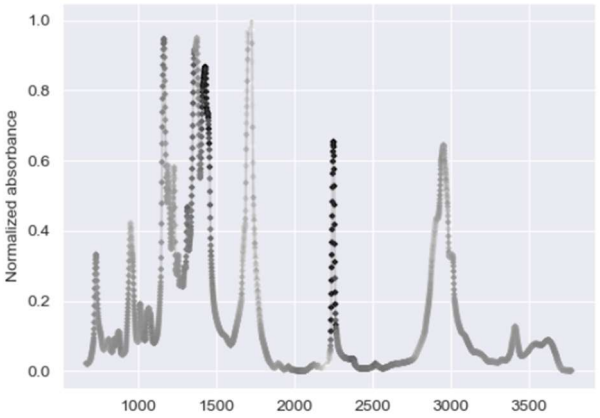
Explainable Artificial Intelligence in physical chemistry. SHAP model explaining a neural network detecting functional groups from FTIR spectra

Tomasz Urbańczyk[1,#], Jakub Bożek[2], Jarosław Koperski[1], Marek Krośnicki[2]

[1]Jagiellonian University, Smoluchowski Institute of Physics, Łojasiewicza 11, 30-348 Krakow
[2]University of Gdansk, Institute of Theoretical Physics and Astrophysics, Wita Stwosza 57, 80-308 Gdansk

In paper [1], a deep learning model for predicting the presence of selected functional groups based on FTIR spectra is described. In the model, each group is predicted by an independent binary classifier using a convolutional neural network with Kolmogorov–Arnold layers (CNN-KAN model). A drawback of deep models is that they operate as a "black boxes," making it unclear on what premises a specific decision is made. The ability to understand these premises is important, as it may increase user confidence in the correct operation of the deep model. Therefore, using explainable AI techniques, we developed an interpretable model utilizing Shapley values to explain the decisions of deep classifiers [2].



The result of an explainable model on a spectrum of molecule containing the group of nitriles. Dark points indicate areas of the spectrum that, according to an explainable model, are especially important for the decision process.

The figure below presents the result of the explainable model for a spectrum of molecule containing the nitrile group. Regions of the spectrum that contribute to the CNN-KAN model's positive decision are marked by darker points. The strong peak, highlighted with dark points in the 2200–2280 cm^{-1} region (associated with C≡N bonding), was correctly interpreted as evidence that the spectrum originates from a molecule containing a nitrile group.

[1] T. Urbańczyk et al. (2025) Chemom. Intell. Lab. Syst., 263, 105421
[2] T. Urbańczyk et al. (2026) Int. J. Mol. Sci., 27, 2004

corresponding author: tomek.urbanczyk@uj.edu.pl