01

Influence of data pre-processing techniques for PLSR model to predict blood glucose by NIR spectroscopy*

© *S.Vasanthadev Suryakala, Shanthi Prince*

Department of Electronics and Communication Engineering,
SRM Institute of Science and technology,
Kattankulathur-603203, Tamil Nadu, India

e-mail: suryakas@srmist.edu.in, shanthip@srmist.edu.in

NIR diffuse reflectance spectroscopic spectra can be mathematically modelled to extract quantitative information by suitable multivariate calibration models. The analysis of spectral data becomes complex as the data is more prone to noise due to light scattering and baseline effects. These errors reduces the robustness and reliability of the developed calibration model. Hence data pre-processing becomes the most important aspect in data analysis. Different mathematical transformations are applied to remove the noise present in the data. This work focuses on the various empirical data pre-processing techniques like baseline correction, multiplicative scatter correction (MSC), robust MSC, extended multiplicative signal correction (EMSC), orthogonal signal correction (OSC) and ($-\log R$) followed by standard normal variate (SNV) techniques for Partial Least Square Regression (PLSR) model in the prediction of blood glucose non-invasively. The performance of the PLSR model for the acquired (raw) spectral data and the same data subjected to different pre-processing techniques is analyzed. The model complexity and robustness is evaluated in terms of the number of latent variables (LVs) required to build the calibration model and obtained mean square prediction error after cross validation. This study utilizes the spectral data collected from 207 subjects from a diabetic center using Diffuse Reflectance Spectrometer (DRS). The analyzed results show that pre-processing based on ($-\log R$) followed by SNV is found to perform well with reduced model complexity and minimum estimated mean square prediction error of 0.23 mg/dl among the other empirical pre-processing techniques.

Keywords: multiplicative scatter correction (MSC), orthogonal signal correction (OSC), standard normal variate (SNV), Diffuse Reflectance Spectrometer (DRS).