Supplementary Information to "Predicting Mechanical Properties of Non-Equimolar High-Entropy Carbides using Machine Learning"

Xi Zhao[a,b*], Shu-guang Cheng[a], Sen Yu[b], Jiming Zheng[a], Rui-Zhi Zhang[c], Meng Guo[c]

[a] *School of Physics, Northwest University, Xi'an 710069, China*

[b] *Northwest Institute for Nonferrous Metal Research, Xi'an 710016, China*

[c] *Jinan Key Laboratory of High-Performance Industrial Software, Jinan Institute of Supercomputing Technology, Jinan, Shandong 250103, China*

* Corresponding author: zhao_xii@stumail.nwu.edu.cn

S1 – The crystal structures for high-throughput density functional calculations of elastic properties for 495 carbides and 123 high-entropy carbides

The crystal structure (left) used for the calculation of the elastic properties of 495 carbides is a single-phase salt rock structure with 8 atoms, and the crystal structure (right) used for the calculation of the elastic properties of 123 high-entropy carbides is a 2x1x1 single-phase salt rock structure containing 16 atoms. For all calculation structures, transition metal atoms Ta, Zr, Hf, V, Nb, Ti, Mo, W, Cr occupy the cation sites and carbon atoms occupy the anion sites.

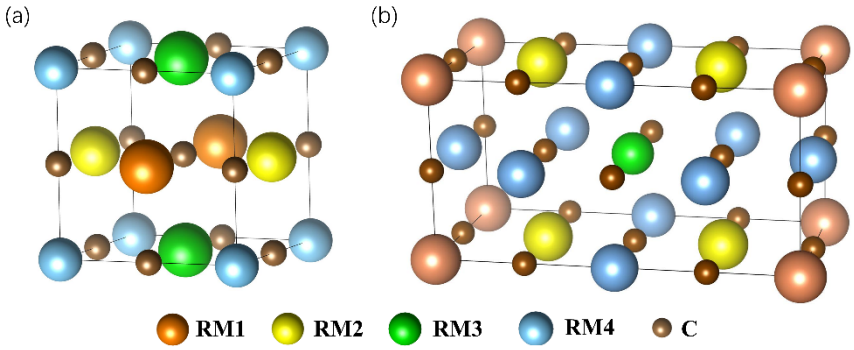


Fig S1 Calculation structure for carbides and high entropy carbides

S2 –Detailed description of the 6 CBFV descriptors

The feature descriptors used in this work are listed on GitHub(https://github.com/ZhaoXi1209/HECs-Mechanical-Properties-prediction).

Table S1 · Description and comparison of CBFV descriptors—Jarvis, Magpie, Mat2vec, Onehot, Oliynyk, Random

Descriptor	Description	Advantages
Jarvis	A collection of feature descriptors used to characterize elements and compounds, including information such as the electronic structure, crystal structure, mechanical properties, thermodynamic properties, etc	Comprehensive description of material properties, suitable for diverse machine learning tasks
Magpie	A set of characteristic descriptors developed by Ward et al., based primarily on the physical and chemical properties of elements, including atomic radius, electronegativity, melting point, boiling point, etc., describes materials through statistical information of these properties (such as mean, variance, minimum, maximum, etc.)	Captures a wide range of element properties, providing useful statistical information.
Mat2vec	Based on natural language processing (NLP) descriptors, word embeddings are used to represent materials. Each material or element is represented as a multi-dimensional vector, which can be used for various machine learning tasks.	Capture the contextual information and implicit characteristics of the materials, useful for tasks requiring semantic understanding.
Onehot	Simple encoding method that converts discrete categories (element symbols) into	easy to implement, provides unique representation for each

	numerical vectors with a length equal to the number of elements.	element.
Oliynyk	Based on physical and chemical properties such as atomic radius, ionic radius, electronegativity, melting point, similar to Magpie but with different feature selection and combination	Focus on more complex and diverse material properties, suitable for complex materials and sophisticated machine learning models
Random	Uses randomly generated feature vectors, does not rely on specific physical or chemical properties, often used for baseline comparison or exploratory analysis	Simple to generate, useful for baseline comparisons and exploratory analysis.

S3 – The elastic properties of monocarbides

The elastic modulus of monocarbides is employed to calculate the ROM values of the elastic properties for 495 carbide systems.

Table S2 · The elastic properties of binary carbides

Binary Carbide	**Young's modulus (GPa)**	**bulk modulus (GPa)**	**shear modulus (GPa)**
ZrC	393.59	222.63	163.27
NbC	436.19	301.31	173.27
HfC	452.36	254.57	187.88
TaC	421.97	338.65	163.26
TiC	438.54	266.74	178.85
CrC	427.71	338.28	165.87
VC	556.30	318.72	230.05
MoC	436.19	301.31	173.27
WC	424.41	381.01	161.45

S4 – The comparison of ROM results and HT-DFT calculations of bulk modulus and shear modulus for 495 carbides (contains 9 monocarbides, 108 ternary carbides, 252 quaternary carbides and 126 equimolar quaternary high entropy carbides)

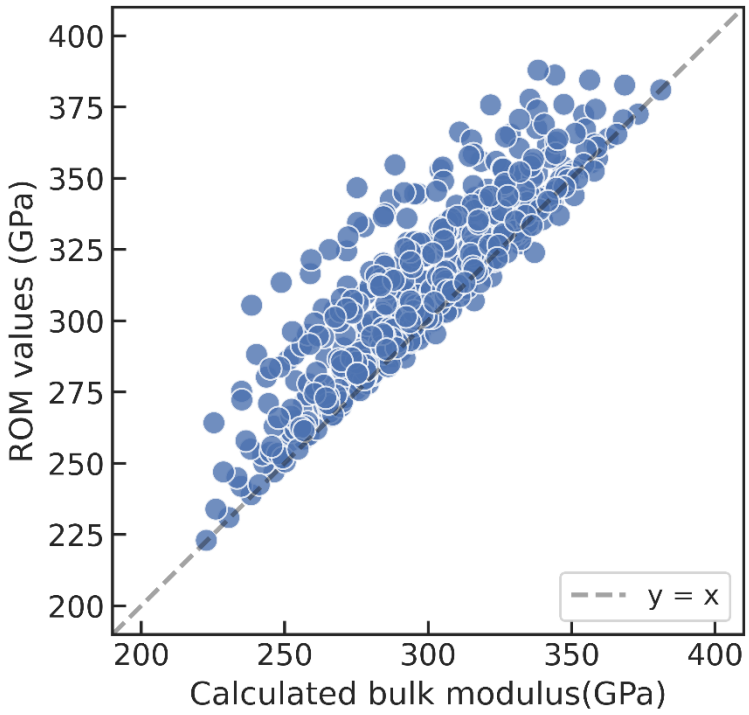


Fig S2 Comparison of ROM results and HT-DFT calculations of bulk modulus for 495 carbides

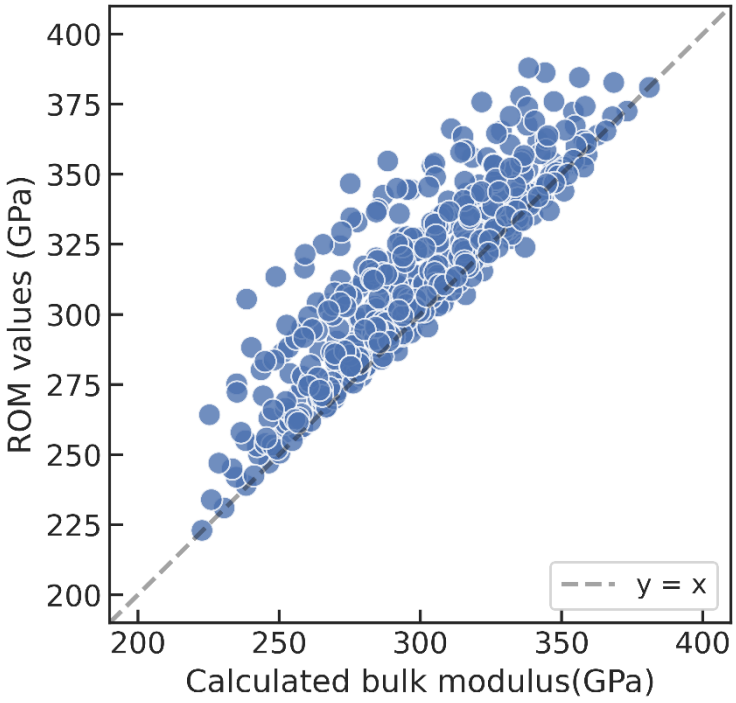


Fig S3 Comparison of ROM results and HT-DFT calculations of shear modulus for 495 carbides

S5 – The machine learning predicted elastic properties of 495 carbides with five kinds of composition-based feature vector descriptors

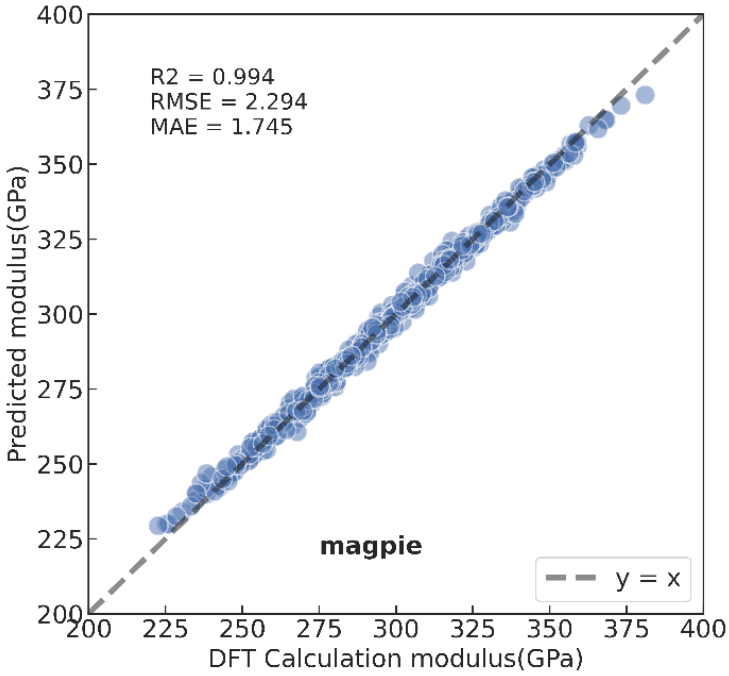
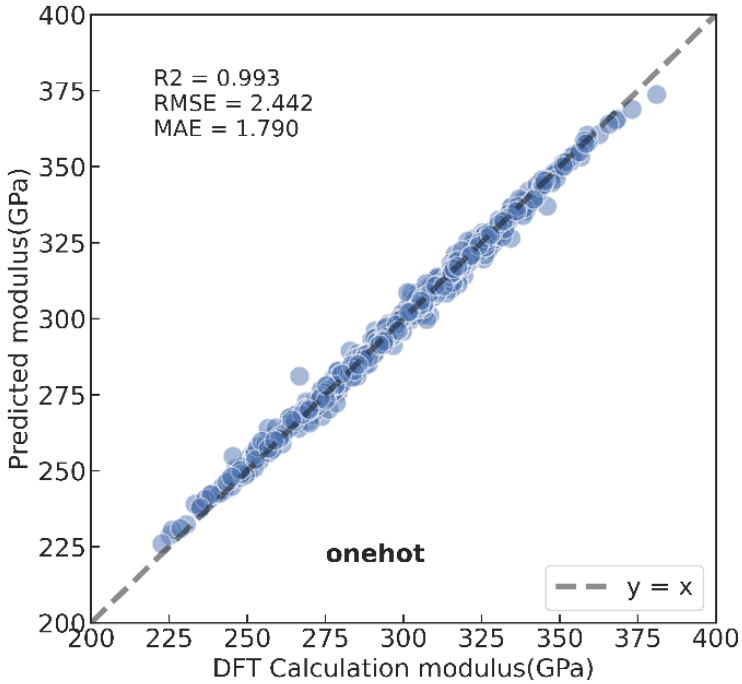


Fig S4 Machine learning prediction results with magpie descriptors on training data

Fig S5 Machine learning prediction results with onehot descriptors on training data



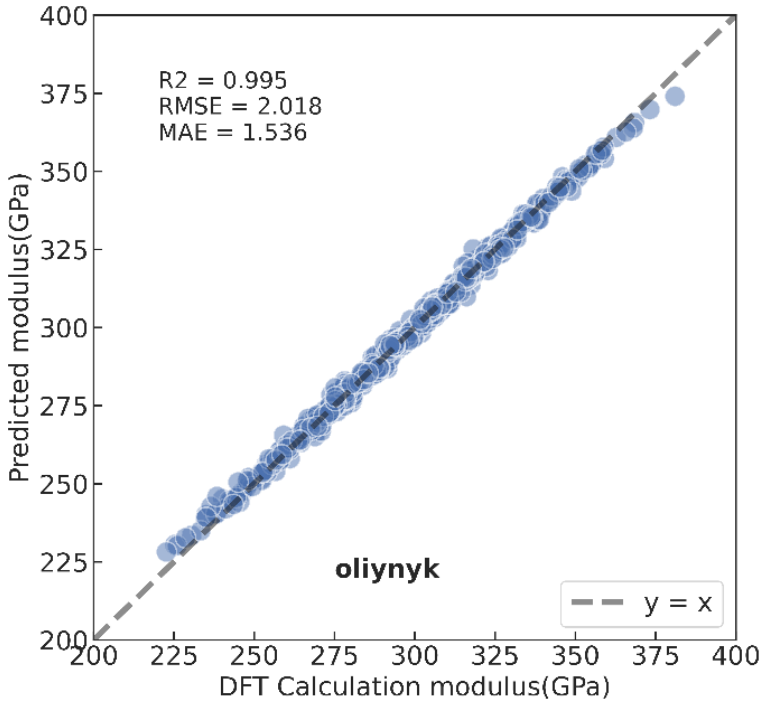


Fig S6 Machine learning prediction results with oliynyk descriptors on training data

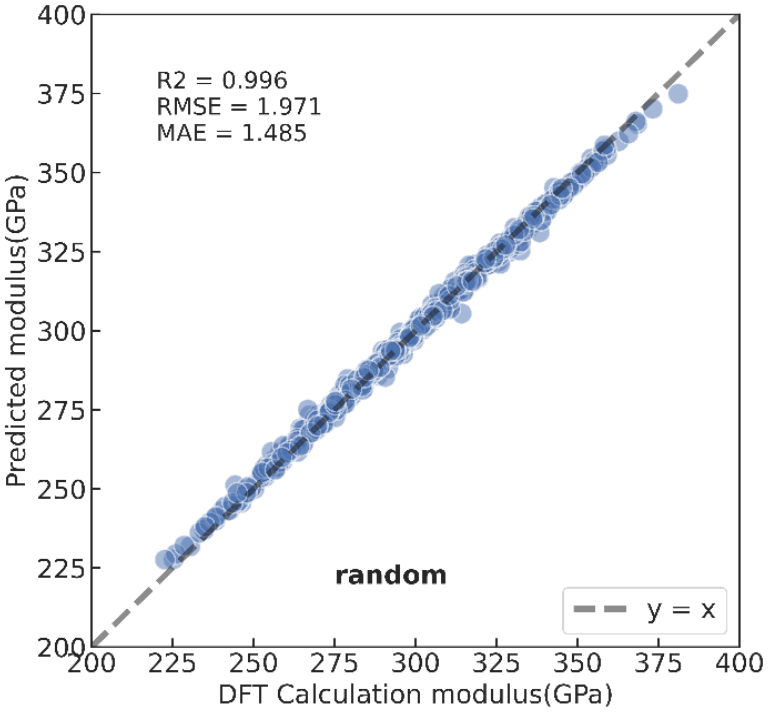


Fig S7 Machine learning prediction results with random descriptors on training data

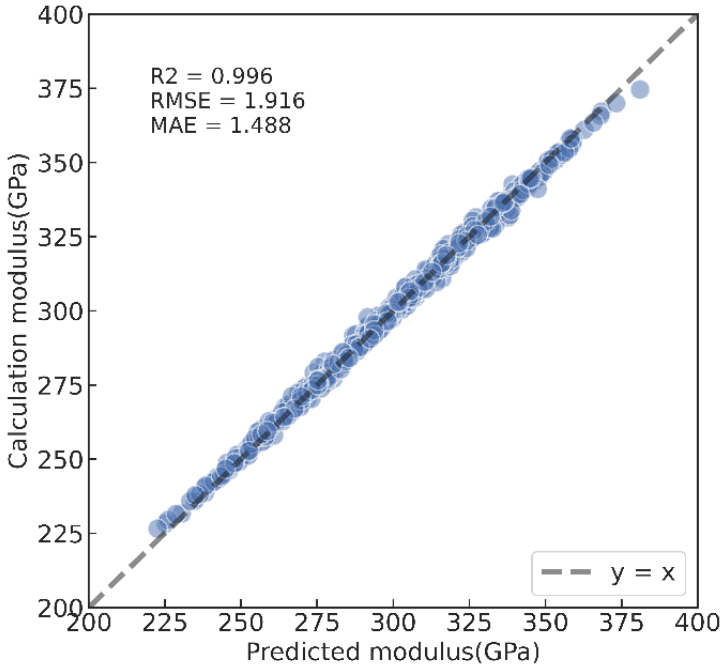


Fig S8 Machine learning prediction results with mat2vec descriptors on training data

S6 –Results of RF model predicting shear modulus

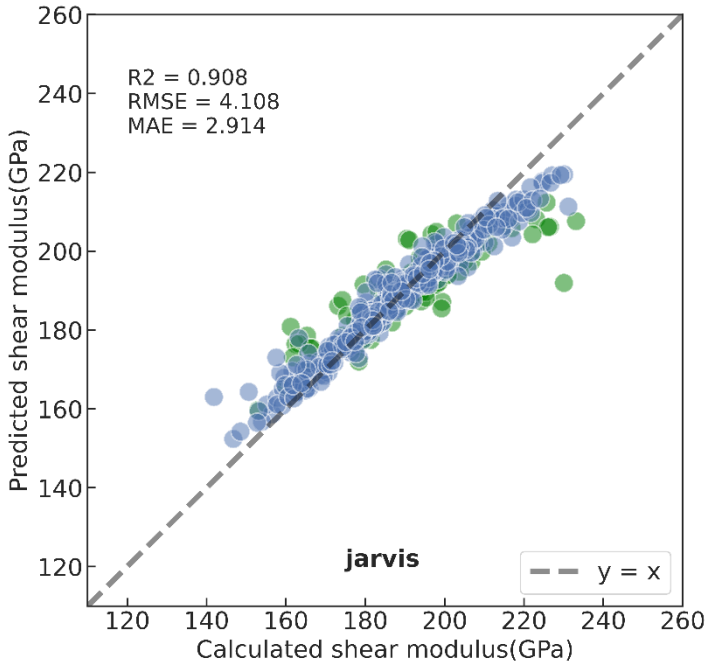


Fig S9 Comparison of DFT calculated shear modulus and predicted shear modulus for both of training (blue) and test (green) datasets with the RF model using Jarvis descriptors;

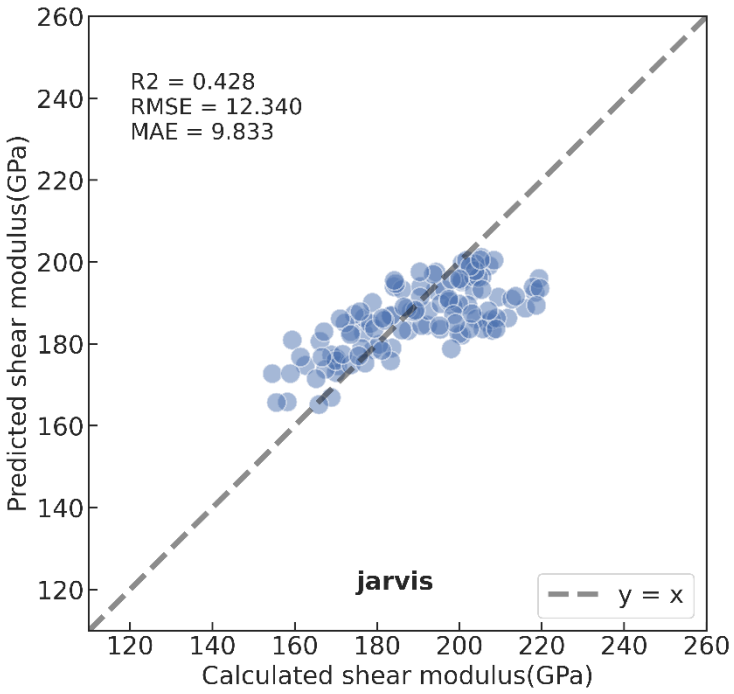


Fig S10 Comparison of DFT results and predicted shear modulus for 123 non-equimolar HECs using the well-trained RF model with Jarvis descriptors;

S7–Results of Crabnet model predicting elastic properties

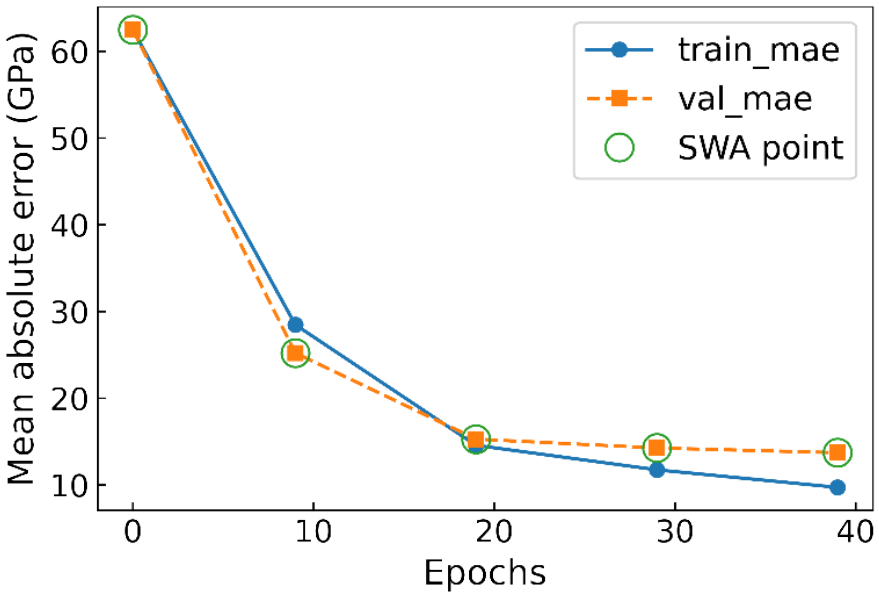


Fig S11 The changes in predicted Young's modulus over the course of the optimization process against the mean absolute error

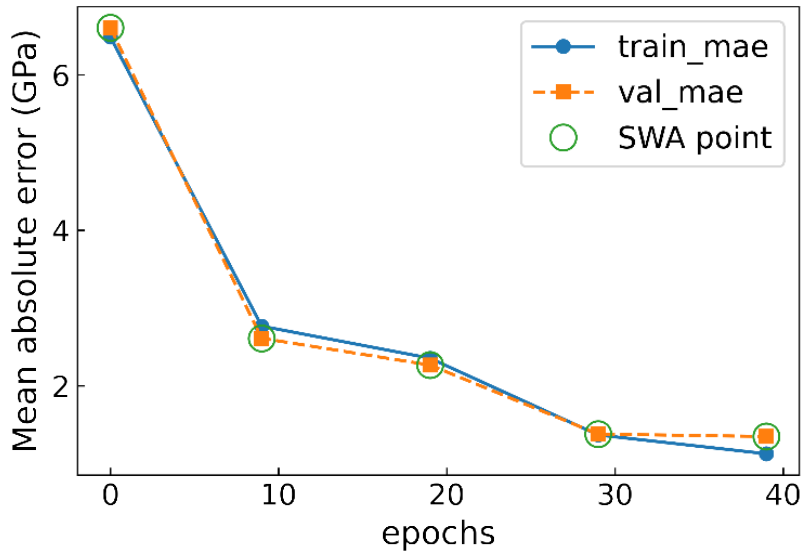


Fig S12 The changes in predicted hardness over the course of the optimization process against the mean absolute error

S8 –Comparison of Crabnet model prediction results with DFT results in Ref

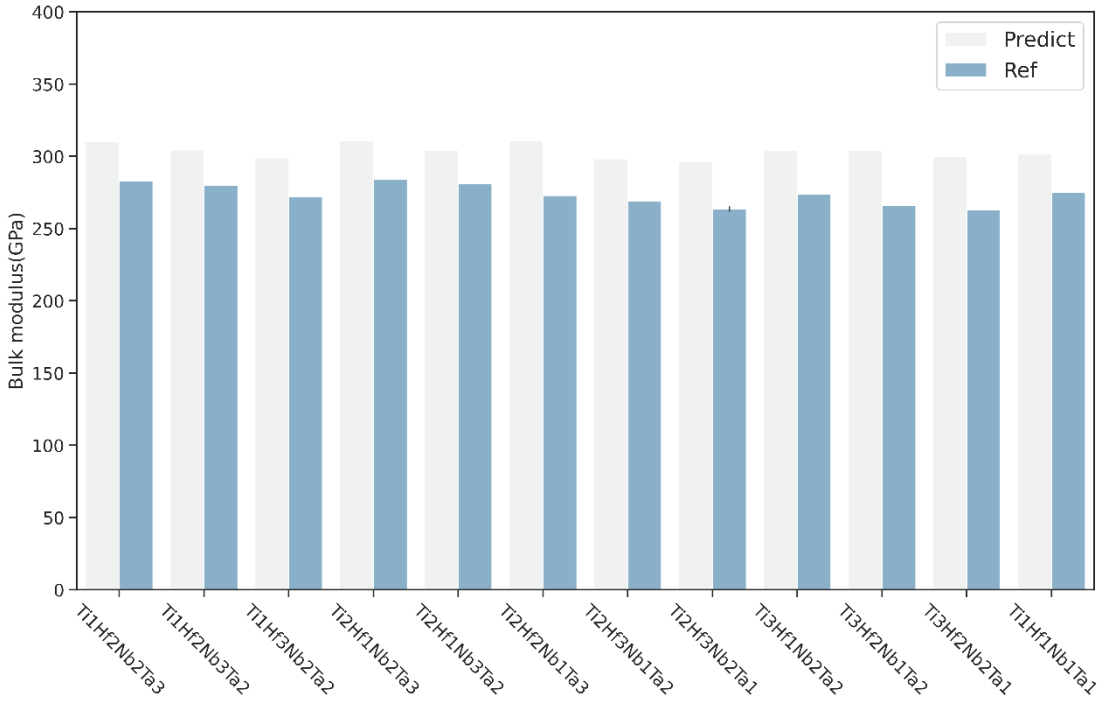


Fig S13 Comparison of bulk modulus of 12 non-equimolar HECs calculated by DFT and predicted by Crabnet model

S9- Results of CrabNet model predicting shear modulus

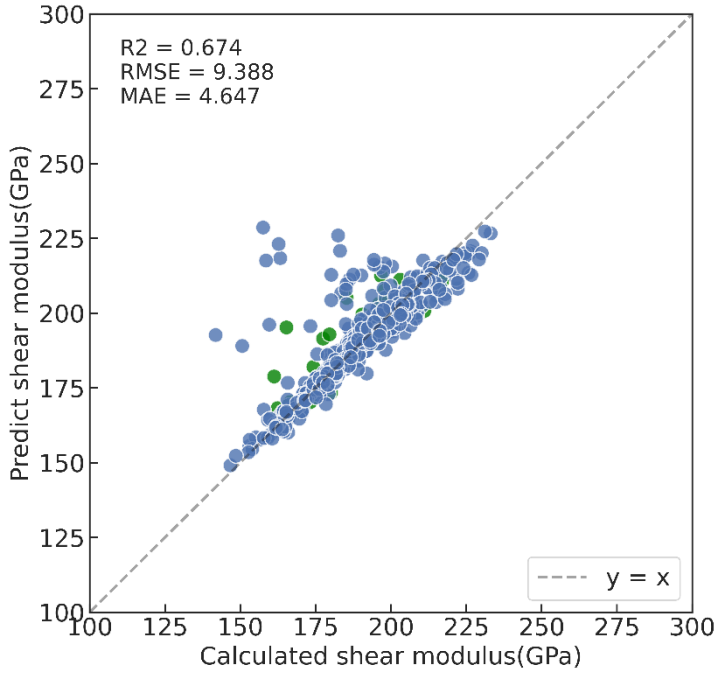


Fig S14 Comparison of DFT calculated shear modulus and predicted shear modulus for both of training (blue) and test (green) datasets with the CrabNet model;

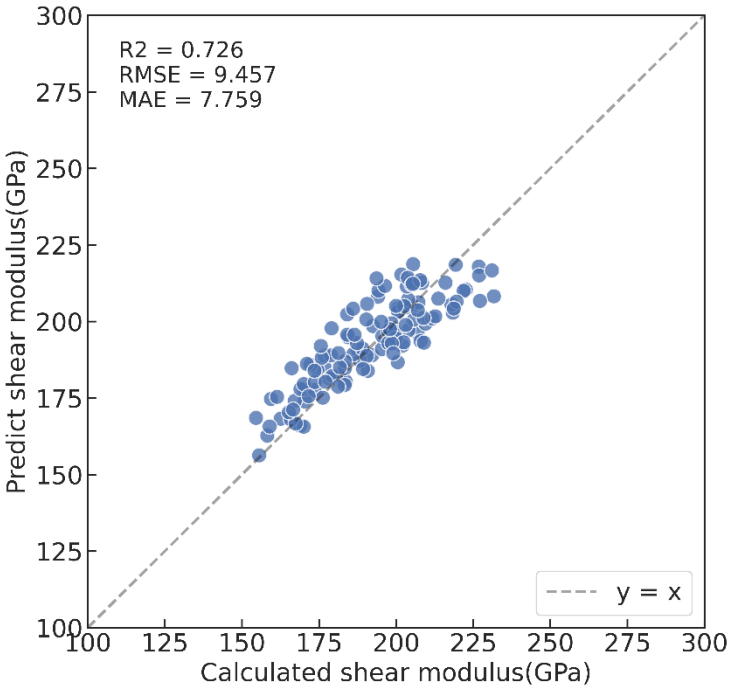


Fig S15 Comparison of DFT results and predicted shear modulus for 123 non-equimolar HECs using the well-trained CrabNet model.