ARTICLE   OPEN

Design high-entropy carbide ceramics from machine learning

Jun Zhang [1], Biao Xu[1], Yaoxu Xiong[1], Shihua Ma[1], Zhe Wang[2], Zhenggang Wu[2✉] and Shijun Zhao [1,3✉]

High-entropy ceramics (HECs) have shown great application potential under demanding conditions, such as high stresses and temperatures. However, the immense phase space poses great challenges for the rational design of new high-performance HECs. In this work, we develop machine-learning (ML) models to discover high-entropy ceramic carbides (HECCs). Built upon attributes of HECCs and their constituent precursors, our ML models demonstrate a high prediction accuracy (0.982). Using the well-trained ML models, we evaluate the single-phase probability of 90 HECCs that are not experimentally reported so far. Several of these predictions are validated by our experiments. We further establish the phase diagrams for non-equiatomic HECCs spanning the whole composition space by which the single-phase regime can be easily identified. Our ML models can predict both equiatomic and non-equiatomic HECs based solely on the chemical descriptors of constituent transition-metal-carbide precursors, which paves the way for the high-throughput design of HECCs with superior properties.

INTRODUCTION

High-entropy materials[1,2] composed of multiple principal elements have attracted extensive attention because of their remarkable properties, including enhanced hardness, mechanical strength, and corrosion resistance[3,4]. Inspired by the concept of entropy stabilization, high-entropy ceramics (HECs), which are single-phase ceramics with no less than four anions or cations[5–7], have been successfully developed. Up to now, different types of HECs have been discovered, including oxides[8,9], borides[10,11], carbides[12–17], nitrides[17–19], sulfides[20], and silicides[21,22], most of which exhibit superior performances in catalysts, thermoelectric conversion, wear resistance, and oxidation resistance[6,23–25]. Among them, high-entropy carbide ceramics (HECCs), which are composed of multiple transition metals in the cation sublattice, have received particular interest due to their high melting points and peculiar high-temperature mechanical properties[14]. In fact, binary transition metal carbides (TMCs) are widely used as ultra-high temperature ceramics in structural applications, such as diffusion barrier layers and protective coatings. TMCs are also considered as potential structural materials used in next-generation nuclear reactors[6]. By incorporating multiple TMCs, including TiC, VC, ZrC, NbC, HfC, TaC, Cr_3C_2, Mo_2C, and WC, the synthesized single-phase multi-principal elemental HECCs can exhibit extraordinary properties compared to their constituent TMCs. Indeed, it has been demonstrated that the mechanical properties of HECCs can surpass their rule-of-mixture (ROM) values[12,26]. In addition, HECCs provide a unique platform to tune the properties of carbides, overcoming the shortcomings of some binary TMCs, such as insufficient fracture toughness and oxidation resistance[26,27]. Though HECCs open a promising avenue to design novel metallic carbides, the vast compositional space makes it difficult to pinpoint the suitable combinations of species and concentrations that can form single-phase HECCs. Because of the complexity in phase formation abilities, only dozens of single-phase HECCs are discovered up to now[5,6]. Since large-scale trial-and-error experiments are extremely expensive, a rational and efficient high-throughput strategy for HECC phase prediction is highly desired.

Empirical rules have played important roles in guiding the search for new materials over the past years[5]. For example, the widely accepted Hume-Rothery rules have been successfully utilized to design binary alloys[28]. In high-entropy materials, it is generally thought that the high configurational entropy is the critical factor to stabilize the single-phase structures[3]. Nonetheless, it has been increasingly realized that other factors, such as mixing enthalpy, can also affect the phase formation[29,30]. Following the essence of classical Hume-Rothery rules[28], other semiempirical descriptors have been proposed to explain the single-phase formability of high-entropy materials, for instance, the atomic-size difference and electronegativity difference[31–34]. While these semiempirical rules provide essential insight on phase stabilities of high-entropy materials, it is still difficult to make predictions accurately based on these parameters. For HECCs, an entropy descriptor termed as entropy forming ability (EFA) has been introduced for the design of single-phase HECCs[12]. The proposed EFA describes the single-phase stability of given HECCs based on the formation energy distribution spectrum from density functional theory (DFT) calculations. Though the descriptor performs well in most cases, it requires the knowledge of the energy distribution of the considered HECC, which calls for thousands of expensive DFT calculations for each composition to reach a good accuracy.

In recent years, machine learning (ML) methods have been a great success in a number of areas, particularly for predicting new materials[35–39]. Instead of mining limited information from a linear combination of different descriptors, ML can make direct predictions with excellent performance through learning the high-dimensional input data (descriptors). In terms of high-entropy materials, especially for high-entropy alloys, ML models, such as artificial neural network (ANN)[40–42], random forest[43], convolutional neural network (CNN)[42], support vector machine (SVM)[40,42,44–46], and K-nearest neighbors (KNN)[40], are successfully applied to predict their single-phase forming abilities. In contrast, the ML approach for HECC design is still at a very early stage. In this regard, Kaufmann et al.[47] developed a regression ML model to predict the EFA parameter of HECCs based on hundreds of

¹Department of Mechanical Engineering, City University of Hong Kong, Hong Kong, China. ²College of Materials Science and Engineering, Hunan University, Changsha, Hunan 410082, China. ³City University of Hong Kong Shenzhen Research Institute, ShenZhen 518057, China. ✉email: zwu9@hnu.edu.cn; shijzhao@cityu.edu.hk

Table 1. Input features for our ML models and corresponding descriptions.

Feature	Description	Feature	Description
ΔH_{mix}	Mixing enthalpy per formula unit	σ_m	Mass deviation of constituent TMCs
ΔV_{mix}	Volume change per formula unit due to mixing	$\bar{\rho}$	Average density of constituent TMCs
ΔS_{mix}	Mixing entropy	σ_ρ	Density deviation of constituent TMCs
$\Delta H_{mix}/\Delta S_{mix}$	Mixing Gibbs free energy parameter per formula unit	$\bar{\chi}$	Average electronegativity of constituent TMCs per formula unit
$\bar{V}$	Average volume of constituent TMCs per formula unit	σ_χ	Deviation of electronegativity of constituent TMCs
σ_V	Volume deviation of constituent TMCs	VEC	Valence electron concentration (VEC) of HECC candidates
$\bar{m}$	Average mass of constituent TMCs per formula unit	σ_{VEC}	VEC deviation of constituent TMCs

chemical attributes and CALPHAD features. The obtained results agree with DFT calculations and experiments, indicating the applicability of the proposed ML model in predicting the EFA parameter for HECCs. Other than this study, there are still no ML models available for predicting single-phase HECCs directly. Therefore, an ML model that can dictate the phase information of HECCs based on the features of possible HECC candidates and their constituent TMCs is highly imperative. Such a model will significantly promote and stimulate the development and design of novel single-phase HECCs.

In this work, we develop ML models to predict the single-phase probability of HECCs based on the chemical attributes of HECC candidates and their constituent binary TMCs. With the precursory information and parameters obtained from DFT calculations, our well-trained SVM and ANN models can predict single-phase HECCs in which the cations are from either IV, V, or VI groups. Although it is well established that Group IV and V TMCs can readily form single-phase HECCs with improved mechanical properties, the phase stability becomes complicated after the introduction of Group VI metals, such as Cr, Mo, and W. Nevertheless, due to the higher valence filling in Group VI metals, it is expected that the incorporation of these elements can further enhance the performance of HECCs[48–51]. Our ML model thus can be used for fast screening of the possible elemental combinations that can form single-phase HECCs. Several of our model predictions are validated by the current experiments, through which new HECCs are successfully discovered and characterized. We further demonstrate that the single-phase formation probability of non-equiatomic HECCs can be quickly evaluated based on our refined ML model. The agreement between our model and experiments suggests that ML is a promising way for rapidly assessing and designing new HECC materials.

RESULTS

Feature selection

After a careful investigation of elemental attributes and the properties of TMCs and HECCs, the descriptors that we choose for our ML models are summarized in Table 1.

To reduce possible overfitting introduced by the strongly correlated features and increase fitting efficiency, we remove those features with Pearson coefficients higher than 0.90. Specifically, the Pearson correlation (r) is calculated by:

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}}, \tag{1}$$

where x_i and y_i are the i-th value of two different input features, respectively; $\bar{x}$ and $\bar{y}$ are the expectations of the two input features. As shown in Fig. 1, the Pearson coefficients between ΔH_{mix} and $\Delta H_{mix}/\Delta S_{mix}$, as well as the average electronegativity ($\bar{\chi}$) and VEC are greater than 0.90. Thus, the Gibbs free energy parameter ($\Delta H_{mix}/\Delta S_{mix}$) and the average electronegativity are excluded in the following training procedure. To converge the ML
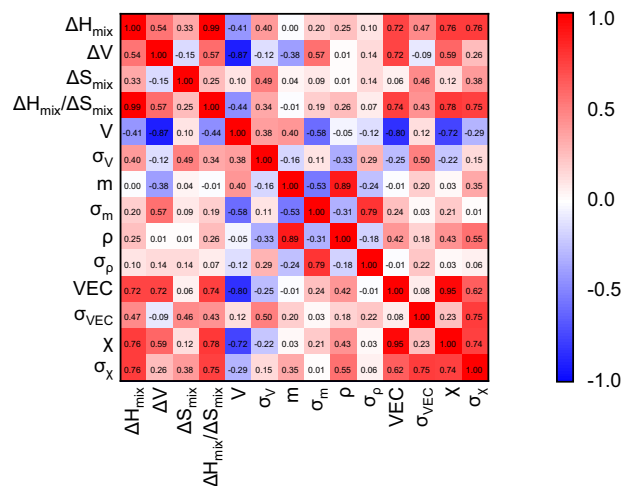

Fig. 1 Feature selection for ML models. Pearson correlation of input features is calculated by Eq. (1).

models faster, we further scale different features into the same range (0–1) with:

$$x_{scaled} = \frac{x_i - x_{min}}{x_{max} - x_{min}} \tag{2}$$

where x_{max}, x_{min}, x_i are the maximum, minimum, and i-th values of the feature x.

Hyperparameter optimization

Two ML models, i.e., ANN and SVM, are adopted in this study because of their excellent performance and flexibility in predicting new materials[40–42,44–46]. For the ANN model, we have optimized the numbers of cross-validation folds, hidden layers, and nodes per hidden layer. A larger number of folds generally uses the input data more efficiently. However, there is no guarantee that training with more folds will lead to better performance on prediction. Additionally, the numbers of hidden layers and nodes per layer determine how many details of the input samples can be learned. The ANN model may only learn inadequate knowledge with insufficient model depth (number of hidden layers). Conversely, unnecessary noise will be induced into the model with too many hidden layers and nodes per layer. For the SVM model, the numbers of cross-validation folds, kernel type, and regularization parameter (C) are optimized. Among these, the C parameter specifies the acceptability of the misclassification; a large C chooses a small-margin hyperplane for classifying the input vectors. To evaluate the performance of these two models, the validation accuracy (acc) is calculated by:
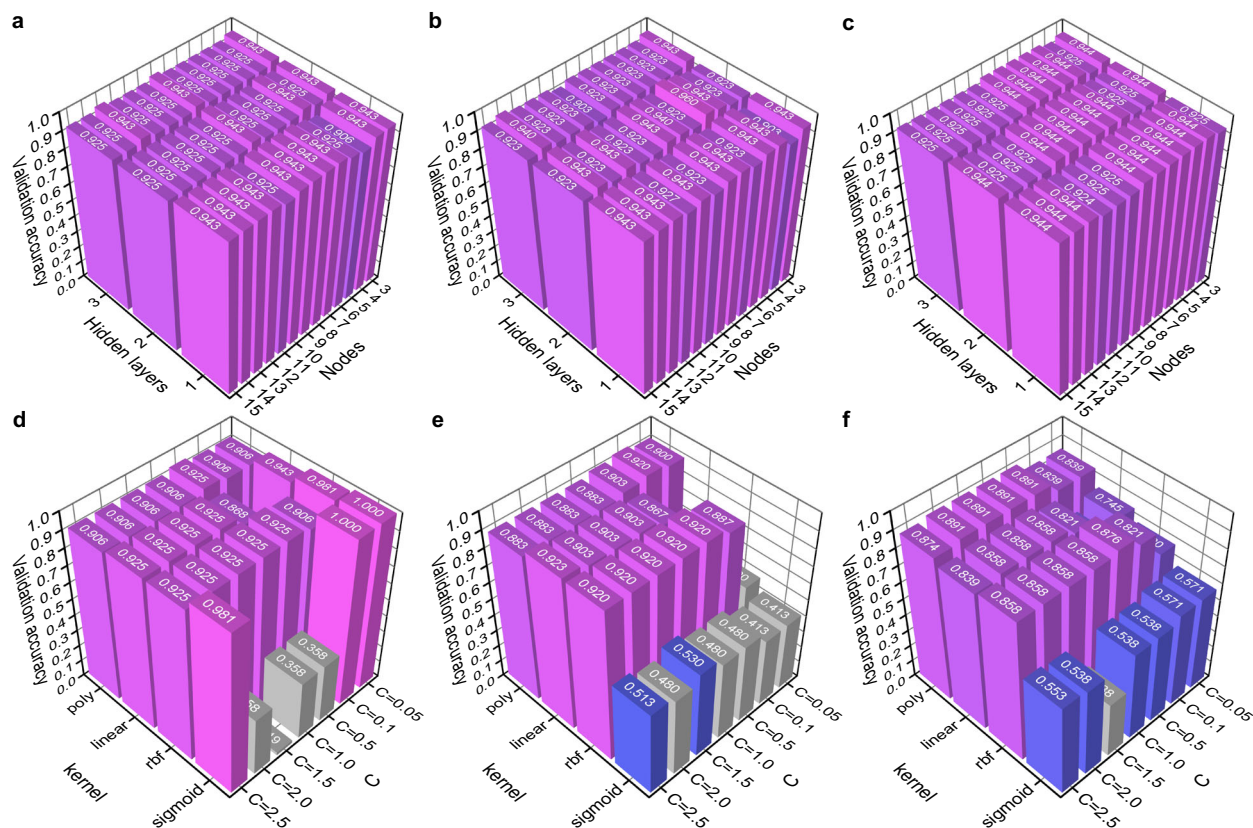
$$acc = \frac{1}{N}\sum acc_i, \tag{3}$$

Fig. 2 Hyperparameters optimization of ML models. Hyperparameters of **a**–**c** ANN and **d**–**f** SVM models, respectively. The first, second, and third columns refer to the LOO, 10-fold, and 5-fold methods, respectively. The validation accuracy is calculated by Eq. (3).

where N and acc_i are the number of folds and the validation accuracy of each fold, respectively. The leave-one-out (LOO), 10- and 5-fold cross-validation methods were deployed to measure the performance of considered ML models. Specifically, the dataset is split into N subsets in the N-fold cross-validation method. By using different subgroups as the test set, N independent models are trained and tested separately. In each model, the variables are trained on $N − 1$ subgroups and tested on the remaining group. For example, 10- and 5-fold cross-validation methods correspond to $N = 10$ and 5, indicating that the validation is performed within ten and five machine learning models, respectively. In the LOO method, the number of folds is equal to the dataset size, and only one data is left in the validation set; while the rest data are included in the training set.

Results in Fig. 2 show that the validation accuracies of ANN models are comparable among different numbers of folds. The best validation accuracy of the ANN model is calculated to be 0.960 (Fig. 2b) with two hidden layers and six neurons in each layer by 10-fold cross-validation. Though SVM models with low C parameters or sigmoid kernel give a high validation accuracy, the number of support vectors is also high (Supplementary Fig. 1), signifying the risk of overfitting. Therefore, if a compromise between the validation accuracy and the number of support vectors is made, the LOO method with the linear kernel function and a C parameter of 1.0 is selected for SVM models, of which the validation accuracy is 0.925.

The confusion matrix, receiver operating characteristic (ROC) curve, and the area under this curve (AUC) are widely used to evaluate the performance of a classifier. In specific, by plotting the true positive rate (TPR) against the false positive rate (FPR) with varied discrimination thresholds, the top-left ROC curve with a larger AUC value generally indicates a better statistical performance[52]. The confusion matrices of these two models exhibit similar patterns (Fig. 3a, b). As indicated by the summation of TPR and true negative rate (TNR), the final accuracy obtained with well-trained ANN and SVM models are 0.982 and 0.944, respectively. Besides, Fig. 3c displays that ANN shows superior classification quality with an AUC score of 0.956 compared to 0.940 of SVM. Therefore, from the above evaluations, ANN shows better performance than SVM.

Prediction

The well-trained models are used to identify the single-phase formation ability of novel HECC candidates. The predictions on the single-phase probability of HECCs are summarized in Table 2. If the classification criterion is set to be 0.5, 38 out of 90 samples are identified as single-phase HECCs by comparing the averaged probability with the classification criterion. Note that all these HECCs have not been experimentally reported before. From the prediction results, we selected 13 HECC candidates with high and low single-phase probabilities to validate our ML models through experiments. As shown from the X-ray diffraction (XRD) spectra in Fig. 4, nine candidates, including $(HfMoNbTaV)C_5$, $(MoNbTaVZr)C_5$, $(HfMoNbTaZr)C_5$, $(CrHfNbTaV)C_5$, $(CrHfNbTaTi)C_5$, $(NbTaVWZr)C_5$, $(MoNbTaTiW)C_5$, $(HfNbTaVW)C_5$, and $(CrHfTaTiZr)C_5$ exhibit single phase with face-centered cubic (FCC) structures, which agrees well with the predictions from our ML models. On the other hand, peaks of the other four samples cannot be indexed to a single FCC phase; specifically, those of $(CrHfMoVW)C_5$ and $(CrMoTiVZr)C_5$ contain two and those of $(CrHfTaVW)C_5$, and $(MoTiVWZr)C_5$ contain three sets of FCC peaks; indicating the presence of multiple types of FCC carbides in these four materials. These observations also agree well with the current ANN and SVM predictions. In Table 2, if there are discrepancies in predictions from the ANN and SVM models, we prefer to rely on the ANN
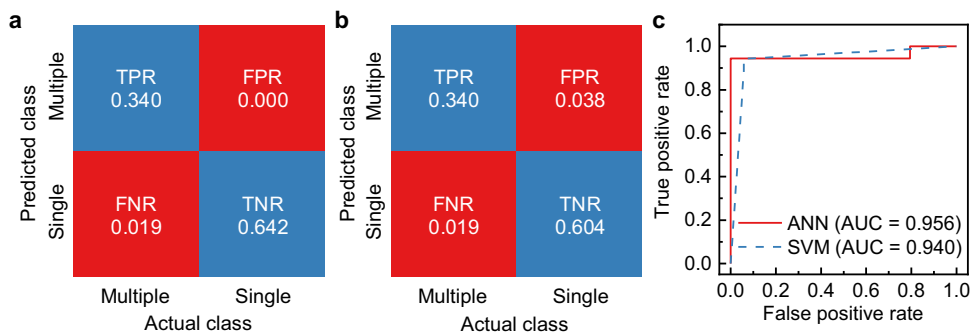
Fig. 3 Performance of ML models. Confusion matrix of **a** ANN and **b** SVM models. **c** ROC curves and AUC scores of ANN and SVM models. TPR, FPR, FNR, and TNR refer to the true positive rate, false positive rate, false negative rate, and true negative rate, respectively.

model due to its higher accuracy, as demonstrated in Fig. 3. The better reliability of the ANN model is also validated by the experimental result of $(CrHfTaTiZr)C_5$.

It is known that metallic elements in groups IV and V can form stable rock-salt binary TMCs with carbon[50,51,53]. In contrast, for those in group VI, their binary carbides usually exhibit different structures at room temperature[54,55]. Based on our ML predictions in Table 2, we have made a statistic of the occurrence of different elements in single-phase HECCs. As provided in the Supplementary Fig. 2, our results reproduce the higher tendency of finding IV/V elements in single-phase HECCs than groups VI. Although group VI elements (Cr, Mo, and W) tend to appear in multi-phase HECCs, they can indeed form single-phase HECCs with suitable combinations of metallic elements from those in groups IV and V, as demonstrated in current experiments and previous studies[47]. As such HECCs usually exhibit superb mechanical performance, accurate prediction of their phase formation is of great significance for practical applications.

To elucidate the governing factors on phase formation, we further evaluate the relative importance of different input features (Δacc) through dropping the input features successively, which can be calculated as:

$$\Delta acc_i = acc_{i,\text{after}} - acc_{i,\text{before}}, \tag{4}$$

where $acc_{i,\text{after}}$ and $acc_{i,\text{before}}$ are the validation accuracy before and after the dropping of the i-th feature, respectively. In general, an important input feature will lead to a significant decrease in accuracy after dropping. The results are shown in Fig. 5, which suggests that σ_{VEC} of constituent TMCs plays the most crucial role in both the ANN and SVM models, as the most significant decrease in Δacc is seen when σ_{VEC} is excluded. In addition, σ_{VEC} is more critical in SVM than in ANN. The top five most important features are listed in Fig. 5c. Except for the mixing entropy, the input features with higher importance are all found to be the differences among properties of constituent binary TMCs rather than their averaged values, which indicates that the ML models are more sensitive to the dissimilarities among precursory TMCs. Nevertheless, Fig. 5d shows that HECCs cannot be differentiated solely by the top few features, including σ_{VEC}, ΔS_{mix}, and σ_χ, emphasizing the superiority of the developed ML tools. This observation also demonstrates the complex relationships among the descriptor variables, which cannot be clearly elucidated by empirical rules-based models.

The VEC parameters have been extensively used in high-entropy alloys to distinguish different phases such as FCC and BCC[56]. Here in HECCs, our results indicate that the absolute values of VEC are less effective in dictating phase formation. Instead, it is the relative differences among different precursors that play decisive roles. Compared to the previous ML model on EFA prediction[47], we find similar importance of σ_{VEC} and σ_χ in determining the phase stabilities of HECCs. These parameters

govern the bonding properties and local atomic environments surrounding metallic elements in HECCs, exerting the most significant influence on the tendency of phase formation or separation.

From equiatomic to non-equiatomic HECCs

Right now, in almost all the reported HECCs, the concentrations of metallic elements are equiatomic or nearly equiatomic. As found in other high-entropy materials, non-equiatomic compositions may exhibit better performance than the equiatomic ones[57–60]. However, the single-phase possibilities of non-equiatomic HECCs are not exploited yet. The exploration from equiatomic to non-equiatomic provides more options for tuning the mechanical and chemical properties of HECCs[57–60].

As demonstrated in Fig. 5, the most important features influencing phase formation are the distributions of VEC, electronegativity, mass, and density, as well as the mixing entropy. All these properties are from the TMC constituent precursors. Therefore, we retrain ML models using descriptors without DFT-calculated mixing energies but only relying on the properties of HECC constituents. The performance of retrained ML models (Supplementary Fig. 3) is very close to the above-discussed models built on all considered descriptors, indicating the good robustness of the newly trained models for high-throughput prediction. Such ML models are extremely useful since they do not require computationally expensive DFT calculations. Besides, the model can make predictions for arbitrary combinations of binary TMCs with different concentrations, which allows the evaluation of the single-phase stability for non-equiatomic HECCs that have different stoichiometric ratios of anions to cations.

To demonstrate the capability of our proposed ML models, we show the predicted phase diagram for (VNbTa)- and (CrMoW)-based HECCs in Figs. 6 and 7, respectively. These two base combinations are chosen as they belong to the group V and VI, respectively. According to our analysis, the three group V elements have the highest probability to appear in single-phase HECCs, whereas the three elements in group VI tend to be found in multi-phase HECCs. However, both single-phase and multi-phase multicomponent HECCs are found when extra elements are incorporated into these two base compositions.

The results in Figs. 6 and 7 are averaged from ANN and SVM predictions, where the single-phase probability is given as colored contour plots. The phase diagrams by ANN and SVM are plotted separately in Supplementary Figs. 4–8. In Fig. 6, the experimentally verified single-phase equiatomic compositions are represented by the triangular and circular labels. Generally, the probability of finding single-phase HECCs with a high concentration of (VNbTa) C_3 is high (the rightmost corner). As shown in Fig. 6a–c, the solubility of Mo/Hf, Mo/Zr, and Cr/Hf carbides in (VNbTa)-based HECCs are similar as these phase diagrams show a similar pattern. In comparison, the Zr/W and Hf/W carbides are relatively difficult

Table 2. Results predicted by current ML models.

Material	ANN	SVM	Expt.	EFA	Material	ANN	SVM	Expt.	EFA
$(CrNbTaTiV)C_5$	0.95	1.00		92[1]	$(CrMoNbVZr)C_5$	0.02	0.68		81[1]
$(CrMoNbTaV)C_5$	0.94	1.00		106[1]	$(CrMoNbTaZr)C_5$	0.47	0.15		74[1]
$(CrNbTaVZr)C_5$	0.94	1.00		88[1]	$(CrHfNbTiZr)C_5$	0.54	0.08		74[1]
$(CrNbTaVW)C_5$	0.94	1.00		100[1]	$(CrHfMoNbTa)C_5$	0.38	0.13		78[1]
$(HfMoNbTaV)C_5$	0.93	1.00	Single	71[2]	$(CrHfNbVZr)C_5$	0.38	0.06		80[1]
$(CrNbTaTiZr)C_5$	0.93	1.00		85[1]	$(HfNbVWZr)C_5$	0.43	0.00		45[2]
$(MoNbTaVZr)C_5$	0.93	1.00	Single	59[2]	$(MoTaVWZr)C_5$	0.40	0.00		48[2]
$(HfMoNbTaZr)C_5$	0.93	1.00	Single	71[2]	$(MoNbVWZr)C_5$	0.36	0.04		48[2]
$(CrHfNbTaV)C_5$	0.93	1.00	Single	94[1]	$(CrMoNbTiW)C_5$	0.33	0.00		78[1]
$(CrHfNbTaTi)C_5$	0.92	1.00	Single	91[1]	$(CrMoTaTiW)C_5$	0.32	0.00		85[1]
$(NbTaVWZr)C_5$	0.92	1.00	Single	56[2]	$(CrNbTaWZr)C_5$	0.27	0.00		59[1]
$(MoNbTaTiW)C_5$	0.91	1.00	Single	77[2]	$(MoTaTiWZr)C_5$	0.23	0.00		48[2]
$(HfNbTaVW)C_5$	0.90	1.00	Single	67[2]	$(HfTiVWZr)C_5$	0.23	0.00		45[2]
$(HfMoTaTiV)C_5$	0.90	1.00		67[2]	$(CrMoTaVZr)C_5$	0.21	0.00		70[1]
$(CrHfNbTaZr)C_5$	0.90	1.00		74[1]	$(HfMoTaTiW)C_5$	0.11	0.08		53[2]
$(HfNbTaWZr)C_5$	0.88	1.00		59[2]	$(HfMoTiVZr)C_5$	0.18	0.00		50[2]
$(MoTaTiVZr)C_5$	0.88	1.00		59[2]	$(CrHfMoTaV)C_5$	0.18	0.00		73[1]
$(MoTaTiVW)C_5$	0.87	1.00		67[2]	$(HfMoNbVW)C_5$	0.17	0.00		56[2]
$(MoNbTiVW)C_5$	0.86	1.00		71[2]	$(MoNbTiWZr)C_5$	0.16	0.00		48[2]
$(CrMoNbTaTi)C_5$	0.85	1.00		96[1]	$(CrHfNbTaW)C_5$	0.14	0.00		63[1]
$(HfMoTaVZr)C_5$	0.83	1.00		50[2]	$(HfMoNbTiW)C_5$	0.11	0.00		53[2]
$(CrMoTaTiV)C_5$	0.83	1.00		96[1]	$(CrHfTiVZr)C_5$	0.10	0.00		73[1]
$(NbTiVWZr)C_5$	0.81	1.00		50[2]	$(CrMoNbTiZr)C_5$	0.02	0.06		67[1]
$(MoNbTaWZr)C_5$	0.79	1.00		63[2]	$(CrHfTaVW)C_5$	0.07	0.00	Multiple	55[1]
$(HfMoNbTiZr)C_5$	0.74	1.00		67[2]	$(CrHfMoTaTi)C_5$	0.06	0.00		72[1]
$(CrHfTaTiV)C_5$	0.89	0.83		89[1]	$(MoTiVWZr)C_5$	0.06	0.00	Multiple	40[2]
$(HfMoNbTiV)C_5$	0.72	1.00		71[2]	$(CrMoTaTiZr)C_5$	0.06	0.00		71[1]
$(CrHfNbTiV)C_5$	0.79	0.91		85[1]	$(CrHfMoNbV)C_5$	0.05	0.00		80[1]
$(HfTaTiVW)C_5$	0.75	0.92		56[2]	$(HfMoNbWZr)C_5$	0.05	0.00		48[2]
$(HfNbTiWZr)C_5$	0.76	0.91		53[2]	$(HfMoTiVW)C_5$	0.04	0.00		45[2]
$(HfMoNbTaW)C_5$	0.58	1.00		71[2]	$(CrMoTaWZr)C_5$	0.03	0.00		54[1]
$(HfMoNbVZr)C_5$	0.45	1.00		50[2]	$(CrMoNbWZr)C_5$	0.03	0.00		62[1]
$(TaTiVWZr)C_5$	0.83	0.53		50[2]	$(CrHfMoTaZr)C_5$	0.03	0.00		61[1]
$(MoNbTiVZr)C_5$	0.34	1.00		59[2]	$(CrHfTaTiW)C_5$	0.03	0.00		56[1]
$(HfNbTiVW)C_5$	0.72	0.55		59[2]	$(CrHfNbTiW)C_5$	0.03	0.00		55[1]
$(CrMoNbTiV)C_5$	0.25	1.00		94[1]	$(CrHfMoNbTi)C_5$	0.03	0.00		71[1]
$(CrNbTiVZr)C_5$	0.15	1.00		85[1]	$(CrMoVWZr)C_5$	0.02	0.00		62[1]
$(CrTaTiVZr)C_5$	0.88	0.25		87[1]	$(CrHfMoTaW)C_5$	0.02	0.00		59[1]
$(HfTaVWZr)C_5$	0.46	0.47		43[2]	$(CrHfMoTiV)C_5$	0.02	0.00		69[1]
$(CrNbTiVW)C_5$	0.77	0.08		73[1]	$(CrHfMoNbW)C_5$	0.02	0.00		53[1]
$(CrHfTaTiZr)C_5$	0.79	0.02	Single	78[1]	$(CrHfMoVW)C_5$	0.02	0.00	Multiple	61[1]
$(CrNbTaTiW)C_5$	0.77	0.04		69[1]	$(CrHfMoNbZr)C_5$	0.02	0.00		63[1]
$(CrTaTiVW)C_5$	0.76	0.00		68[1]	$(CrHfMoVZr)C_5$	0.01	0.00		66[1]
$(HfMoTaVW)C_5$	0.22	0.53		56[2]	$(CrHfMoTiZr)C_5$	0.01	0.00		58[1]
$(CrHfTaVZr)C_5$	0.69	0.02		78[1]	$(CrMoTiVZr)C_5$	0.01	0.00	Multiple	66[1]

[1] EFA values adopted from reference[47].
[2] EFA values adopted from reference[12].
The fractions denote the predicted single-phase probability of HECC candidates. The phases observed in the current experiments are denoted.

to be incorporated into (VNbTa)-based HECCs to form single-phase carbides. Though the rock-salt binary TMCs with cations in group VI are unstable, they can form single-phase HECCs by combining with (VNbTa) base (Fig. 6f, h, i). Interestingly, the combinations of $(Mo,Nb,Ta,V,W)C_5$ can be stable in a single phase within a broad range of concentrations of constituent precursors (Fig. 6f). Therefore, it would be instructive to explore the HECCs composed of V, Nb, Ta, Mo, and W elements through experiments.
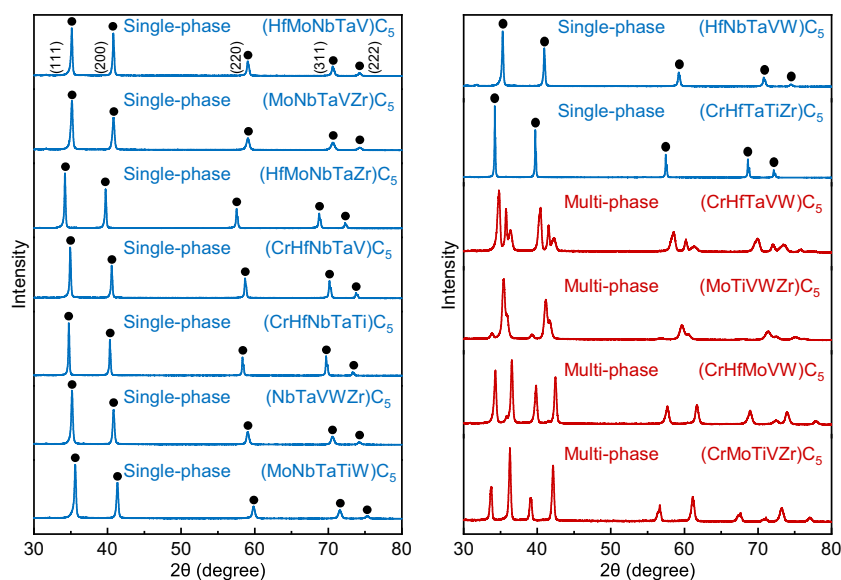
Fig. 4 X-ray diffraction patterns for experimentally validated HECCs. Nine materials including $(HfMoNbTaV)C_5$, $(MoNbTaVZr)C_5$, $(HfMoNbTaZr)C_5$, $(CrHfNbTaV)C_5$, $(CrHfNbTaTi)C_5$, $(NbTaVWZr)C_5$, $(MoNbTaTiW)C_5$, $(HfNbTaVW)C_5$, and $(CrHfTaTiZr)C_5$ are single-phase HECCs with one FCC phase; others are observed with multiple FCC peaks. The HECCs shown in the figure are in the order of the averaged single-phase probability based on the current ML models.
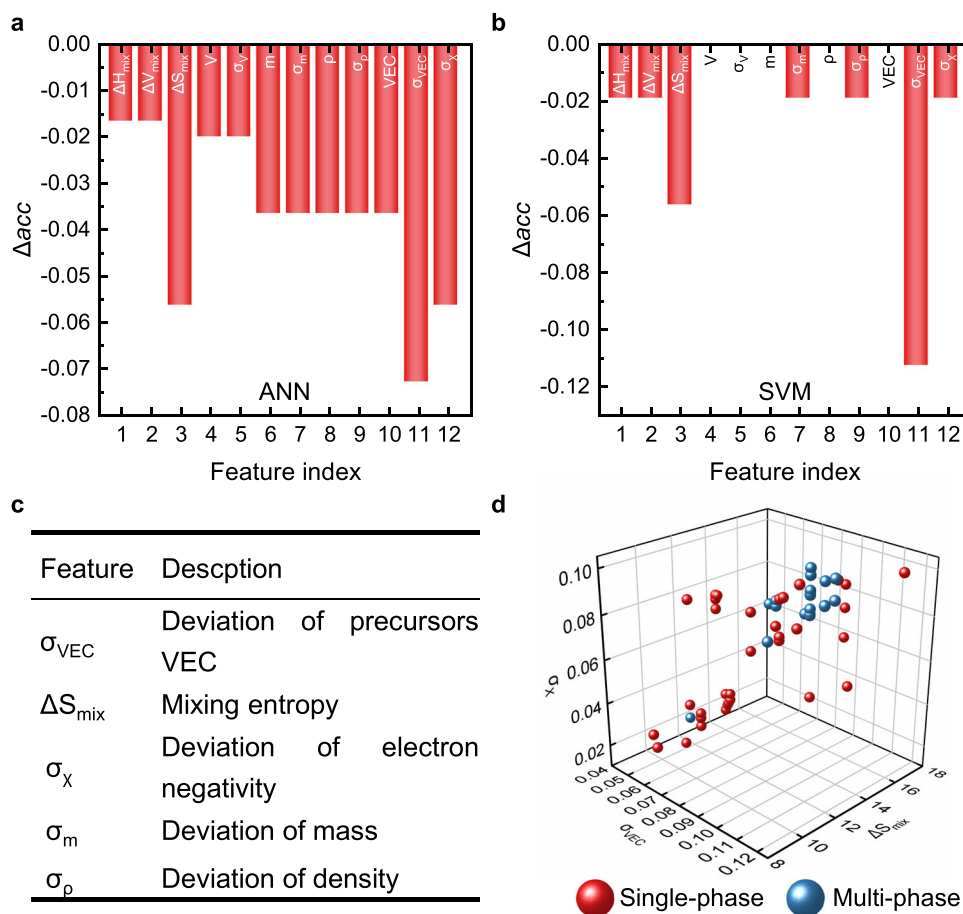

Feature	Descption
σ_{VEC}	Deviation of precursors VEC
ΔS_{mix}	Mixing entropy
σ_χ	Deviation of electron negativity
σ_m	Deviation of mass
σ_ρ	Deviation of density

Fig. 5 Relative importance of different features. Feature importance of a ANN and b SVM models. c Top five most important features. d 3D plot of top three most important features, including σ_{VEC}, ΔS_{mix}, and σ_χ, against the phase classification.
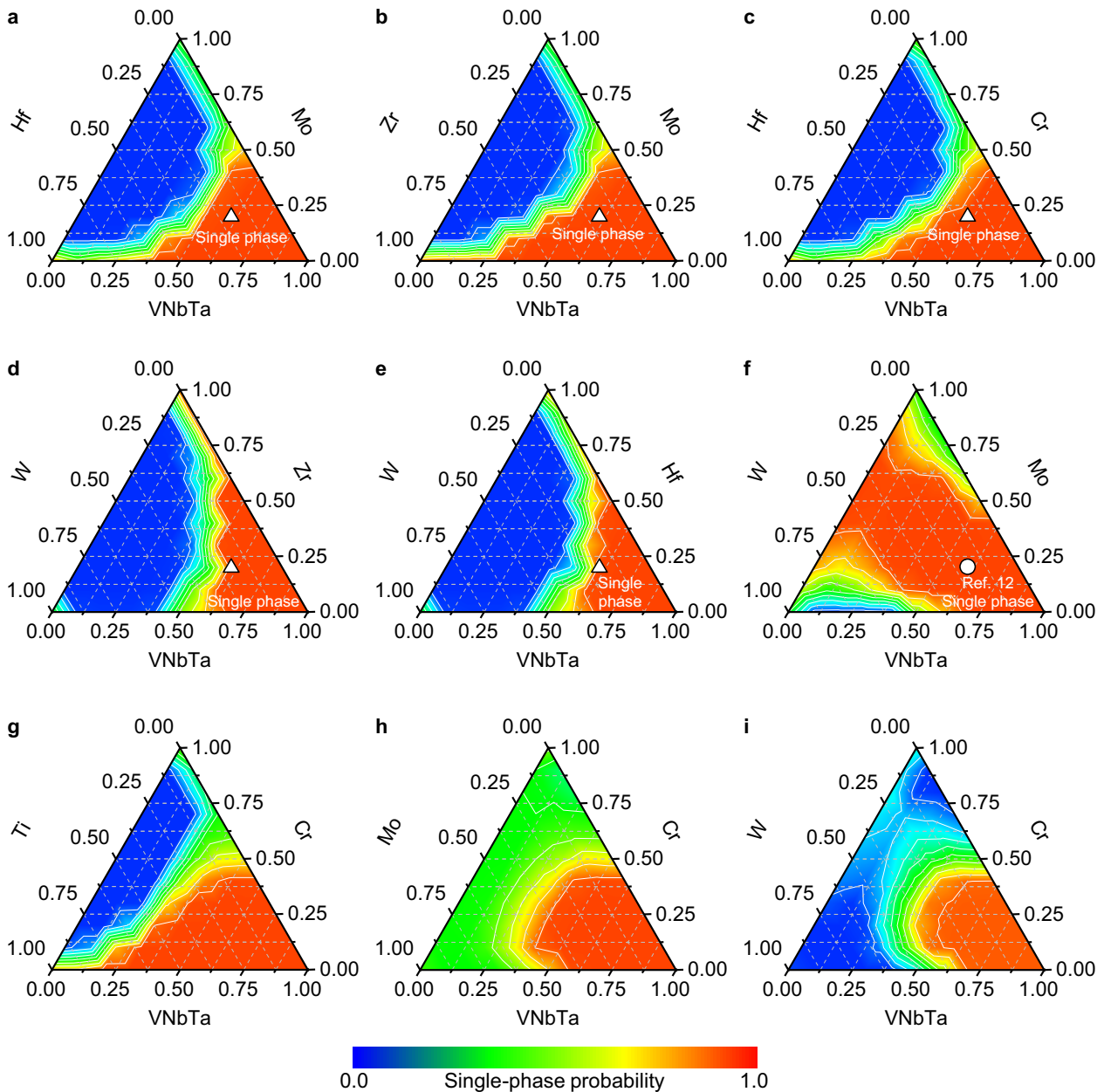
Fig. 6 Phase diagrams of (VNbTa)-based HECCs from ML. a $(Hf, Mo, Nb, Ta, V)C_5$. **b** $(Mo, Nb, Ta, V, Zr)C_5$. **c** $(Cr, Hf, Nb, Ta, V)C_5$. **d** $(Nb, Ta, V, W, Zr)C_5$. **e** $(Hf, Nb, Ta, V, W)C_5$. **f** $(Mo, Nb, Ta, V, W)C_5$. **g** $(Cr, Nb, Ta, Ti, V)C_5$. **h** $(Cr, Mo, Nb, Ta, V)C_5$. **i** $(Cr, Nb, Ta, V, W)C_5$. Numbers in each axis denote the atomic percentages in the cation sublattice. Three elements are included in the horizontal axis with equiatomic ratios. Scattered data with triangular and circular labels are experimentally validated by this work and ref. [12], respectively.

For the (CrMoW)-based HECCs, the single-phase structure can be achieved through incorporating binary carbides with cations in group V (Fig. 7a–c). Though equiatomic HECCs of $(CrMoNbVW)C_5$, $(CrMoNbTaW)C_5$, and $(CrMoTaVW)C_5$ have been experimentally determined to be single phase[47], there is still a large area predicted as single phases in the non-equiatomic region, which deserves further experimental exploration. In contrast to (VNbTa)-based HECCs, if the cation elements in group IV are included in (CrMoW)-based systems, the phase space for single-phase HECCs is relatively narrow (Fig. 7d–f).

DISCUSSION

Since the discovery of high-entropy materials, their phase formation abilities have been a field of intensive research. While the inclusion of multiple principal elements allows effectively tuning the properties of this class of materials, the extreme chemical disorder presents significant obstacles for predicting their stable phases. Generally, the mixing Gibbs free energy is widely accepted to be the driving force of phase formation. Besides that, the mixing enthalpy also plays a key role in governing the phase stabilities[61,62]. In order to check the feasibility of identifying single-phase HECCs with these two parameters, we plot their values against the sample indices in Fig. 8. Here the Gibbs free energy is represented by $\Delta H_{mix}/\Delta S_{mix}$ as the synthesis temperature of HECCs is not available. It is suggested that no clear threshold of $\Delta H_{mix}/\Delta S_{mix}$ or ΔH_{mix} can be drawn between the single-phase and multi-phase samples. For instance, though most single-phase HECCs exhibit a mixing enthalpy below 0.2 eV per formula unit, the mixing enthalpies of multi-phase $(HfTaTiZr)C_4$[31]
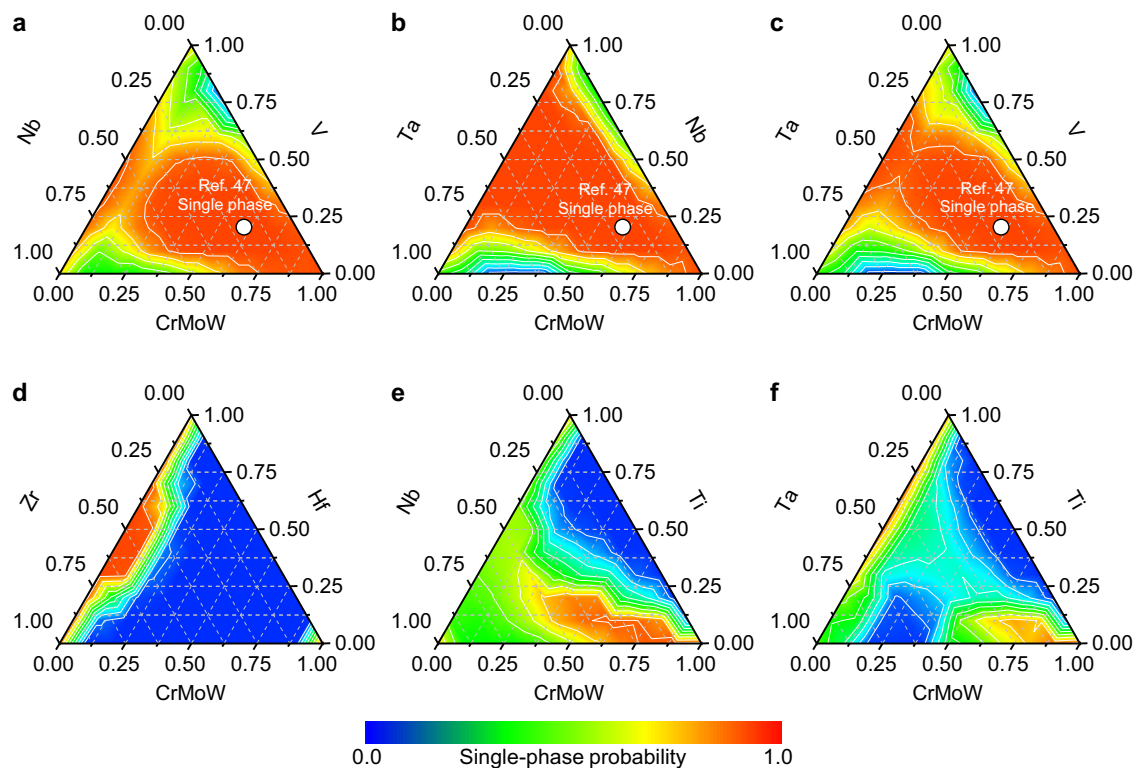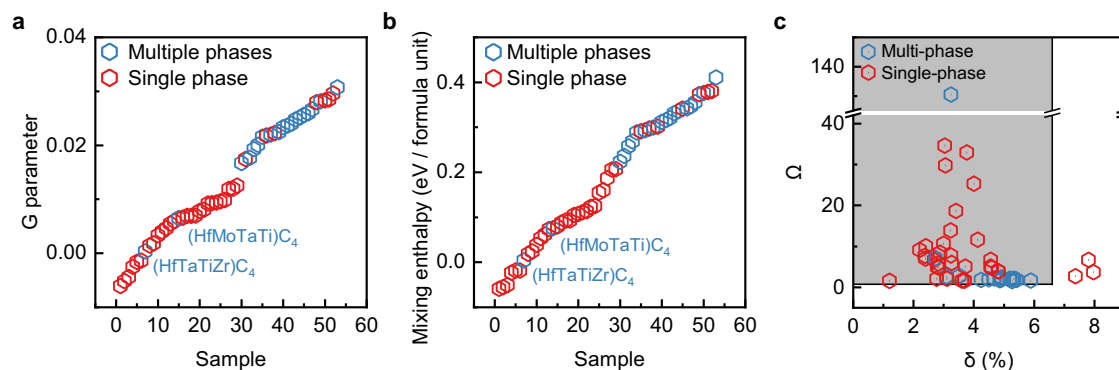
Fig. 7 Phase diagrams of (CrMoW)-based HECCs from ML. a (Cr, Mo, Nb, V, W)C$_5$. **b** (Cr, Mo, Nb, Ta, W)C$_5$. **c** (Cr, Mo, Ta, V, W)C$_5$. **d** (Cr, Hf, Mo, W, Zr)C$_5$. **e** (Cr, Mo, Nb, Ti, W)C$_5$. **f** (Cr, Mo, Ta, Ti, W)C$_5$. Numbers in each axis denote the atomic percentages in the cation sublattice. Three elements are included in the horizontal axis with equiatomic ratios. Scattered data with circular labels are results from ref. [47].

Fig. 8 Conventional descriptors for HECCs. a Gibbs free energy parameter, **b** mixing enthalpy, and **c** Ω–δ map.

and (HfMoTaTi)C$_4$[63] are found to be 0.004 and 0.075 eV per formula unit, respectively.

Another important empirical rule proposed in high-entropy materials is the Ω-δ criterion[64], where δ is the mean square deviation of the lattice constant that can be calculated as:

$$\delta = \sqrt{\sum_1^n c_i \left(1 - \frac{a_i}{\bar{a}}\right)^2}. \qquad (5)$$

Here c_i and $\bar{a}$ are the concentration i-th precursory carbide and the average lattice constant, respectively. Ω is defined by:

$$\Omega = \frac{T_m \Delta S_{mix}}{|\Delta H_{mix}|}, \qquad (6)$$

where T_m is the melting point, which is set to be 4,500 K as T_m of most HECCs is in the range of 3,750–5,000 K using the rule of mixture[65]. The Ω–δ criterion has been successful in predicting

high-entropy alloys, where solid solutions are expected for $\Omega \geq 1.1$ and $\delta \leq 6.6\%$[64]. Very recently, Liu et al. extrapolated this Ω–δ criterion to the HECC systems[65]. By comparison, the calculated Ω–δ map for the collected samples in the present study is summarized in Fig. 8c, with the expected single-phase region shown in the shaded area. It is obvious that the Ω–δ criterion can only partly predict the phase formation for HECCs. For example, the experimentally synthesized single-phase (TiZrHfTaLaY)C$_6$, (TiZrHfTaNbLaY)C$_7$, and (TiZrHfTaNbMoWLa)C$_8$ exhibit[66] large volume misfit (δ) of near 8%, which is greater than the proposed criterion[64,65].

From the above discussion, one can see that the empirical rules have their limitations in the HECC field. In comparison, the ML models are able to mine the complex relationships among descriptor variables in a high-dimensional space, manifesting superior advantages for discovering the complex high-entropy ceramics. Particularly, our ML models can predict phase formation

for both equiatomic and non-equiatomic HECCs, which greatly expands the search regime for high-performance ceramics based on the entropy concept. Notably, we find that there are large single-phase regions that are not explored so far, especially for HECCs composed of Group VI metals. Verification of these novel HECCs would lead to significantly enhanced material performance due to its increased number of valence electrons. Considering that the TMC precursors from Group VI metals are orthorhombic Cr_3C_2, hexagonal Mo_2C, and hexagonal WC, it is challenging to envisage the phase stability of HECCs comprised of these TMCs from experience. Therefore, the information on non-equiatomic compositions is highly instructive.

In summary, we successfully build ML models (ANN and SVM) with high accuracy of up to 0.982 to predict the single-phase formation tendency of novel HECCs. Our results indicate that the absolute values of VEC, which is an important parameter to determine the dominant phase in high-entropy alloys, are less effective here to dictate the phase formation in HECCs. Among the assessed descriptors, σ_{VEC} and σ_χ are found to impart the most significant influence on the phase formation of HECC. Using these well-trained models, we study the single-phase probability of ~90 un-synthesized HECCs containing five cations from groups IV, V, and VI metals, among which 38 single-phase HECCs are predicted. The predictions are in good agreement with current experimental findings. Specifically, nine of the 13 experimentally investigated HECCs, namely $(HfMoNbTaV)C_5$, $(MoNbTaVZr)C_5$, $(HfMoNbTaZr)C_5$, $(CrHfNbTaV)C_5$, $(CrHfNbTaTi)C_5$, $(NbTaVWZr)C_5$, $(MoNbTaTiW)C_5$, $(HfNbTaVW)C_5$, and $(CrHfTaTiZr)C_5$, are verified as single phases with rock-salt structures. Through training solely on the properties of constituent precursors, our ML models are able to predict the phase formation probabilities of non-equiatomic HECCs. Our obtained diagrams of (VNbTa)- and (CrMoW)-based HECCs show large unexploited regimes where non-equiatomic single phases are expected. Our developed ML models can accelerate the discovery of equiatomic and non-equiatomic HECCs, which opens the avenue for rational HECC design within the immersive phase space, thus allowing for effectively tuning the properties of HECCs.

METHODS
Encoding samples
The training dataset, including 34 single-phase and 19 multi-phase carbides, was collected from available publications (Supplementary Table 1). As shown in Fig. 9, the cations of most collected samples are in the IV, V, VI groups of the periodic table. To train the ML models, samples with pure phase and mixed phases were labeled as "0" and "1", respectively. The k-fold cross-validation method was adopted to evaluate the performance of ML models. In such a procedure, the dataset was split into k parts. In particular, the LOO method was deployed when the number of folds (k) was equal to the number of samples.

The samples were represented by encoding the properties of possible HECC candidates and their constituent TMCs with 14 input features (Table 1), some of which were proven effective in predicting other high-entropy materials[40–46]. The total energies and geometrical structures of binary TMCs were optimized by DFT calculations. The space group and calculated properties of precursor TMCs were listed in Supplementary Table 2. In accordance with the stoichiometry of equiatomic high-entropy carbides, the binary carbides used for calculations are all at equiatomic compositions. From the calculated properties of binary TMCs and HECC candidates, the enthalpy and volume changes due to the mixing were computed according to:

$$\Delta prop = prop_{HECC} - \sum_i c_i prop_{TMC,i} \tag{7}$$

where $prop_{HECC}$ and $prop_{TMC,i}$ are the characteristic values of HECC candidates and i-th binary constituent TMC, respectively. c_i is the concentration of i-th binary constituent TMC. The mixing entropy was calculated as:

$$\Delta S = -R \sum_i (c_i \ln c_i) \tag{8}$$

where R and c_i are the molar gas constant and molecular concentration of i-th TMC, respectively. The mean value of a property ($\overline{prop}$) was calculated

Fig. 9 Frequencies of different metallic elements appeared in the collected HECCs. Numbers in this figure denote elemental frequencies.

by:

$$\overline{prop} = \sum_1^n c_i prop_i \tag{9}$$

where $prop_i$ is the value of the i-th constituent TMC. To feed our models with the variation of constituent TMCs, deviation (σ_{prop}) of considered properties was calculated as:

$$\sigma_{prop} = \sqrt{\sum_1^n c_i \left(1 - \frac{prop_i}{\overline{prop}}\right)^2}. \tag{10}$$

Machine learning framework
As depicted in Fig. 10, features were retrieved from the collected samples and transformed into vectors. The input vectors were shuffled and fed into both ANN and SVM models. After training, the phase formation abilities of HECCs were predicted by these two models with well-trained ML structures and weights. At last, selected predictions were validated by experiments. The experimental results can, in turn, be used to enhance the robustness of ML models by adding them to the training dataset in the future.

The LOO, 10- and 5-fold cross-validation methods were deployed to find the best performance of considered ML models. Because of the unbalanced number of single- and multi-phase training samples, more attention is required to the class with fewer samples, i.e., the multi-phase samples. Therefore, we balanced the data by automatically adjusting their weights inversely proportional to their class frequencies. The weight of each class (w_i) was calculated by:

$$w_i = \frac{n_{tot}}{n_{class}n_i}, \tag{11}$$

where n_{tot}, n_i, and n_{class} are the total number of samples, number of classes, and number of i-th samples.

The ANN model was constructed with high-level Keras API as implemented in the TensorFlow package. In this model, we included two densely connected hidden layers with six nodes per layer. In each layer, the output vector was computed by:

$$\mathbf{x}_{l+1} = \sigma(\mathbf{w}\mathbf{x}_l + \mathbf{b}), \tag{12}$$

where $\mathbf{x}_{l+1}$ and $\mathbf{x}_l$ are the feature representation in the $l+1-$ and l-th layer, respectively. The $\mathbf{w}$ and $\mathbf{b}$ in Eq. (12) refer to the weight matrix and bias vector, respectively. $\sigma(\mathbf{z})$ is the softmax function which was used to activate the outputs of perceptrons into a non-linear space. Such a function is given by:

$$\sigma(z_i) = \frac{e^{z_i}}{\sum_i e^{z_i}}, \tag{13}$$

where z_i is the raw output of the i-th neuron in each dense layer. The output layer gives a probability of single- and multi-phase HECCs with two nodes. In the training phase, the cross-entropy loss function was minimized iteratively with Adam optimizer[67]. The loss [$H(p,q)$] between
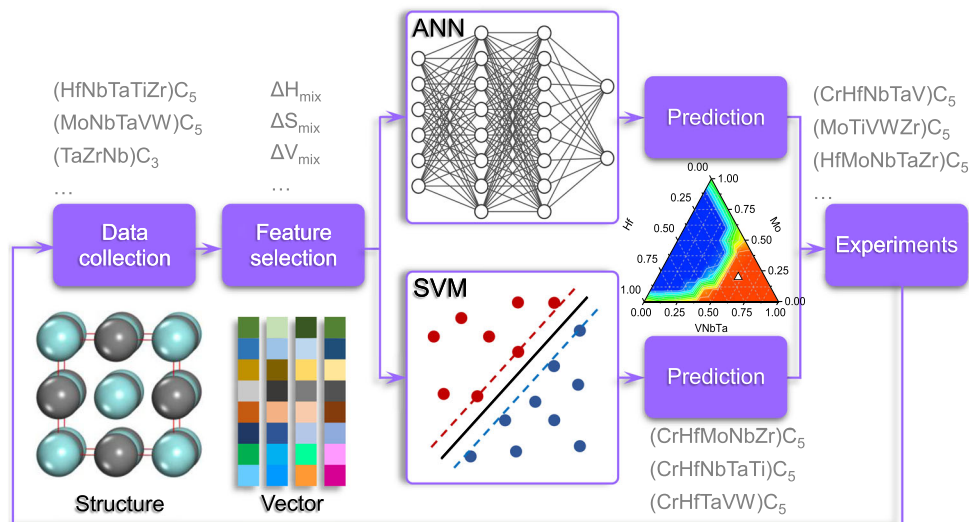
Fig. 10 Machine learning framework. ANN and SVM refer to artificial neural network and support vector machine, respectively.

label and prediction was calculated as:

$$H(p,q) = -\sum_x [p(x)\log q(x)], \quad (14)$$

where $p(x)$ and $q(x)$ are true and predict distribution, respectively.

For the SVM model, we adopted the scikit-learn Python package[68] to fit the best hyperplane ($\mathbf{w}^T x + \mathbf{b} = 0$). The linear kernel was implemented with a C parameter of 1.0. By mapping the input vectors into higher dimensions, SVM can effectively find a good hyperplane to separate the input samples based on the maximum margin strategy.

DFT calculations

The Vienna Ab initio Simulation Package (VASP)[69–71] was adopted to perform electronic structure calculations. We used Perdew-Burke-Ernzerhof (PBE) functional[72] to describe the exchange and correlation interactions in generalized gradient approximation (GGA). The energy cutoff was selected up to 600 eV to converge the energy uncertainty below 1 meV/atom. Automatic k-mesh generation was adopted to build a $3 \times 3 \times 3$ gamma-centered grid. The pseudopotentials recommended by the VASP manual were adopted with the projector augmented wave (PAW) method[73]. Partial occupancies of orbitals were determined by the tetrahedron scheme with Bloch corrections and an energy smearing of 0.1 eV. For the geometrical optimization, the electronic self-consistency was reached when the energy difference was lower than 10^{-6} eV. The ionic optimization was stopped when the force acting on atoms was fewer than 0.01 eV/Å. Spin polarization was taken into consideration in all DFT calculations. For the geometrical optimization of considered carbides, atomic positions, super-cell volume, and supercell shapes were fully relaxed.

Special quasirandom structures (SQS) of HEC carbides were built with Monte-Carlo simulations[74,75]. The random swap was confined within cation sublattice by prohibiting the swap between cations and anions. For each HECC, three independent SQS structures were calculated to converge the variation of lattice constants and energies with respect to the local environments.

Experiments

The raw constituent powders, include graphite (C), TiC (~2.0 μm), VC (~1.5 μm), Cr_3C_2 (~4 μm), ZrC (~1.35 μm), NbC (~1.20 μm), Mo_2C (~2 μm), HfC (~1.10 μm), TaC (~1.10 μm), and WC (~1.25 μm), were weighed and mixed to obtain 13 selected HECCs, including $(HfMoNbTaV)C_5$, $(MoNbTaVZr)C_5$, $(HfMoNbTaZr)C_5$, $(CrHfNbTaV)C_5$, $(CrHfNbTaTi)C_5$, $(NbTaVWZr)C_5$, $(MoNbTaTiW)C_5$, $(HfNbTaVW)C_5$, $(CrHfTaTiZr)C_5$, $(MoTiVWZr)C_5$, $(CrHfMoVW)C_5$, and $(CrMoTiVZr)C_5$. The powder mixtures were ball milled at 250 rpm for 20 h in a cemented carbide jar. The ball (cemented carbide, diameter: 5 mm) to powder mixture was 10:1. Before milling, the jar was filled with Ar gas. Subsequently, the powder mixtures were sintered using the spark plasma sintering (SPS) method (LABOX-325R Spark plasma sintering system, Japan); the sintering temperature was set to 2200 °C for 10 min in a vacuum, and the pressure was maintained at 30 MPa during the sintering process. Noted that the powder mixtures containing low-melting-temperature Cr_3C_2, namely $(CrHfTaTiZr)C_5$, $(CrHfNbTaTi)C_5$, $(CrHfNbTaV)C_5$, $(CrMoTiVZr)C_5$, $(CrHfMoVW)C_5$, and $(CrHfTaVW)C_5$, were first sintered at a lower temperature of 1600 °C to prevent the possible loss of Cr, followed by two-stage annealing (1800 °C for 3 h and 2000 °C for 3 h) to attain the chemical homogeneity. The phase components and lattice parameters of the sintered parts were determined using X-ray diffraction (XRD, Japan, Cu Kα, 40 kV, 40 mA, 5°/min, 30–80° range).

REFERENCES

1. Yeh, J. W. et al. Nanostructured high-entropy alloys with multiple principal elements: novel alloy design concepts and outcomes. *Adv. Eng. Mater.* **6**, 299–303 (2004).
2. Cantor, B., Chang, I. T. H., Knight, P. & Vincent, A. J. B. Microstructural development in equiatomic multicomponent alloys. *Mater. Sci. Eng. A* **375-377**, 213–218 (2004).
3. George, E. P., Raabe, D. & Ritchie, R. O. High-entropy alloys. *Nat. Rev. Mater.* **4**, 515–534 (2019).
4. Praveen, S. & Kim, H. S. High-entropy alloys: potential candidates for high-temperature applications—an overview. *Adv. Eng. Mater.* **20**, 1700645 (2018).
5. Zhang, R. Z. & Reece, M. J. Review of high entropy ceramics: design, synthesis, structure and properties. *J. Mater. Chem. A* **7**, 22148–22162 (2019).
6. Oses, C., Toher, C. & Curtarolo, S. High-entropy ceramics. *Nat. Rev. Mater.* **5**, 295–309 (2020).
7. Wang, Z., Li, Z. T., Zhao, S. & Wu, Z. G. High-entropy carbide ceramics: a perspective review. *Tungsten* **3**, 131–142 (2021).
8. Rost, C. M. et al. Entropy-stabilized oxides. *Nat. Commun.* **6**, 8485 (2015).

9. Chen, T. K. & Wong, M. S. Structure and properties of reactively-sputtered Al_x-CoCrCuFeNi oxide films. *Thin Solid Films* **516**, 141–146 (2007).

10. Gild, J. et al. High-entropy metal diborides: a new class of high-entropy materials and a new type of ultrahigh temperature ceramics. *Sci. Rep.* **6**, 37946 (2016).

11. Zhang, Y. et al. Microstructure and mechanical properties of high-entropy borides derived from boro/carbothermal reduction. *J. Eur. Ceram. Soc.* **39**, 3920–3924 (2019).

12. Sarker, P. et al. High-entropy high-hardness metal carbides discovered by entropy descriptors. *Nat. Commun.* **9**, 4980 (2018).

13. Yan, X. et al. (Hf_{0.2}Zr_{0.2}Ta_{0.2}Nb_{0.2}Ti_{0.2})C high-entropy ceramics with low thermal conductivity. *J. Am. Ceram. Soc.* **101**, 4486–4491 (2018).

14. Zhou, J. et al. High-entropy carbide: a novel class of multicomponent ceramics. *Ceram. Int* **44**, 22014–22018 (2018).

15. Harrington, T. J. et al. Phase stability and mechanical properties of novel high entropy transition metal carbides. *Acta Mater.* **166**, 271–280 (2019).

16. Zhao, S. Lattice distortion in high-entropy carbide ceramics from first-principles calculations. *J. Am. Ceram. Soc.* **104**, 1874–1886 (2021).

17. Dippo, O. F. et al. Bulk high-entropy nitrides and carbonitrides. *Sci. Rep.* **10**, 21288 (2020).

18. Jin, T. et al. Mechanochemical-assisted synthesis of high-entropy metal nitride via a soft urea strategy. *Adv. Mater.* **30**, 1707512 (2018).

19. Lai, C. H., Lin, S. J., Yeh, J. W. & Chang, S. Y. Preparation and characterization of AlCrTaTiZr multi-element nitride coatings. *Surf. Coat. Technol.* **201**, 3275–3280 (2006).

20. Zhang, R. Z. et al. Data-driven design of ecofriendly thermoelectric high-entropy sulfides. *Inorg. Chem.* **57**, 13027–13033 (2018).

21. Gild, J. et al. A high-entropy silicide: (Mo_{0.2}Nb_{0.2}Ta_{0.2}Ti_{0.2}W_{0.2})Si_2. *J. Materiomics* **5**, 337–343 (2019).

22. Qin, Y. et al. A high entropy silicide by reactive spark plasma sintering. *J. Adv. Ceram.* **8**, 148–152 (2019).

23. Jiang, B. et al. High-entropy-stabilized chalcogenides with high thermoelectric performance. *Science* **371**, 830 (2021).

24. Li, Z. et al. Phase, microstructure and related mechanical properties of a series of (NbTaZr)C-based high entropy ceramics. *Ceram. Int* **47**, 14341–14347 (2021).

25. Liu, B. et al. Application of high-throughput first-principles calculations in ceramic innovation. *J. Mater. Sci. Technol.* **88**, 143–157 (2021).

26. Wang, Y. & Reece, M. J. Oxidation resistance of (Hf-Ta-Zr-Nb)C high entropy carbide powders compared with the component monocarbides and binary carbide powders. *Scr. Mater.* **193**, 86–90 (2021).

27. Wang, H., Han, X., Liu, W. & Wang, Y. Oxidation behavior of high-entropy carbide (Hf_{0.2}Ta_{0.2}Zr_{0.2}Ti_{0.2}Nb_{0.2})C at 1400–1600 °C. *Ceram. Int* **47**, 10848–10854 (2021).

28. Hume Rothery, W., Mabbott, G. W., Channel Evans, K. M. & Carpenter, H. C. H. The freezing points, melting points, and solid solubility limits of the alloys of silver and copper with the elements of the b sub-groups. *Philos. Trans. R. Soc. A* **233**, 1–97 (1934).

29. Troparevsky, M. C. et al. Criteria for predicting the formation of single-phase high-entropy alloys. *Phys. Rev. X* **5**, 011041 (2015).

30. Otto, F., Yang, Y., Bei, H. & George, E. P. Relative effects of enthalpy and entropy on the phase stability of equiatomic high-entropy alloys. *Acta Mater.* **61**, 2628–2638 (2013).

31. Castle, E. et al. Processing and properties of high-entropy ultra-high temperature carbides. *Sci. Rep.* **8**, 8609 (2018).

32. Liu, R. et al. Entropy as a gene-like performance indicator promoting thermoelectric materials. *Adv. Mater.* **29**, 1702712 (2017).

33. Zhang, Y. et al. Solid-solution phase formation rules for multi-component alloys. *Adv. Eng. Mater.* **10**, 534–538 (2008).

34. Troparevsky, M. C. et al. Beyond atomic sizes and Hume-Rothery rules: understanding and predicting high-entropy alloys. *JOM* **67**, 2350–2363 (2015).

35. Schmidt, J., Marques, M. R. G., Botti, S. & Marques, M. A. L. Recent advances and applications of machine learning in solid-state materials science. *npj Comput Mater.* **5**, 83 (2019).

36. Jablonka, K. M., Ongari, D., Moosavi, S. M. & Smit, B. Big-data science in porous materials: materials genomics and machine learning. *Chem. Rev.* **120**, 8066–8129 (2020).

37. Chen, C. et al. A critical review of machine learning of energy materials. *Adv. Energy Mater.* **10**, 1903242 (2020).

38. Liu, H. et al. Machine learning for glass science and engineering: a review. *J. Non-Cryst. Solids* **557**, 119419 (2021).

39. Hart, G. L. W., Mueller, T., Toher, C. & Curtarolo, S. Machine learning for alloys. *Nat. Rev. Mater.* **6**, 730–755 (2021).

40. Huang, W., Martin, P. & Zhuang, H. L. Machine-learning phase prediction of high-entropy alloys. *Acta Mater.* **169**, 225–236 (2019).

41. Islam, N., Huang, W. & Zhuang, H. L. Machine learning for phase selection in multi-principal element alloys. *Comput Mater. Sci.* **150**, 230–235 (2018).

42. Zhou, Z. et al. Machine learning guided appraisal and exploration of phase design for high entropy alloys. *npj Comput Mater.* **5**, 128 (2019).

43. Kaufmann, K. & Vecchio, K. S. Searching for high entropy alloys: a machine learning approach. *Acta Mater.* **198**, 178–222 (2020).

44. Zhang, L. et al. Machine learning reveals the importance of the formation enthalpy and atom-size difference in forming phases of high entropy alloys. *Mater. Des.* **193**, 108835 (2020).

45. Zhang, Y. et al. Phase prediction in high entropy alloys with a rational selection of materials descriptors and machine learning models. *Acta Mater.* **185**, 528–539 (2020).

46. Li, Y. & Guo, W. Machine-learning model for predicting phase formations of high-entropy alloys. *Phys. Rev. Mater.* **3**, 095005 (2019).

47. Kaufmann, K. et al. Discovery of high-entropy ceramics via machine learning. *npj Comput Mater.* **6**, 42 (2020).

48. Balasubramanian, K., Khare, S. V. & Gall, D. Valence electron concentration as an indicator for mechanical properties in rocksalt structure nitrides, carbides and carbonitrides. *Acta Mater.* **152**, 175–185 (2018).

49. Kindlund, H. et al. A review of the intrinsic ductility and toughness of hard transition-metal nitride alloy thin films. *Thin Solid Films* **688**, 137479 (2019).

50. Ma, J. et al. Low temperature synthesis of vanadium carbide (VC). *Mater. Lett.* **63**, 905–907 (2009).

51. Smith, J. F., Carlson, O. N., De & Avillez, R. R. The niobium-carbon system. *J. Nucl. Mater.* **148**, 1–16 (1987).

52. Marzban, C. The ROC curve and the area under it as performance measures. *Weather Forecast* **19**, 1106–1114 (2004).

53. López de la Torre, L. et al. Elastic properties of tantalum carbide (TaC). *Solid State Commun.* **134**, 245–250 (2005).

54. Clougherty, E. V., Lothrop, K. H. & Kafalas, J. A. A new phase formed by high-pressure treatment: face-centred cubic molybdenum monocarbide. *Nature* **191**, 1194–1194 (1961).

55. Kavitha, M., Sudha Priyanga, G., Rajeswarapalanichamy, R. & Iyakutti, K. Structural stability, electronic, mechanical and superconducting properties of CrC and MoC. *Mater. Chem. Phys.* **169**, 71–81 (2016).

56. Guo, S., Ng, C., Lu, J. & Liu, C. T. Effect of valence electron concentration on stability of fcc or bcc phase in high entropy alloys. *J. Appl Phys.* **109**, 103505 (2011).

57. Pradeep, K. G. et al. Non-equiatomic high entropy alloys: approach towards rapid alloy screening and property-oriented design. *Mater. Sci. Eng. A* **648**, 183–192 (2015).

58. Xu, Z. et al. Microstructural and mechanical behavior of a CoCrFeNiCu_4 non-equiatomic high entropy alloy. *J. Mater. Sci. Technol.* **60**, 35–43 (2021).

59. Yao, M. J., Pradeep, K. G., Tasan, C. C. & Raabe, D. A novel, single phase, non-equiatomic FeMnNiCoCr high-entropy alloy with exceptional phase stability and tensile ductility. *Scr. Mater.* **72-73**, 5–8 (2014).

60. Ma, D. et al. Phase stability of non-equiatomic CoCrFeMnNi high entropy alloys. *Acta Mater.* **98**, 288–296 (2015).

61. Ye, B. et al. First-principles study, fabrication and characterization of (Zr_{0.25}Nb_{0.25}Ti_{0.25}V_{0.25})C high-entropy ceramics. *Acta Mater.* **170**, 15–23 (2019).

62. Wang, Y. et al. Enhanced hardness in high-entropy carbides through atomic randomness. *Adv. Theory Simul.* **3**, 2000111 (2020).

63. Zhang, H. & Akhtar, F. Processing and characterization of refractory quaternary and quinary high-entropy carbide composite. *Entropy* **21**, 474 (2019).

64. Yang, X. & Zhang, Y. Prediction of high-entropy stabilized solid-solution in multi-component alloys. *Mater. Chem. Phys.* **132**, 233–238 (2012).

65. Liu, S. Y. et al. Phase stability, mechanical properties and melting points of high-entropy quaternary metal carbides from first-principles. *J. Eur. Ceram. Soc.* **41**, 6267–6274 (2021).

66. Sun, Y. et al. Synthesis of rare earth containing single-phase multicomponent metal carbides via liquid polymer precursor route. *J. Am. Ceram. Soc.* **103**, 6081–6087 (2020).

67. Kingma D. P., Ba J. Adam: a method for stochastic optimization. ICLR (2015).

68. Pedregosa, F. et al. Scikit-learn: machine learning in Python. *J. Mach. Learn Res* **12**, 2825–2830 (2011).

69. Kresse, G. & Hafner, J. Ab initio molecular dynamics for liquid metals. *Phys. Rev. B* **47**, 558–561 (1993).

70. Kresse, G. & Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B* **54**, 11169–11186 (1996).

71. Kresse, G. & Furthmüller, J. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Comput Mater. Sci.* **6**, 15–50 (1996).

72. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865–3868 (1996).

73. Blöchl, P. E. Projector augmented-wave method. *Phys. Rev. B* **50**, 17953–17979 (1994).

74. Zhao, S., Stocks, G. M. & Zhang, Y. Stacking fault energies of face-centered cubic concentrated solid solution alloys. *Acta Mater.* **134**, 334–345 (2017).

75. Zhang, J. et al. Elemental partitions and deformation mechanisms of L12-type multicomponent intermetallics. *Acta Mater.* **219**, 117238 (2021).

AUTHOR CONTRIBUTIONS

J.Z. performed the calculations and drafted the initial manuscript. Z.Wa. carried out the experiments. Z.Wu. supervised the experiments and reviewed the manuscript. S.Z. supervised the project, reviewed, and edited the manuscript. All authors contributed to the manuscript writing and discussion.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41524-021-00678-3.

Correspondence and requests for materials should be addressed to Zhenggang Wu or Shijun Zhao.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.