Invited Lecture

CrystalMELA: A machine learning-based web platform for polycrystalline characterization

R. Rizzi[1], N. Corriero[1], N. Del Buono[2], G. Settembre[2], D. Diacono[3]

[1] Institute of Crystallography (IC-CNR), Bari, Italy, [2]Department of Mathematics, University of Bari, Bari, Italy, [3]National Institute for Nuclear Physics, Bari, Italy

rosanna.rizzi@cnr.it

In the past few years, remarkable advancements in data-driven models and the accessibility of extensive experimental data from various sources have facilitated the development and implementation of Artificial Intelligence in materials science, particularly machine learning (ML) algorithms for diffraction data analysis. These algorithms have proven to be particularly useful and effective in characterizing polycrystalline compounds.

A powder diffraction pattern is mainly affected by peak overlaps, difficulty in the correct background estimation, presence of preferred orientation effects, and limited experimental resolution. These factors collectively complicate the process of structure solution. Importantly, establishing initial steps such as pattern indexation and space group determination can be challenging, and their accurate determination is particularly critical, especially if more than one chemical phase is present in the compound. Furthermore, an incorrectly defined unit cell does not lead to the structural solution. It happens despite the progress, availability, and variety (in terms of strategies and methods implemented) of automatic indexing software such as DICVOL [1], N-TREOR09 [2] and ITO [3].

A new web platform based on machine learning (ML), named CrystalMELA (Crystallographic MachinE LeArning) [5], has been developed for crystal system classification. The aim is to try to overcome the challenges encountered in the structure solution process from powder diffraction data, and to complement traditional indexing approaches. The platform is currently designed for classifying the seven crystal classes. In the current version, the platform can run three different and complementary ML models: a Convolutional Neural Network (CNN), a Random Forest (RF) and an Extremely randomized trees (ExRT). These models have been trained using theoretical powder diffraction patterns from over 280,000 crystal structures of inorganic, organic, organo-metallic compounds and minerals as collected in the POW_COD database [6]. A classification accuracy of 70% has been achieved, which improves to 90% when considering the top-2 accuracy.

CrystalMELA is freely available to the scientific community, and its homepage is shown in Fig. 1. All classification options in the CrystalMELA platform are designed to be powerful and easy to use, supported by a user-friendly graphic interface. The key features, along with several real-case application examples, will be showcased.

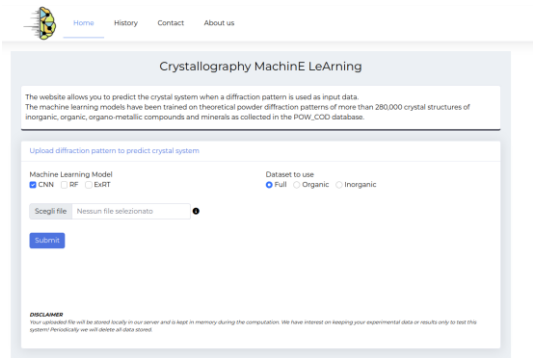

Figure 1. The Home web page of CrystalMELA platform.

[1] Boultif, A., Louer, D. (2004). *J. Appl. Cryst*., **37(5)**, 724.

[2] Altomare, A., Campi, G., Cuocci, C., Eriksson, L., Giacovazzo, C., Moliterni, A., Rizzi, R., Werner, P.-E. (2009). *J. Appl. Cryst*., **42(5)**, 768.

[3] Visser, J. W. (1969). *J. Appl. Cryst*., **2(3)**, 89.

[4] Mueller, T., Kusne, A. G., Ramprasad, R. (2016). *Machine Learning in Materials Science*, John Wiley and Sons, Ltd., chap. 4, pp. 186–273.

[5] Corriero, N., Rizzi, R., Settembre, G., Del Buono, N., Diacono, D. (2022). *J. Appl. Cryst*., Submitted.

[6] Altomare, A., Corriero, N., Cuocci, C., Falcicchio, A., Moliterni, A., Rizzi, R. (2015). *J. Appl. Cryst*., **48**, 598.