Automated Lattice Constant Estimation of X-ray Diffraction Patterns by Ensemble Learning

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

1. Tokyo University of Science, Tokyo, Japan., 2. The Institute of Statistical Mathematics, Tokyo, Japan., 3. High Energy Accelerator Research Organization, Ibaraki, Japan.

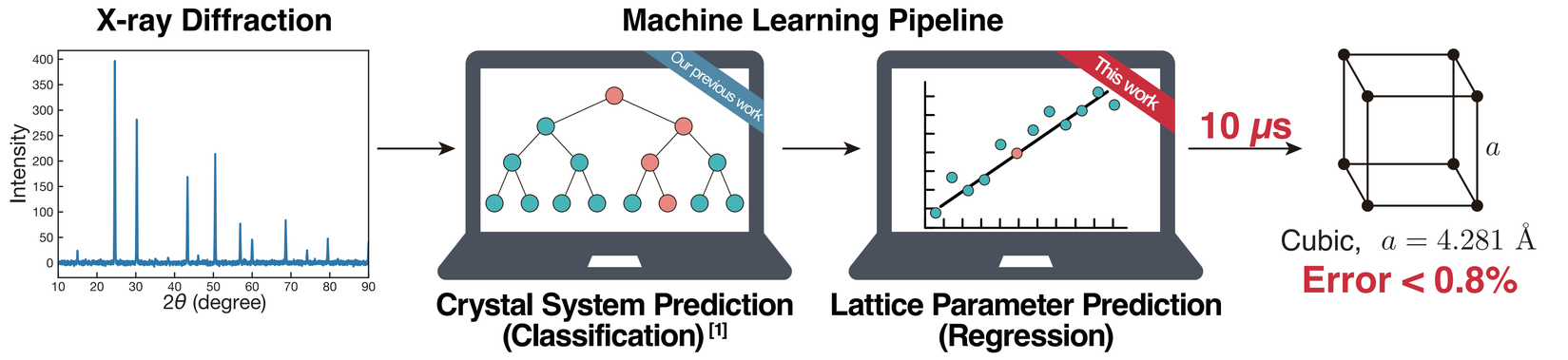
Introduction

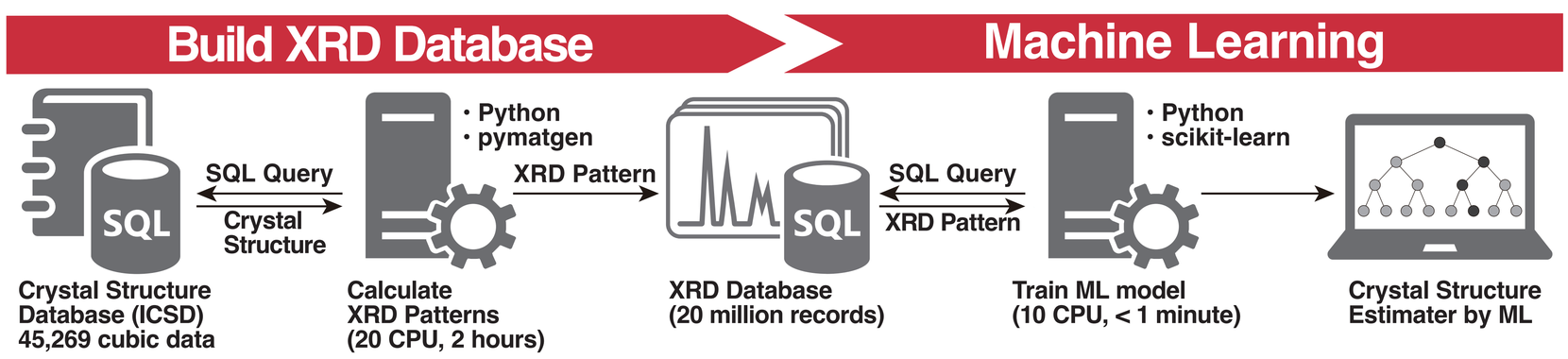
- **Materials Informatics (MI)**
  Accelerates materials discovery and obtains the knowledge by statistical learning.
- **Keys of accelerated materials discovery**
  1. Materials design by machine learning
  2. High-throughput experiments
  3. On-the-fly data analysis
- **X-ray Diffraction (XRD)** is one of the most important techniques for materials characterization. But, the conventional analysis is performed manually, and it could be a bottleneck of materials discovery workflow.
- Estimation of crystal systems tand lattice parameters from XRD pattern is difficult, since try-and-errors are required. **If we build machine learning(ML) model to predict the crystal structure from a XRD pattern, the data analysis will be automated and accelererated.**

Objectives

- **Automated estimation of crystal structures and lattice parameters from XRD patterns with machine learning.**
- We are aiming to accelererate materials discovery by the integtaration of the high-throughput materials measurement and on-the-fly data analysis.

Our strategy and Methods





- International Crystal Structure Database (ICSD) was used for the data source. Out of 200k entries of all crystal systems including invalid information, 45k valid entries of cubic systems were used.
- XRD patterns were calculated using pymatgen [2] middleware. The wavelength and 2θ range were set to CuKa radiation (1.5418 Å) and 0˚-90˚, respectively.
- **We chose the 2θ positions of first ten peaks for the descriptor of the XRD patterns.** The ML model was trained with the dataset of ten peak positions and the lattice parameter paired with them.
- Random Forests (RF) [3] was employed for the machine learning algorithm.
- RF is flexible and able to express complex nonlinear functions. It offers stable estimation result by the ensemble of many decision tree estimaters. From these favorable characteristics, we applied RF to this research.
- The generalization performance, the response to new data was examined with out-of-bag (OOB) samples of RF, and new materials added in ICSD Ver.2018.1. OOB validation is approximately the same as N-fold cross-validation.

Results and Discussion

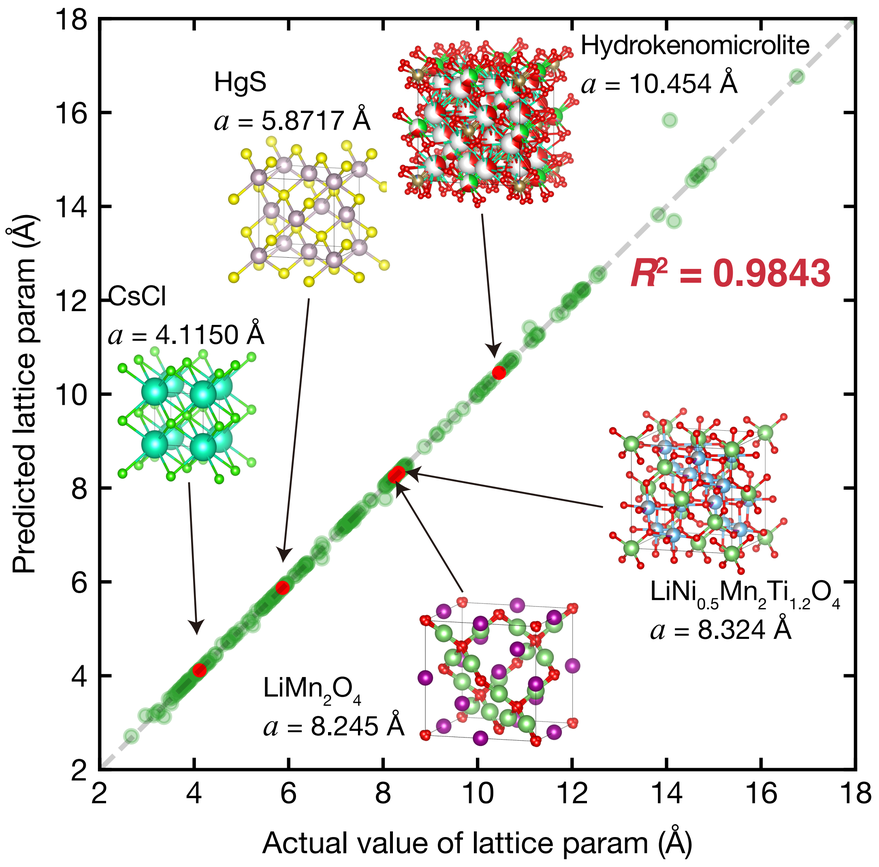


Fig.1 The prediction results of lattice parameters.

- **The prediction error of lattice parameter** (mean relative absolute error, **MRAE**) **was 0.83 %, and correlation coefficient R^2 = 0.9843.** 92.6 % of used structures showed MRAE of less than 1 % and 72.6 % marked less than 0.01 % (Fig. 1, 2).
- Although this model is a prototype only for cubic systems, it shows the possibility that the crystal systems and lattice parameters can be estimated quickly and automatically from XRD patterns by machine learning.

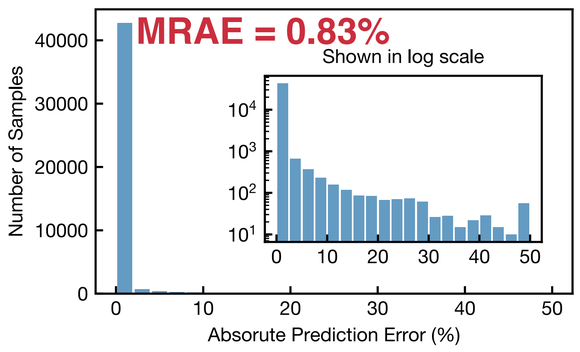


Fig.2 The histogram of prediction error.

$$\mathrm{MRAE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{a_{pred}}{a_{actual}} - 1\right|$$ n is the number of data.

Summary

- **We applied machine learning to crystal system estimation from powder diffraction patterns and showed our ML model achieved promising performance.**
- **Further improvement could be made by combining other established ML methods with our model and introducing crystallographic knowledge.**
- **Our result suggests that automatic analysis of XRD data is possible in a data-driven fashion, if the model is successfully extended to other crystal systems.**
- **We are now working on ML-based automated peak-indexing and impurity peak identification.**

Previous Work[1]: Crystal System Classification

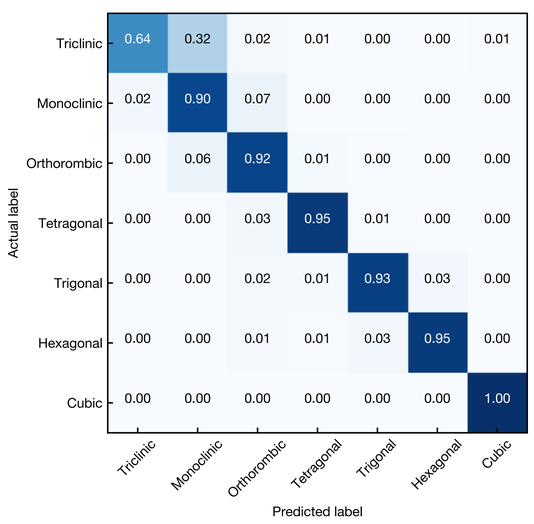


Fig.3 The confusion matrix of crystal system prediction from XRD pattern with RF. Each number indicates the percentage of samples predicted for that area.

Crystal System	Accuracy (%)
Cubic	99.48
Tetragonal	95.59
Hexagonal	95.36
Monoclinic	90.73
Trigonal	93.10
Orthorombic	91.65
Triclinic	63.65
All	93.34

Tb.1 The accuracy of crystal system prediction.

- **The prediction accuracy for crystal systems (seven classes) was 93.34 % (Fig. 3, Tb. 1).**
- **99.48 % accuracy for cubic system was obtained and the estimation took less than 1 ms per XRD pattern.**
- **The prediction accuracy for space groups (230 classes) was 84.03 %.**
- The relatively poor prediction performance for triclinic system was caused by the insufficiency of data. ICSD contains only 7300 triclinic materials (4 %) which is not enough to train ML model. Data augmentation might help this problem.

Resources

- Please visit our GitHub and website for download this poster and contact us.

github.com/resnant/ENGE2018

resnant.github.io

References

1. Suzuki. Y. et al. Microsc. Microanal. **24**, 144–145 (2018).
2. Ong, S. P. et al. Comput. Mater. Sci. **68**, 314–319 (2013).
3. Breiman, L. Random Forests. Mach. Learn. **45**, 5–32 (2001).

- This research is partly supported by JST ACT-I (JPMJPR18UE).