tallow using infrared radiation: Process evaluation through response surface methodology and artificial neural network, *Applied Energy*, 114 (2014) 827-836.