Detecting glucose levels by means of Raman spectroscopy, support vector machine and principal component analysis

CASTRO-RAMOS, J.' *, VILLA-MANRÍQUEZ, J. F.', GÓMEZ-GIL, P.', GONZÁLEZ-VIVEROS, N.', NAREA-JIMÉNEZ, F.', SÁNCHEZ-ESCOBAR J. J.''and MUÑOZ-LÓPEZ, J.'

' *Instituto Nacional de Astrofísica Óptica y Electrónica, Luis Enrique Erro 1, Tonantzintla, Puebla, México, C.P. 72000.*
'' *Centro de Enseñanza Técnica Industrial, Nueva Escocia 1885 Fraccionamiento Providencia, Guadalajara, Jalisco, México, C. P. 44620.*

_____

Abstract

In this paper we present the results of automatically detecting the glucose levels in chemical solutions. We prepared a set of different solutions, each with different concentrations of anhydrous glucose, which was increased by 10% with respect to the previous one; these dissolutions were diluted in water. Using the whole Raman spectra as features representing the samples, and principal component analysis with support vector machines as a classification method, we predicted the concentration level, getting an accuracy of classification of 79.16% measured using AUC (Area Under the ROC) Curve.

Glucose Level Solution, Raman Spectroscopy, PCA, SVM

_____

_____

_____

*Correspondance to Author (email: jcastro@inaoep.mx)

Introduction

Diabetes mellitus is an important health problem diagnosed whether glucose level is greater than 126 mg/ml. nowadays almost all commercial methods to detect either the level of glucose or glycosylated haemoglobin in humans are invasive. A huge quantity of devices measuring levels of glucose require a drop of blood, which is obtained through a puncture with a needle; it is painful and in some cases traumatic.

In the other hand, it is possible that people living in developing countries does not check their glucose levels due to cost of clinical analysis are high.Several studies have been reported before, using Raman especterograpy for measuring glucose. For example, Wei Gao et al. (2016), worked on a noninvasive device, which measures sweat metabolites such as glucose, lactate and electrolytes.

Barman et al. (2010), developed a model for monitoring blood glucose using Support Vector Machines and Raman Spectroscopy; the authors showed the nonlinear effects in the relationship between the concentrations of the solute of interest and the mixture Raman spectra. Prominent Raman bands associated with glucose (926, 1302, 1125 cm^{-1}) were mention by Zephania Birech et al. (2017), likewise Jingwei Shao et al (2012), mention 796, 1060, 1125 and 1366 like Raman bands. In the work of Mircea Oroian et al. (2017) Raman spectroscopy is used to detect honey adulterated with fructose, and other substances.

In this paper we present a non-invasive method to quantify the levels of glucose from several levels of concentration. The proposed method is based on Raman spectroscopy, principal component analysis and support vector machine.

It is important to point out that in a previous work Villa-Manríquez et al. (2017) applied a similar method was successfully applied to detect the glycated hemoglobin in vivo, without the need for blood. In such work we showed the capability of the method to distinguish between diabetic and non-diabetic patients.

Raman spectroscopy

The Raman Effect or Raman scattering was discovered by Chandrasekara Venkata (1928), Raman Effect is inelastic; the energy of photons scattered from a sample are either higher or lower than the energy of the incident light. There is a type of vibrational spectroscopy that measures the fundamental vibrational modes of chemical bonds and it is used to generate molecular specific information about a sample Colthup et al (1990), thereby it allows the detection of molecular bonds of glucose.

Principal Component Analysis and Support vector Machine

Support vector machines (SVMs) is a discriminative and supervised method, with associated learning algorithms that analyzes data. It is used for classification and regression analysis which is done finding a separating hyperplane.

Principal component analysis (PCA) is a statistical procedure, used to reduce the dimensionality of a data set. It is formed by a large number of interrelated variables. An orthogonal transformation converts a set of observations into a set of values of linearly uncorrelated variables, called principal components.

Methodology

We prepared the samples by dissolving anhydrous glucose in distilled water. We got a set of samples, each with a different concentration with differences of 10% each.

We used the mass concentration definition, it means the constituent in gr (mass), is divided by the mixture in liter (volume). Samples were put in cuvettes. A test probe was placed at the top of the container to acquire Raman signal. We use the experimental setup shown in Figure 1.

We used PCA, to get principal components which are used as features to input to the classifier. The use of PCA allows to reduce the dimension of the data, without losing relevant information. It is important to point out that the complete Raman spectra was used for calculating the principal components.

The reduced features were used for SVM training, with K-fold cross-validation iterations, that is, data was divided into K subsets (folds). One of the subsets was used as a test set and the rest (K-1 folds) as training set. The cross-validation process was repeated for K iterations, using each of the available subsets of test data, with K = 10.

The error was calculated as the arithmetic mean of the errors of each iteration. The average obtained accuracy was 79.16%, wich was obtained using an SVM with a quadratic kernel, implemented in the statistical programming language R (2016).
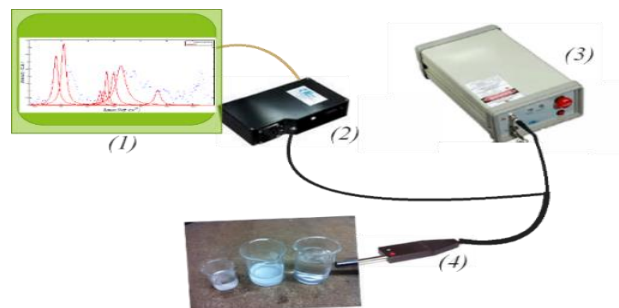

Figure 1. Experimental setup: (1) the computer monitor, (2) the Raman spectrometer, (3) the excitation light source, (4) the sample measurement of Raman spectra.

Results

Figure 2, shows the Raman spectra of samples at different concentrations in a range from 900 to 1200 cm^{-1}.
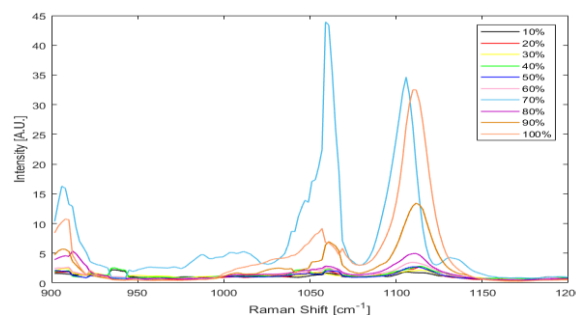

Figure 2. Raman spectra of different solutions with concentrations from 10 to 100 %.

Figure 3 a), shows a map of the decision regions. In that we can differentiate between each concentration, hence it is possible to classify these samples using PCA and SVM. Classification results are shown in figure 3 b). Notice that class labeled "30%" was not well classified (see figure 4 b, the confusion matrix), due to systematic errors in measurements. In future work we will take more measurements to see if this issue is corrected.
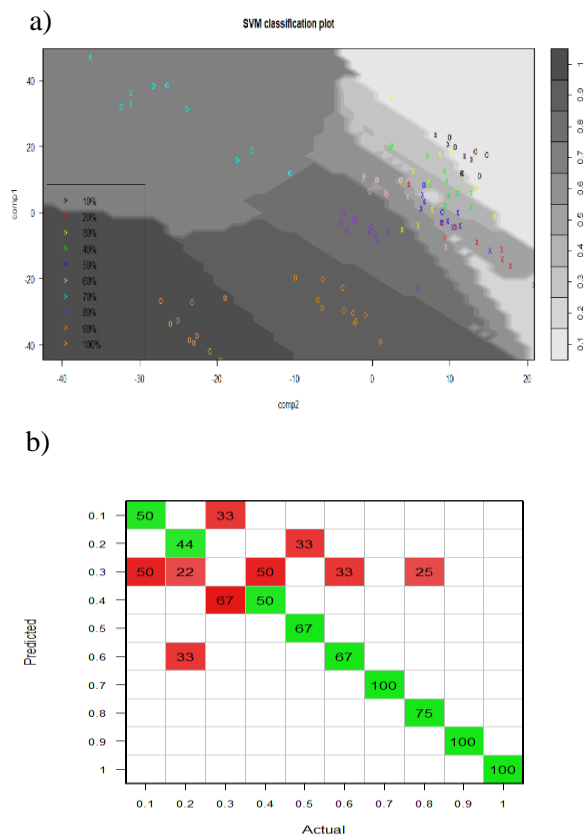
a)

b)

Figure 3. (a) SVM and PCA results, (b) confusion matrix

Conclusions

We analysed the results of a method to quantify glucose in dissolutions, using PCA-SVM, which was tested using k-fold cross validation. We found quadratic kernel results to be a good method to select the optimum level of glucose from the set of trained samples used in this experiment.

Still, more experiments are required, but this results are promising and therefore this method could be applied to know the presence of different concentrations of glucose, considering that in average it was found a classification accuracy of 80%.

Our final goal is to apply automatic classification for measuring glucose in blood.

References

GAO, W., Emaminejad, S., Hnin Yin Yin Nyein, Challa, S., Chen, K., Peck, A., Fahad, H. M., Ota, H., Shiraki, H., Kiriya, D., Lien, D.-H., Brooks, G. A. (2016). Fully integrated wearable sensor arrays for multiplexed in situ perspiration analysis. *Nature* 529, pp. 509–514

Barman, I., Kong, C.-R., Dingari, N. C., Dasari, R. R. and Feld, M. S. (2010). Development of Robust Calibration Models Using Support Vector Machines for Spectroscopic Monitoring of Blood Glucose. *Analytical Chemistry* **82**(23), pp. 9719–9726.

Birech, Z., Mwangi, P. W., Bukachi, F., & Mandela, K. M. (2017). Application of Raman spectroscopy in type 2 diabetes screening in blood using leucine and isoleucine amino-acids as biomarkers and in comparative anti-diabetic drugs efficacy studies. *PLoS ONE*, *12*(9), 1–12. https://doi.org/10.1371/journal.pone.0185130

Shao, J., Lin, M., Li, Y., Li, X., Liu, J., Liang, J., & Yao, H. (2012). In Vivo Blood Glucose Quantification Using Raman Spectroscopy. *PLoS ONE*, *7*(10), 1–6. https://doi.org/10.1371/journal.pone.0048127.

Oroian, M., Ropciuc, S., & Paduret, S. (2017). Honey Adulteration Detection Using Raman Spectroscopy, (July).

Villa-Manríquez, J. F. Castro-Ramos, J., Gutiérrez-Delgado, F., Lopéz-Pacheco, M. A. and Villanueva-Luna, A. E. (2017). Raman spectroscopy and PCA-SVM as a non-invasive diagnostic tool to identify and classify qualitatively glycated hemoglobin levels in vivo. *Journal of Biophotonics* 10(8), pp. 1074-1079.

CASTRO-RAMOS, J., VILLA-MANRÍQUEZ, J. F., GÓMEZ-GIL, P., GONZÁLEZ-VIVEROS, N., NAREA-JIMÉNEZ, F., SÁNCHEZ-ESCOBAR J. J. and MUÑOZ-LÓPEZ, J. Detecting glucose levels by means of Raman spectroscopy, support vector machine and principal component analysis. ECORFAN Journal Taiwan. 2017

C. V. Raman & K. S. Krishnan, A New Type of Secondary Radiation, (1928), Nature 121, 501–502.

Colthup, N. B. Daly, L. H., Wiberley, S. E. (1990). Introduction to infrared and Raman spectroscopy. Academic Press, Inc.R Core Team (2016). R: A language and environment for statistical computing. R

Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/