Automated Multi-Element composition analysis of X-Ray Fluorescence Spectra via Vision Transformers

Francesco La Porta[1], Anass Bellachehab[1], Olga Roudenko[1], Emmanuel Farhi[1]
[1]Soleil Synchrotron – L'Orme des Merisiers, 91190 Saint-Aubin, France

francesco.la-porta@synchrotron-soleil.fr

Modern synchrotron facilities are increasingly adopting Artificial Intelligence (AI) to manage the vast amounts of experimental data they generate [1]. AI not only helps to process this data efficiently but can also provide users with preliminary insights to guide ongoing experiments.
This project aims to develop algorithms that transform raw data into immediate results under a wide variety of experimental conditions. X-ray fluorescence (XRF) spectroscopy is used as the initial case study due to its extensive use at the facility and the relative simplicity of its simulation. The primary technical goal is to develop an AI model capable of determining elemental stoichiometry directly from XRF spectra.
A key limitation is that deep learning models require large, labeled datasets, which are often impractical to obtain experimentally at synchrotron. To address this, the project uses simulated datasets for initial model training. Simulations are performed with McXtrace, a Montecarlo ray-tracing software specifically designed for modeling synchrotron X-ray experiments[2]. For this study, a comprehensive dataset of 44’500 spectra was generated, covering 74 elements with atomic numbers between 19 (Potassium) and 92 (Uranium). The simulations were performed with different incident photon energies (from 3.5 to 35 keV) and with different detector resolutions. This ensures that the AI model can generalize more effectively across vary experimental conditions.
Various model architectures were tested. Convolutional Neural Networks (CNNs), commonly used in spectroscopy tasks, struggled with this XRF problem and it could not effectively predict the stoichiometry across multiple elements simultaneously.
Success came with the Vision Transformer (ViT)[3], an architecture adapted from large language models. In ViT, XRF spectra are divided into small sections called "patches" treated like words in a sentence. The Multi-Head Attention (MHA) mechanism allows the model to focus on the most informative spectral regions, efficiently capturing complex patterns. The resulting ViT model, with nearly 110,000 trainable parameters, achieved an R^2 (persentage of variance explained by the model) of 89.1% on test data, and a small mean squared error, demonstrating high predictive accuracy. Figure 1 shows an example of inference of stoichiometry from a simulated spectrum, Figure 2 shows the averaged R^2 for each element of the test set. Beyond accuracy, the ViT model is highly flexible, capable of handling different detector energy ranges, incident photon energies, and spectral resolutions. Its inference speed reaches 400,000 spectra per second.
To bridge the gap between simulation and real-world measurements, fine-tuning the pre-trained model will be applied on small experimental datasets with labelled data. In this approach, most model layers (pre-trained on simulated data) are blocked, and they act as feature extractors, while only the final layers are retrained. Literature shows that even a few experimental spectra can significantly improve the model performances[4].
A complementary direction of the project is to apply transfer learning on the pre-trained model on XRF, to predict some metrics of other techniques (e.g. Raman or IR spectroscopy). This method can be particularly powerful when the dataset of the second technique is limited[5].
Ultimately, the project aims to deliver a flexible, reliable AI tool for real-time analysis, enhancing both efficiency and scientific output across the Synchrotron facilities. Although this model has been implemented for XRF, it can be readily extended to other techniques, both at synchrotron beamlines and in conventional laboratory settings.

References

[1] Wang et al., Small, 14.46 (2018): 1802291.
[2] Bergbäck Knudsen et al., Applied Crystallography 46.3 (2013): 679-696.
[3] Dosovitskiy et al., arXiv:2010.11929 (2020).
[4] Koblischke and Bovy., *arXiv:2411.04750* (2024).
[5] Chen et al., Digital Discovery 3.2 (2024): 369-380.

Figures

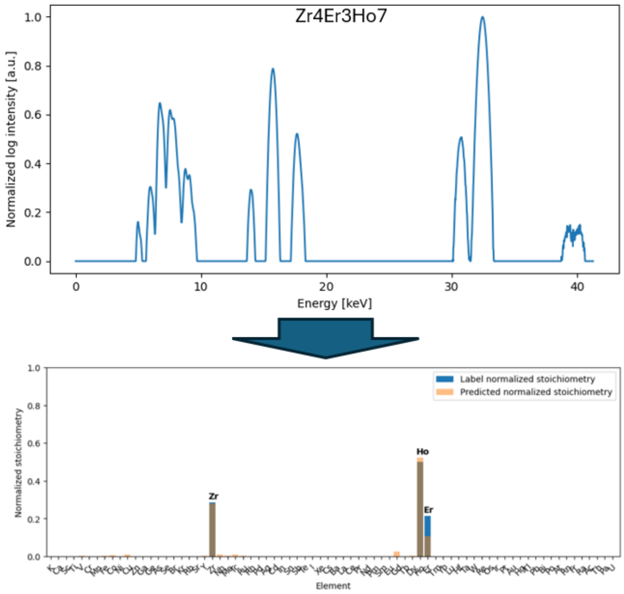

Figure 1. Example of inference of the trained ViT model. The model receives a simulated XRF spectra as input. The spectrum is in logarithmic scale and normalized between 0 and 1, as shown in the upper panel. As output, ViT predicts a concentration vector of all the 74 elements. In the lower panel, the normalized stoichiometry used to simulate the spectrum (serving as ground-truth label) is shown in blue, while the stoichiometry predicted by the model in semi-transparent orange. The overlap between the two bars (blue and orange) results in brown.

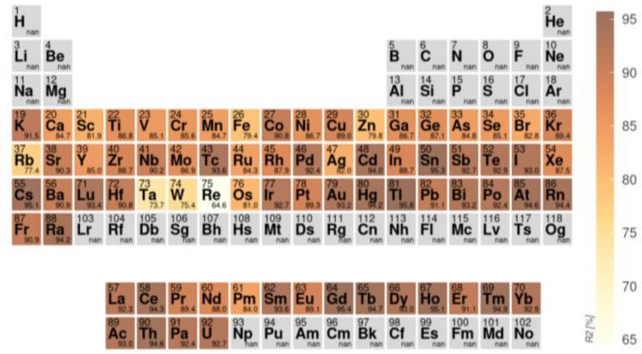

Figure 2. The average R^2 value for each element was calculated from the model inference on the test dataset. The figure shows that the model achieves excellent agreement with the true stoichiometry across all elements, indicating robust predictive performance.