The use of Artificial Neural Network for lipid and glycaemic profiles quantification through infrared spectroscopy

Henrique Hesse[1], Rejane Frozza[1], Valeriano Corbellini[2], Cézane Reuter[2], Miria Burgos[2]

[1]Mestrado em Sistemas e Processos Industriais – Universidade de Santa Cruz do Sul (UNISC) – Santa Cruz do Sul, RS – Brazil

[2]Mestrado em Promoção a Saúde – Universidade de Santa Cruz do Sul (UNISC) – Santa Cruz do Sul, RS - Brazil

henriquehesse@me.com, frozza@unisc.br, valer@unisc.br, cpreuter@hotmail.com, mburgos@unisc.br

Abstract. *This paper aims to look at the viability of the use of artificial neural networks to solve nonlinear correlations between infrared spectra and biochemical quantification tests, to build a computational system to predict the levels of glycaemic and lipid profiles using infrared spectroscopy. The studies of one of the parameters was modelled and showed signs of viability to quantify all parameters with the suggested methodology. Therefore, more complex and larger data sets are going to be tested with this technique.*

1. Introduction

The monitoring of the glycaemic and lipid profiles are important, since the human vascular system plays a variety of roles, mainly providing nutrients and oxygen to tissues and cells and removing waste from the human system. Basically, each blood component can be quantified, through biochemical analysis. Thresholds are useful to diagnostics and to monitor the subjects' health. Despite the fact that there is equipment to do such analysis, they still rely on the usage of chemicals and they need at least 10 mL of blood to perform glycaemic and lipid analysis [Low-Yinga et al., 2002].

Infrared spectroscopy is a technique that can be used for most organic liquids, solids and gases. Providing information related to the atoms and its bonds, revealing what type of organic function it is and indirectly its quantification [Barbosa, 2007]. This method does not require any chemicals to be executed and the sample's volume is about 5 μL. Both of these factors are imperative when another factor is introduced to the equation of analysis routines, the analysis time [Irudayaraj, 2002]. For infrared technique the analysis time is less than a minute, covering all substances, while the standard biochemical quantification has to be done for each parameter one at the time, with specific reactions [Jessen et al., 2014; Song, Lee e Kim, 2015].

One infrared spectrum from 5 μL of blood provides a great amount of information, since each wavelength from the spectrum is a variable. The mid infrared works between 4,000 and 400 cm-1, therefore every spectrum generates 3,500 variables [Mohd et al., 2015]. Normally the quantification using infrared data is done through multivariate analysis, providing good results for linear variables. The blood spectra are not linear to the amount of cholesterol, HDL (high-density lipoprotein), LDL (low-density lipoprotein), triglycerides and glucose. Hence the non-linear information requires more advanced set of tools for its quantification [Yadav et al., 2015; Shahin, 2014].

An artificial neural network (ANN) is capable to resolve non-linear problems due its different functions. The ANN is built over a group of sample that the values for each parameter of the glycaemic and lipid profiles were already measured. Once the network topology was established, the neural network is trained to reach the minimum error associated with the real values. After the training the ANN can ready new samples and quantify all parameter at once and absorb the information of this new sample to increase its precision [Lima, Pinheiro & Santos, 2014; Piotrowski *et al*., 2015].

2. Experimental

For the analyses were collected 92 blood samples from subjects in Santa Cruz do Sul, Rio Grande do Sul, Brazil, with open consent and well informed about the research, which is registered in the ethics committee in research with human beings of Universidade de Santa Cruz do Sul (reg. number 469916).

The mid-infrared spectra of the total peripheral blood samples were analysed fresh or lyophilized and acquired in real triplicate using diffuse reflectance accessory (DRIFTS) using a spectrophotometer Perkin Elmer (Spectrum 400) with resolution of 4 cm^{-1} and 16 scans.

For the first trial was chosen the total cholesterol, due its importance. The total cholesterol gives a general idea of the ratio between the HDL and LDL. In fact most of the quantification of the LDL is done by the equation of Friedewald, which uses the amount of total cholesterol, HDL and triglyceride to calculate the LDL amount.

The infrared spectrum of the cholesterol in blood samples are hard to distinguish, because of the densitiy variations are very distinct and not very perceptible. The others parameters from the profiles are more easily correlated to certain organic functional groups.

It was used a matrix with size of 95x2331 and a vector of 95x1 in a feed-forward backpropagation ANN type, configured with four layers containing 10, 5, 20 and 1 neurons respectively containing 94 input attributes (due the dimensionality reduction using principal component analysis algorithm, PCA) and 1 output (total cholesterol) (Figure 1). The MATLAB 2010 (Mathworks Inc., Natick, MA) software was used for the implementation of PCA and RNA.

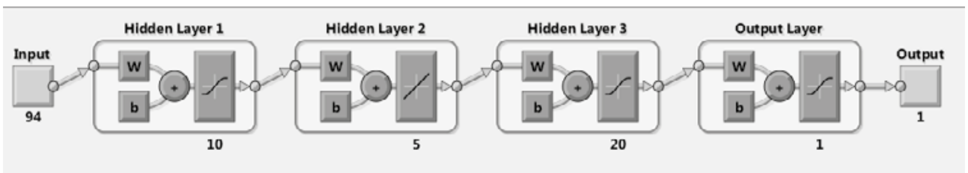

Figure 1. Topology for the ANN.

3. Preliminary results

The initial results of ANN showed a good response to nonlinearity of the data set, achieving an overall correlation coefficient (r) of 0.80985 for this preliminary model and the root-mean-square error (RMSE) of 30.14, (Figure 2).

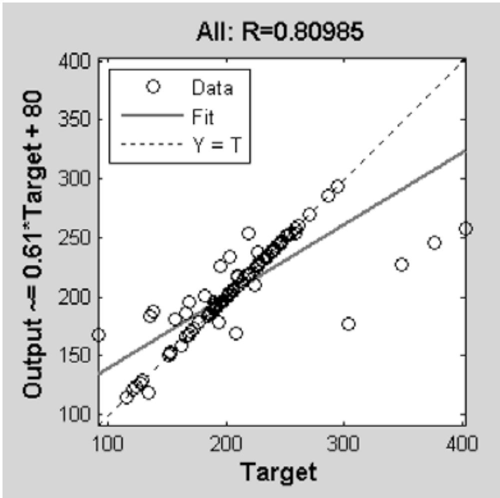

Figure 2. ANN overall regression, showing the correlation between all samples

As can be seen on Figure 2, some samples with higher concentrations of total cholesterol are far from the trending line. Usually those samples would be considered outliers, although it was decided to keep all the samples. Based on the fact that more samples are going to be added to this ANN, making higher the chances to the weight of those samples to be recalculated.

The results for all stages of the ANN are shown in Figure 3. It can be noticed that in all steps as in, training, validation and test the outputs were close enough to the target values to proceed with this research and enlarge the training data set to cover all the cholesterol levels and rise the correlation coefficient to above 95% of correlation between outputs and targets.

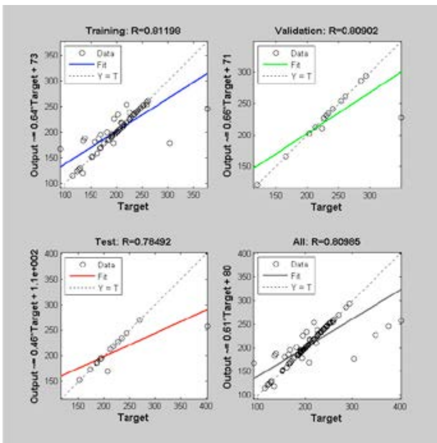

Figure 3. ANN regression, displaying all stages of the network usage

4. Conclusion

The first study of one of the components from the glycaemic and lipid profiles, showed that the artificial neural network was capable to solve the nonlinearity concerning the infrared spectra and the biochemical levels. From this stage of the research, it is the intention to build a full neural network with the entire components from the profiles with a larger data set (500 samples). A larger group of samples will provide more information for every level of each parameter (e.g. total cholesterol, HDL, LDL, triglycerides and glucose).

References

Barbosa, Luiz Cláudio de Almeida. *Espectroscopia no infravermelho na caracterização de compostos orgânicos*. Viçosa: Universidade Federal de Viçosa, 2007.

Irudayaraj, J. (2002). Rapid determination of vitamin C by NIR, MIR and FT-Raman techniques. J. *Pharm. Pharmacol*., London, v.54, n.9, p.1247-1255, 2002.

Jessen, T. E., Hoskuldsson, A. T., Bjerrum, P. J., Verder, H., Sorensen, L., Bratholm, P. S., . . . Jensen, M. A. (2014). Simultaneous determination of glucose, triglycerides, urea, cholesterol, albumin and total protein in human plasma by Fourier transform infrared spectroscopy: direct clinical biochemistry without reagents. *Clin Biochem, 47*(13-14), 1306-1312. doi: 10.1016/j.clinbiochem.2014.05.064

Karmowski, K. A. S., Jacek Majda, Piotr Rubisz, Stanisław Han e Andrzej Karmowski. (2014). Lipid index changes in the blood serum of patients with hyperplastic and early neoplastic lesions in the ovaries. *Journal of Ovarian Research, 7*(1), 1-4.

LIMA, ISAÍAS. PINHEIRO, A. M. CARLOS, SANTOS, A. O. FLÁVIA. *Inteliência artificial*. 1. ed. Rio de Janeiro: Elsevier, 2014.

Low-Yinga, S., R. Anthony Shawa, Michael Lerouxb, Henry H. Mantscha. (2002). Quantitation of glucose and urea in whole blood by mid-infrared spectroscopy of dry films. *Vibrational Spectroscopy, 28*, 111-116.

Mohd Ali, J., Hussain, M. A., Tade, M. O., & Zhang, J. (2015). Artificial Intelligence techniques applied as estimator in chemical process systems – A literature survey. *Expert Systems with Applications*. doi: 10.1016/j.eswa.2015.03.023

Piotrowski, A. P., Napiorkowski, M. J., Napiorkowski, J. J., & Osuch, M. (2015). Comparing various artificial neural network types for water temperature prediction in rivers. *Journal of Hydrology, 529*, 302-315.

Song, S. Y., Lee, Y. K., & Kim, I. J. (2016). Sugar and acid content of Citrus prediction modeling using FT-IR fingerprinting in combination with multivariate statistical analysis. *Food Chem, 190*, 1027-1032. doi: 10.1016/j.foodchem.2015.06.068

Shahin, M. A. (2014). State-of-the-art review of some artificial intelligence applications in pile foundations. *Geoscience Frontiers*. doi: 10.1016/j.gsf.2014.10.002

Yadav, J., Rani, A., Singh, V., & Murari, B. M. (2015). Prospects and limitations of non-invasive blood glucose monitoring using near-infrared spectroscopy. *Biomedical Signal Processing and Control, 18*, 214-227.