doi:10.1017/S1431927618013107
Microsc. Microanal. 24 (Suppl 2), 2018

Machine Learning-based Crystal Structure Prediction for X-Ray Microdiffraction

Yuta Suzuki[1,3], Hideitsu Hino[2], Yasuo Takeichi[3], Takafumi Hawai[3], Masato Kotsugi[1], and Kanta Ono[3,*]

[1.] Tokyo University of Science, Department of Materials Science and Technology, Tokyo, Japan.
[2.] The Institute of Statistical Mathematics, Tokyo, Japan.
[3.] High Energy Accelerator Research Organization, Institute of Materials Structure Science, Ibaraki, Japan.
* Corresponding author, kanta.ono@kek.jp

Introduction:
X-ray diffraction (XRD) is an essential tool for phase identification and the characterization of crystal structure, impurity, strain, and texture. Conventional XRD yields only averaged results, and recent developments in micro- and nano-X-ray diffraction (μXRD) with highly focused X-ray beams has enabled diffraction mapping characterization [1]. However, the bottleneck in the workflow of μXRD experiments is the analysis of large XRD pattern dataset, most of the experiments using μXRD are limited to point analyses, while the instruments are capable of diffraction mapping characterization that outputs tens of thousands of XRD patterns. A typical structural characterization using XRD needs to extract structural descriptors from the obtained patterns; the standard structural descriptors are crystal symmetry, space group, and lattice parameters. We attempted to predict the crystal structure from XRD patterns using machine learning (Fig. 1). This machine learning-based automated crystal structure prediction will contribute to the realization of on-the-fly data analysis in materials science.

Experimental:
A dataset with 188,607 XRD patterns was prepared by simulating the patterns of all materials in the Inorganic Crystal Structure Database (ICSD) [2] using Pymatgen middleware [3]. The XRD calculation was performed using Cu Kα radiation (wavelength = 1.5418 Å) with the 2θ range set to 0°–90°. We used the 2θ positions of the ten peaks in ascending order of 2θ, and the number of diffraction peaks as the descriptor for the patterns. We used random forests [4] to classify the XRD patterns into crystal systems and space groups. The prediction accuracy was estimated using 10-fold cross-validation.

Results and Discussion:
The prediction accuracy of the classification of the XRD patterns into space groups (230 classes) was 83.62%, which exceeded that achieved in a recent study using deep learning (81.14%) [5], although we used only ten 2θ positions and the number of peaks as descriptors. The computational time for constructing the classification model using random forests was several minutes, and the classification of one XRD pattern took less than 1 ms. Further, we classified the XRD patterns based on crystal systems (seven classes) using random forests and achieved a prediction accuracy of 93.07% (Fig. 2).

In this study, we demonstrated that the crystal structures could be estimated from the peak positions of the XRD patterns using machine learning. It was shown that the performance of ensemble learning such as random forests was comparable to that of deep learning, with the former having a lower computational cost and higher speed. We also successfully predicted the lattice constants using random forest regression, which will be discussed in detail in the near future. Using the proposed method, a machine learning model could be constructed according to a specific experimental setup (i.e., depending on the 2θ range or X-ray wavelength). Thus, we streamlined the materials discovery workflow, and the

accelerated characterization of materials will be realized by combining high-throughput experiments through optimal measurements [6] and on-the-fly data analysis (as shown in this research) and knowledge acquisition from the data [7].

References:

[1] H B Qin *et al*, Environ. Sci. Technol. **51** (2017), p. 6027–6035.
[2] https://www.fiz-karlsruhe.de/en/leistungen/kristallographie/icsd.html
[3] S P Ong *et al*, Comput. Mater. Sci. **68** (2013), p. 314–319.
[4] L Breiman, Mach. Learn. **45** (2001), p. 5–32.
[5] W B Park *et al,* IUCrJ **4** (2017), p. 486–494.
[6] T Ueno *et al*, npj Comput. Mater. **4** (2018), p. 4.
[7] This work was partly supported by JST CREST JPMJCR1761.

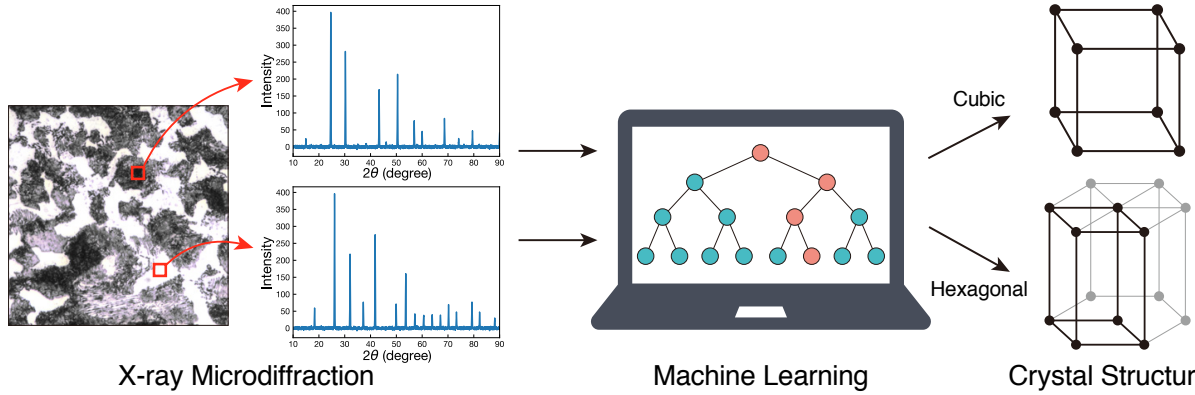

Figure. 1. Schematic figure of machine learning methodology for crystal structure prediction of X-ray diffraction patterns. This methodology enables us the on-the-fly analysis of the obtained data, and spatially map a distribution of crystal structure combined with μXRD.

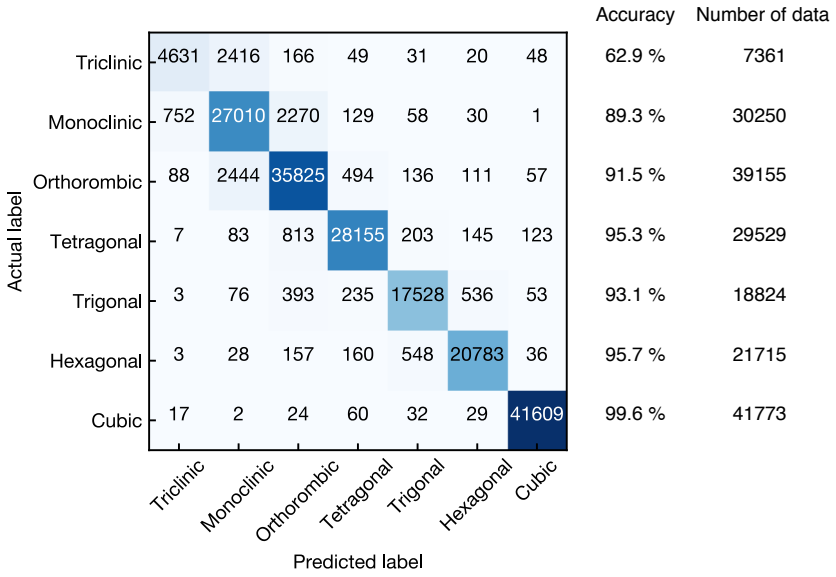

Figure. 2. The accuracy of predicting crystal systems. The accuracy was evaluated with 10-fold cross-validation of the whole data. This performance is sufficient for the on-the-fly data analysis in high-throughput measurement such as μXRD. It should be noted that 99.6 % classification accuracy in a cubic system was achieved.

https://doi.org/10.1017/S1431927618013107 Published online by Cambridge University Press