Machine Learning For High Entropy Alloy Design And Discovery: Perspective, Challenges, And Future Opportunities

*A. Debnath, W.F. Reinhart

The design and discovery of advanced materials for applications in energy generation and storage, water purification, and carbon sequestration is necessary for meeting the needs of a growing world population. However, design of advanced materials with desired value of properties is challenging due to the large and mostly uncharted design space and the difficulty in predicting the non-linear relationships between structure, property, and processing parameters. While conventional design approaches have been successfully used in the past, they are often time-consuming and inefficient as they require a combination of chemical intuition and serendipity. The advent of machine learning has opened a new paradigm of materials design and has increasingly been adopted by the scientific community. However, their application in the materials science domain is impacted by several challenges that need consideration. We will illustrate some of these using High Entropy Alloy (HEA) as our material of interest.

Data sparsity is one of the biggest challenge due to the destructive nature of most mechanical tests and high costs associated with manufacturing multiple samples with varied requirements, resulting in limited number of property measurements performed on each sample. Training machine learning or deep learning models on such small datasets could lead to overfitting and unreliable predictions. However, the inherent correlation between property, chemistry, and processing, can be leveraged using methods like transfer learning and multi-task learning. The second challenge involves the principle of inverse design – the process of configuring material chemistry given target material properties. While generative models have been successfully applied for inverse design of HEAs, a major unanswered question is how they compare against the state-of-the-art forward design methods like high throughput screening and optimization. Finally, as the importance of capturing the underlying data distribution is important for generating new synthetic data, we also probe several popular synthetic data generation methods.