Interpretation of Raman spectra of biological media based on deep learning

Lyudmila A. Bratchenko[1*], Elena N. Tupikova[2], Sahar Al-Sammarraie[1], Valery P. Zakharov[1], Ivan A. Bratchenko[1]

1. Department of Laser and Biotechnical Systems, Samara University, 34 Moskovskoe Shosse, Samara, 443086, Russia
2. Department of Chemistry, Samara University, 34 Moskovskoe Shosse, Samara, 443086, Russia

ABSTRACT

The study of the metabolic profile of the human body tissues can provide information about the state of the body and possible pathological-associated changes in it. The analysis of the structural features and molecular vibrations in biological media is of certain interest in the goals of metabolomics. Various analytical methods based on Raman techniques are now widely used to analyse the structural features of biological media. It should be noted that the Raman spectra are characterized by multicollinearity and contain noise, Raman shift drift, and the contribution of extraneous sources of spectral noise. Obtaining statistically reliable results when analyzing the full Raman spectrum is possible using machine learning methods. In this work, the Raman spectra of the skin, serum and blood plasma, as well as individual metabolites are analyzed. The experimental setup for registration of Raman characteristics includes a spectrometric system (EnSpectr R785, Spektr-M, Chernogolovka, Russia) and a microscope (ADF U300, ADF, China). Focusing the exciting radiation and collecting the scattered radiation were implemented using 50x Objective LMPlan. The stimulation of collected spectra was performed by the laser module with central wavelength 785 nm. A nonlinear method for recognition of Raman spectra of biological media in the range of 500-1900 cm^{-1} under the conditions of a nonlinear dependence of the spectral characteristics of the medium on the features of its composition based on a one-dimensional convolutional neural network is proposed. The proposed method with the involvement of deep learning makes it possible to increase the accuracy of spectral analysis for determining the component features of biological media with multiple overlaps of the spectral bands of analytes by 7–13% compared to the bilinear method without involving deep learning. The possibility of identifying the features of the composition of polycomponent colloidal media and associated spectral features using the distribution of the importance of variables in constructing a model based on the permutation algorithm is established. The application of the proposed method of distribution of the importance of variables ensured the determination of the spectral contribution of creatinine in the bands 630-650 cm^{-1}, 720-750 cm^{-1}, 1380-1415 cm^{-1} and the spectral contribution of urea in the bands 720-750 cm^{-1}, 990-1030 cm^{-1}.