MOL2NET, **2022**, 8, **ISSN**: 2624-5078 **1**
https://mol2net-08.sciforum.net/

CHEMBIOINFO-08: Cheminfo., Chemom., Bioinfo., Comput. & Quantum Chem. Congress München, GR-Cambridge, UK-Ch. Hill, USA, 2022

The application of Machine Learning to Raman spectroscopy

Andrea Ruiz Escudero
Department of Organic and Inorganic Chemistry, Faculty of Science and Technology, University of the Basque Country.
IKERDATA S.L INVESTIGO Next-generation Researcher.

Abstract
In analytical science, data extraction from complex or extensive datasets can be a laborious task that takes a long time to complete. Machine learning (ML) offers a pioneering opportunity to rapidly extract information from chromatography, spectroscopy, and mass spectrometry datasets, among others. Over the past few years, new approaches have been developed for the rapid processing of Raman spectra using ML. This review will discuss different applications of ML techniques employed in Raman.

Raman spectroscopy is a technique based on the inelastic scattering of monochromatic light that allows the unequivocal characterization of a material. It is used in a high variety of applications including chemical synthesis, crystallization processes and polymeric reactions. The application of machine learning techniques to automatically extract underlying knowledge in the data provides a good opportunity to develop accurate, fast and inexpensive methods in various areas. Next, the application of ML and Raman-based techniques in three different fields will be discussed.

Within the food and nutrition area of expertise, Berghian-Grosan et al. developed a new method for edible oil authentication.[1] For this approach, they analyzed seven types of less common edible oils by ML and Raman to obtain an accurate and rapid detection of any adulteration together with an estimation of the impurity degree. They also demonstrated that some of the oils were incorrectly labelled claiming a lower proportion of the cheaper ingredient.

Another area where Raman and machine learning have been combined is microbiology. An example of this is a method created by Yang et al. for the differentiation and classification of bacterial endotoxins.[2] Here, by using silver nanorod array substrates, they acquired the surface-enhanced Raman scattering (SERS) spectra of eleven bacterial endotoxins. Furthermore, by combining the identified characteristic SERS peaks with different classical ML and deep learning (DL) algorithms, they classified the endotoxins with high accuracy and precision.

Lastly, in the medical field, Guevara et al. applied machine learning tools and Raman spectroscopy to screen type 2 diabetes mellitus (DM2).[3] In this work, they compared principal component analysis (PCA) and artificial neural networks (ANN) performances in discerning between DM2 and control groups. The ANN-based method showed a greater classification accuracy, demonstrating a high potential for the precise and rapid screening of diabetic disease in the general population.

In conclusion, the use of machine learning in analytical techniques, such as Raman, can offer the opportunity to create rapid and accurate methods, while saving time and resources during their development.

References

1. Berghian-Grosan, C., & Magdas, D. A, Raman spectroscopy and machine-learning for edible oils evaluation. *Talanta*, **2020**, vol. 218, p. 121176.

2. Yang, Y., Xu, B., Haverstick, J., Ibtehaz, N., Muszyński, A., Chen, X. et al., Differentiation and classification of bacterial endotoxins based on surface enhanced Raman scattering and advanced machine learning. *Nanoscale*, **2022**, *14*, 8806-8817.

3. Guevara, E., Torres-Galván, J. C., Ramírez-Elías, M. G., Luevano-Contreras, C., & González, F. J., Use of Raman spectroscopy to screen diabetes mellitus with machine learning tools. *Biomedical Optics Express*, **2018**, 9, p. 4998-5010.