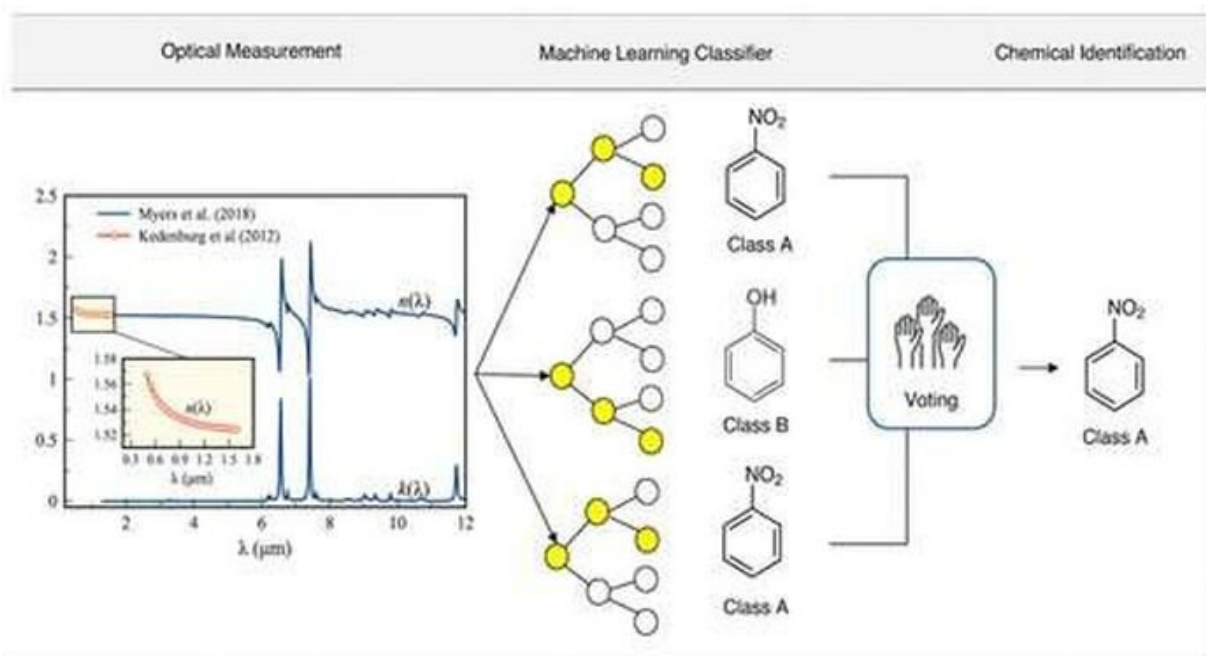# Identifying organic compounds with visible light

March 17 2023, by Justin Jackson

Graphical abstract. Credit: *The Journal of Physical Chemistry A* (2023). DOI: 10.1021/acs.jpca.2c07955

Researchers from the Universidad de Santiago de Chile and the University of Notre Dame, working with machine learning, have devised a method to identify organic compounds based on the refractive index at a single optical wavelength. The technique could have research and industrial applications for automated chemical analysis that is cheaper,

safer and requires less expertise to operate.

In the paper, "Machine learning identification of organic compounds using visible light," published in The *Journal of Physical Chemistry A*, the researchers document the creative and novel way in which they acquired a unique data set and the steps they used to build a proof of concept organic chemistry detector.

Machine learning was trained on a publicly available database of past optical experiments with published data from scientific literature dating back to 1940. In this database, researchers found all the parameters needed to compile identification profiles for 61 organic molecules; group velocity and group velocity dispersion, the measurement wavelength range and the state of matter of the samples, refractive indexes and extinction coefficients over a wide range of wavelengths. In all, 194,816 spectral records of refractive index and extinction curves of the 61 organic compounds and polymers were applied.

In a typical infrared (IR) molecular classification detector, molecule identity is confirmed by absorption and Raman scattering peaks, creating a fingerprint of combined features matched to a database. The static refractive index of organic compounds is a single-valued feature that does not have the same encoded information. The same applies to refractive index databases at single wavelengths away from the ultraviolet and infrared absorption resonances, which is perhaps why visible light has not been used to classify organic molecules.

Initial testing with raw data reached 80%, and the researchers attempted to increase it from there. The original database was not intended for optimizing machine learning as much of it came from research conducted before the first home computer had been invented. There was a tremendous amount of information on wavelengths in the UV and IR range, which the AI was cross-training on. So, the researchers decided to

take a more focused approach.

Several data preprocessing strategies were employed to simulate a more idealized learning environment for the AI. The goal was to create a balanced data set so that the AI did not preferentially give weight to certain features over others just by the volume of information. Oversampling and undersampling and data physical-based augmentation techniques were used to essentially reduce the impact of IR wavelengths in the overall data set. By training with preprocessed balanced data, the researchers achieved molecular classification testing accuracies in the visible regions better than 98%.

The researchers state that additional work is needed to expand and generalize the classifier to identify the structural and other chemical features of the molecules that are present in the Refractive Index Database. In summary, they write that the work is a good starting point for developing remote chemical sensors.

  More information: Thulasi Bikku et al, Machine Learning Identification of Organic Compounds Using Visible Light, *The Journal of Physical Chemistry A* (2023). DOI: 10.1021/acs.jpca.2c07955

© 2023 Science X Network

Citation: Identifying organic compounds with visible light (2023, March 17) retrieved 10 May 2024 from https://phys.org/news/2023-03-compounds-visible.html