Machine Learning-Based X-Ray Diffraction Analysis for Nanostructure Characterization

Nguyen Khac Huy
MSc Applied Mathematics / ELTE

Semester 2025/26 I Report

1 Introduction and Problem Statement

Modern materials research increasingly relies on combinatorial approaches, where a large number of different material compositions are synthesized and tested within a single sample [1]. To characterize the structure of such compositionally varying materials, X-ray diffraction (XRD) is one of the fastest and most widely used experimental techniques. The width and shape of diffraction peaks contain detailed information about the material's microstructure, including crystallite size, morphology, and lattice defects such as dislocations and planar faults [2]. The established method for extracting this information is X-ray Line Profile Analysis (XLPA), which determines microstructural parameters by analyzing diffraction peak profiles.

However, a major limitation of conventional XLPA techniques emerges in high-throughput experimental settings [1]. When spatially resolved XRD measurements are performed—such as during combinatorial synthesis or synchrotron experiments—researchers may collect hundreds or thousands of diffractograms. Interpreting such large datasets manually using traditional XLPA methods becomes extremely time-consuming [2].

Conventional approaches, including Whole Powder Pattern Fitting (WPPF), Whole Powder Pattern Modeling (WPPM), and Convolutional Multiple Whole Profile (CMWP) fitting, rely on iterative, non-linear regression to minimize the difference between measured and theoretically modeled diffraction patterns. These conventional XLPA techniques, such as CMWP, fit experimental patterns by using theoretical functions derived from models of the microstructure. While highly accurate, these methods are computationally intensive and impractical for rapid, large-scale structural mapping, creating a significant bottleneck in materials discovery [2]. This limitation is especially pronounced in high-throughput experimental settings (like synchrotron experiments or combinatorial synthesis), where researchers may collect hundreds or thousands of diffractograms.

To address this challenge, recent research has introduced **Machine Learning–based X-ray Line Profile Analysis (ML-XLPA)** [1]. After an initial training phase, ML-XLPA enables rapid prediction of microstructural parameters directly from XRD patterns, significantly accelerating the determination of microstructural parameters from experimental patterns. This speed drastically increases efficiency and reduces analysis time from hours or days to seconds for large-area analysis. This makes it possible to generate detailed maps of microstructure as a function of chemical composition, supporting high-throughput materials characterization [2]. The ML-XLPA framework combines physically motivated diffraction modeling (used to create the labeled training set), robust data preprocessing (such as the ARPLS method for background subtraction), and high-performance machine learning regression techniques (such as gradient boosting implemented in XGBoost).

So far, ML-XLPA methods have been developed primarily for cubic crystal structures, such as FCC materials [1]. However, many technologically relevant materials exhibit hexagonal crystal structures, which require different diffraction modeling and cannot be reliably analyzed using

cubic-based approaches. The capability of the ML-XLPA technique needs extension to other structures, such as hexagonal close-packed (HCP) structures, which is cited as a subject of future work [2]. In addition, real experimental data introduce further challenges, including small peak shifts, variations in relative peak intensities, and measurement noise, all of which can degrade prediction accuracy if not properly addressed [1–6].

The overarching goal of this project is to extend the ML-XLPA framework to overcome these limitations. By generating synthetic diffraction data using physically motivated modeling and developing robust machine-learning models (such as gradient boosting), the aim is to enable fast and reliable microstructural characterization of complex material systems. Ultimately, applying such methods to real combinatorial material libraries will facilitate the exploration of structure–property relationships and support the efficient discovery of new materials. The ability to quickly determine maps of microstructural parameters (like crystallite size, dislocation density, and twin fault probability) across combinatorial samples is essential for studying the correlation between chemical composition and microstructure.

2 Scope and Contribution of the Current Semester

The work carried out during this semester primarily focused on an in-depth study and critical understanding of the theoretical and methodological foundations of machine learning–based X-ray line profile analysis. This included a detailed examination of conventional XLPA techniques, the physical modeling of diffraction peak broadening, the generation of synthetic training data, advanced background correction methods, and the application of machine learning regression models. Considerable effort was devoted to understanding the underlying assumptions, definitions, and limitations of the existing ML-XLPA framework, as well as the practical challenges associated with data availability and reproducibility. Due to the complexity of the methods and the limited accessibility of suitable experimental and synthetic datasets, no independent implementation or validation was performed during this period. Instead, this semester established the necessary conceptual and technical groundwork required for the methodological extensions and experimental investigations planned for the next phase of the project.

3 Plan for Next Semester

Building on the theoretical understanding established this semester, the next phase of the project will focus on methodological development and practical implementation. The first objective is to extend the ML-XLPA framework beyond single-phase FCC materials by developing diffraction pattern generation and learning strategies for hexagonal close-packed (HCP) crystal structures. In parallel, efforts will be made to improve data preprocessing to handle non-smooth experimental backgrounds, including weak secondary-phase peaks and other artifacts that currently limit the accuracy of microstructural parameter estimation, particularly for dislocation density. A further goal is to generalize the ML-XLPA approach to multiphase materials with strongly overlapping diffraction peaks, enabling realistic analysis of complex compositional libraries. These developments will be accompanied by systematic testing on synthetic and, where possible, experimental datasets.

References

[1] Péter Nagy, Bálint Kaszás, István Csabai, Zoltán Hegedűs, Johann Michler, László Pethö, and Jenő Gubicza. Machine learning-based characterization of the nanostructure in a combinatorial co-cr-fe-ni compositionally complex alloy film. *Nanomaterials*, 12(24):4407, 2022.

[2] Bálint Kaszás, Péter Nagy, and Jenő Gubicza. Diffault: Simulation of diffraction patterns of faulted crystals. *SoftwareX*, 27:101860, 2024.

[3] Gábor Ribárik, Jeno Gubicza, and Tamás Ungár. Correlation between strength and microstructure of ball-milled al–mg alloys determined by x-ray diffraction. *Materials science and engineering: A*, 387:343–347, 2004.

[4] Sung-June Baek, Aaron Park, Young-Jin Ahn, and Jaebum Choo. Baseline correction using asymmetrically reweighted penalized least squares smoothing. *Analyst*, 140(1):250–257, 2015.

[5] Péter Nagy, Nadia Rohbeck, Zoltán Hegedűs, Johann Michler, László Pethö, János L Lábár, and Jenő Gubicza. Microstructure, hardness, and elastic modulus of a multibeam-sputtered nanocrystalline co-cr-fe-ni compositional complex alloy film. *Materials*, 14(12):3357, 2021.

[6] Jenő Gubicza. *X-ray line profile analysis in materials science*. IGI global, 2014.