

Building a machine learning model for properties-based selection of High-Entropy Alloy compositions

Marzuk Kamal

AeonX AI SAS, Paris, France

Machine learning models were developed with the aim to both predict the properties of High-Entropy Alloys (HEAs) from an input chemical composition selected over an element pool that includes Fe, Mn, Cr, Ni, Al, Co, Cu, Mn, Mo, Si and Ti, and, more importantly, to solve the inverse problem of identifying HEA compositions that meet specific requirements on their properties. Specifically, the equilibrium phase composition at specified temperatures, the stacking fault energy (SFE), the hardness, the pitting potential, and the sliding wear rate were the selected properties.

A variety of models and approaches were integrated to this end, since very different datasets were available. Equilibrium phase composition data can be generated at high-throughput through the Calculation of Phase Diagrams (CalPhaD) method; hence, a pool of almost 4000 data for 4- and 5-element compositions could be generated. An artificial neural network (ANN) model was trained onto a subset of 80% of the dataset and validated against the remaining 20%, achieving an excellent root mean square error (RMSE) validation loss of 0.027. The model was further expanded through an explainable AI model to identify which element(s) contribute the most to obtaining a specific result in terms of phase composition. To incorporate a physical understanding of the system, the training of the ANN included not only the chemical composition of each alloy, but also some fundamental structural and thermodynamic alloy parameters such as the enthalpy and entropy of mixing and their ratio, the valence electron concentration (VEC) and the atomic radius asymmetry, which are known to have a significant influence on the phase composition of multi-element alloys-

A similar approach could be employed for SFE, which is closely related to the deformation mechanism of the alloy and, therefore, its ductility and the possibility of toughening mechanisms such as TWIP and TRIP plasticity. Density Functional Theory (DFT) calculations provided a dataset of almost 400 datapoints, which were employed to train a corresponding ANN.

On the other hand, computational models for properties such as wear resistance and corrosion rate cannot provide large datasets in high-throughput mode. For the corrosion resistance of the materials, therefore, we trained a ML model on a literature dataset of more than 550 pitting potential values of HEAs measured in NaCl aqueous solution, using an experimental dataset of ~140 values obtained during the project for the validation. For the sliding wear rate, we employed a simpler ML model, Support Vector Machine (SVM) with radial basis function (RBF) kernel, on a more limited dataset generated entirely experimentally within the CoBRAIN project using a ball-on-disc method in unidirectional dry sliding configuration against a hard, inert counterbody (Al_2O_3).

Very importantly, for each property, the resulting ML models provide not only a predicted value, but also a reliability index calculated through the Mahalanobis distance, which indicates how far the specific modelled value is from the training range.

The models have been integrated into an API which can be used to solve the inverse problem, i.e., given desired targets in terms of one or more of the properties listed above, and possible restrictions on the compositional range, it returns all compositions (in 5wt.% steps for each element) that meet the requirements, including a prediction of their properties and the related Mahalanobis distance.

Thus, the resulting ML API can be integrated into a Decision Support System to help a user identify the most suitable compositions for a given application, provided that the application requirements can be translated into property targets.