Software for Comprehensive Batch Analysis of Raman/SERS Spectra with Real-time Interactive Visualization and Machine Learning Predictions

Venkat Suprabath Bitra *

International Institute of Information Technology Bangalore, Electronic City, Bengaluru-560100, India

** Email: venkat.suprabath@iiitb.ac.in*

Raman spectra provide characteristic fingerprint of various organic molecules. Manifold enhancement of Raman signal in presence of gold or silver nanoparticles, known as Surface Enhanced Raman scattering (SERS) technique makes it highly useful for various analytical applications. At present, one of the research interests is integration of SERS/Raman technique with different machine learning (ML) and deep learning (DL) strategies for developing various real life applications [1,2]. These studies involve processing (i.e., training, validation and testing) of large number of Raman spectra having different level of intensities, noise etc., for ensuring reliable prediction of compounds. In this respect, a software for quick and comprehensive analysis with dynamic interactive visualisation of Raman spectra obtained from different instruments along with user selectable pre-implimented ML/DL methods, without need of additional programming is currently not available.

In view of this, a versatile and user-friendly Raman spectra analysis software tool for simultaneous processing of large number of spectra in single batch (hundreds of spectra) with real-time interactive display using advanced plotting tools has been developed. This software is developed with help of different open-source software tools and consists of two main parts, frontend and backend. The frontend is written in Typescript and built using React, Recoil, React Router, Material UI, Recharts and Apache eCharts. The backend is written in Python, uses Numpy, Pandas, Scipy, Scikit-learn, Flask and Flask-RESTful. This tool can perform various data processing operations: 1) automatic background removal with user defined parameters, 2) spectral noise reduction, 3) auto peak identification with data table (database) saving, 4) auto plot signal variation vs analyte concentration, 5) heat-maps or bar charts of all peaks for spatial intensity mapping, 6) spectra comparison with normalization techniques (standard normal variate, min-max normalization), 7) ratio-metric analysis for quantification, 8) auto identification of compound from user customized database library, 9) peaks area analysis, 10) 2D and 3D multi-plots, 11) principal component analysis (PCA) with different data normalization techniques with interactive 2D and 3D plots, 12) compound classification with different ML models and 13) Prediction table of test compounds. For example, Fig. 1(a-d) presents four different typical graphs obtained using the developed tool. In a single webpage, the software displays multiple graphs and data tables interactively within few mouse clicks and user can select various parameters to process the spectra. In all plots, mouse over on lines or data points displays the corresponding file name. This tool significantly decreases workflow execution time for analysis of Raman spectra from few weeks to few hours and can play important role for developing various societal applications based on Raman spectra. Typically, for hundred spectra, all above mentioned analysis can be performed within an hour. The developed software is in use for ML/DL based SERS studies at Laser Materials Processing Division, RRCAT, Indore. The tool is built with different algorithms, and can be further upgraded for more advanced upcoming ML/DL methods.

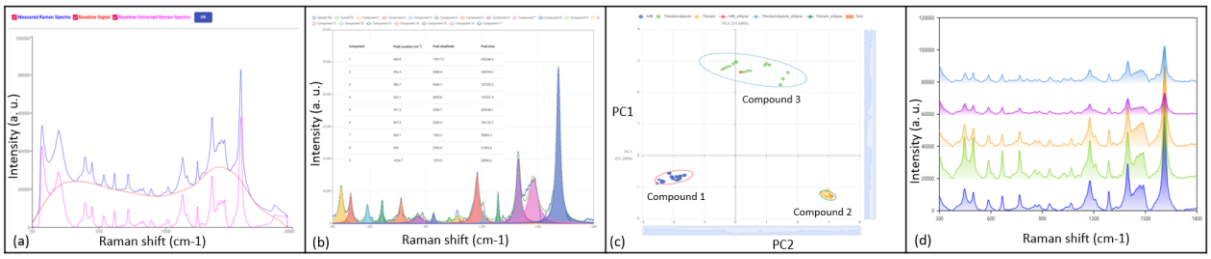

Fig. 1: (a) Background removal, (b) Peak area determination, (c) PCA analysis and (d) 2D multi plot.

In conclusion, the developed software tool processes large number of Raman spectra in various aspects within few mouse clicks. It can process spectral data of all spectrometers and useful for both R&D and industrial users. The tool has capability for compounds identification by classical means of spectra comparison as well as advanced methods based on machine learning algorithms. The software is useful for understanding molecular adsorption driven spectral variation using different ML and DL methods and their performance comparison.

References

[1] F. Lussier, V. Thibault *et al*. "Deep learning and artificial intelligence methods for Raman and surface-enhanced Raman scattering," TrAC Trends in Analytical Chemistry **124**, 115796 (2020).

[2] Y. X. Leong, et al, "Surface-Enhanced Raman Scattering (SERS) Taster: A Machine-Learning-Driven Multireceptor Platform for Multiplex Profiling of Wine Flavors," Nano Letters **21**, 6, 2642–2649 (2021).