Available online at www.sciencedirect.com
ScienceDirect
Trans. Nonferrous Met. Soc. China 35(2025) 1855−1874
Transactions of Nonferrous Metals Society of China
www.tnmsc.cn

Phase classification of high entropy alloys with composition, common physical, elemental-property descriptors and periodic table representation

Shuai LI[1], Jia YANG[2], Shu LI[3], Dong-rong LIU[1], Ming-yu ZHANG[3]

1. School of Materials Science and Chemical Engineering, Harbin University of Science and Technology, Harbin 150080, China;
2. School of Computer Science and Technology, Harbin University of Science and Technology, Harbin 150080, China;
3. Key Laboratory of Engineering Dielectric and Applications (Ministry of Education), School of Electrical and Electronic Engineering, Harbin University of Science and Technology, Harbin 150080, China

Received 16 October 2023; accepted 3 June 2024

Abstract: Phase classification has a clear guiding significance for the design of high entropy alloys. For mutually exclusive and non-mutually exclusive classifications, the composition descriptors, commonly used physical parameter descriptors, elemental-property descriptors, and descriptors extracted from the periodic table representation (PTR) by the convolutional neural network were collected. Appropriate selection among features with rich information is helpful for phase classification. Based on random forest, the accuracy of the four-label classification and balanced accuracy of the five-label classification were improved to be 0.907 and 0.876, respectively. The roles of the four important features were summarized by interpretability analysis, and a new important feature was found. The model extrapolation ability and the influence of Mo were demonstrated by phase prediction in $(CoFeNiMn)_{1-x}Mo_x$. The phase information is helpful for the hardness prediction, the classification results were coupled with the PTR of hardness data, and the prediction error (the root mean square error) was reduced to 56.69.

Key words: high entropy alloy; phase classification; feature engineering; periodic table representation; convolutional neural network; hardness prediction

1 Introduction

High entropy alloys (HEAs) are a new type of alloys that generally contain five or more elements in mole fractions of between 5% and 35% [1,2]. Due to their unique compositions and structures, they have excellent properties, such as superior corrosion and wear resistance, high hardness, and high-temperature thermal stability [1−4]. The study of phase formation is meaningful for the design of HEAs. Due to the large and complex composition and structure space of HEAs, experiments and some theoretical calculation methods, such as first-principles calculation and calculation of phase diagram (CALPHAD), are difficult and time-consuming for large-scale exploration [2,5].

Data-driven science has become the fourth paradigm in materials research, after experiment, theory and computer simulation [6]. With the increase of materials data and the improvement of algorithms [7−10], machine learning (ML) could efficiently and automatically predict various material properties and inversely design desirable materials [11−14]. Thus, many researchers built ML models to predict the phase formation of HEAs [11,15−18].

Corresponding author: Shu LI, Tel: +86-18686840603, E-mail: lishu@hrbust.edu.cn;
Dong-rong LIU, Tel: +86-18104517785, E-mail: dongrongliu@hrbust.edu.cn
https://doi.org/10.1016/S1003-6326(25)66787-2

With an appropriate set of material descriptors, the ability of the ML model to learn the internal laws of the material data would be improved [19,20]. The common descriptors for HEAs phase classification are composition descriptors and empirical physical parameter descriptors [15−18]. The common empirical physical parameters include the mean valence electron concentration (VEC), the configuration entropy change (ΔS), and so on. Their prediction effect has been demonstrated in previous HEAs studies [16−18]. However, the HEAs phase classification still faces research problems. How to make the information contained in the features further enriched? How to select and verify the appropriate feature combinations? How to uncover different roles of the features? How to couple phase classification with the study of mechanical properties, such as hardness prediction? Exploring these problems and providing solutions based on ML are also the purpose of this work.

To enrich the information and improve the representation ability of the descriptors, it is worth exploring to add more phase classification related descriptors to the feature pool, and select an appropriate feature combination. Almost all common physical parameters for HEAs are proposed based on domain knowledge, and calculated based on the composition and physical properties of the constituent elements. Thus, more elemental-property parameters that are not suggested by domain knowledge could be considered. Furthermore, to avoid the limitations of human-defined computation methods of parameters, other new and promising descriptors could be extracted from the "periodic table representation (PTR)" of the material by the convolutional neural network (CNN) model [21]. The properties of the elements, such as relative atomic mass and atomic radius, are related to their positions in the periodic table. Therefore, in addition to the composition information, the PTR also implies the potential laws of the periodic table. The CNN model is then trained based on the PTR to extract the material descriptors or to predict the target property. In this work, the CNN model with the PTR input is referred to as CNN+PTR, and the descriptors extracted from the PTR by CNN model are named as CNN+PTR descriptors.

The CNN is an end-to-end model that includes an automatic feature extractor and a neural network model [22,23], and it is powerful for image processing tasks in materials science [24]. Based on the combination of the CNN model and PTR, some material properties have been successfully predicted [20,25−27], such as the critical casting diameter of metallic glasses [25]. In the HEAs studies, based on the CNN+PTR model, FENG et al [21] used the model trained on a big dataset of alloy glass-forming ability to extract descriptors of HEAs for HEAs phase prediction, and GUO et al [27] achieved three binary classifications of HEAs phases. In our previous work, a CNN+PTR regression model, which has a PTR with various information additions, is adopted to predict the hardness of HEAs [28]. Therefore, the descriptors derived from CNN+PTR are promising for automatically extracting useful information for the phase classification of HEAs. However, most CNN+PTR works have used the CNN+PTR models to predict material properties, and only a few have adopted models to extract features for further exploration. As the information contained in the descriptors should be enriched, there is great potential for combining these different types of descriptors for feature selection to achieve phase classification of HEAs.

To the authors' knowledge, most studies on HEAs phase classification by ML focused on mutually exclusive classification [15,17,18] and a few on non-mutually exclusive classification [16]. While some studies have adopted interpretable analysis methods, few studies have analyzed the specific roles of different feature ranges for different phases. There is also a need for further in-depth research on HEAs phase classification in practical applications, such as extrapolation for alloy systems outside the training set and deep coupling of phase classification and material property prediction. The main innovation of this work is the proposal and application of a new ML framework for mutually exclusive and non-mutually exclusive phase classifications of HEAs. Different from pervious works [15−18], a new feature pool containing multiple types of features was proposed for HEAs phase classifications, and appropriate feature combinations were selected. The new important features were analyzed, their roles in different phase formations were revealed, and the model extrapolation performance was explored. In the HEAs research, the present work can be used to

construct phase classification models and analyze the factors affecting phase formation. Furthermore, this work proposed and validated a method to couple phase classification with hardness prediction, which can be used to improve the HEAs hardness prediction and analyze the structure-property relationship.

In this work, four types of descriptors with different information were combined as the original features pool. Then, feature selection based on genetic algorithm (GA) was implemented for mutually exclusive four-label and non-mutually exclusive five-label HEAs phase classifications. The phase classification model with selected feature combination was constructed. The interpretability analyses were carried out to attain cognition of the mapping mode of the phase classification model and the rules of phase formation in HEAs. In order to understand the physical factors influencing phase formation and to guide the exploration of HEAs with the desired phase, it is important to analyze important features in this new and large feature pool and to determine different feature ranges that promote or inhibit the formation of different phases. The phase classification model with the selected descriptors was validated on additional and representative HEAs system of $(CoFeNiMn)_{1-x}Mo_x$, and the influence of the Mo content on phase formation was revealed. In order to deeply couple the structure information and property prediction of HEAs, the predicted phase classification results were added into the PTR to improve the HEAs hardness prediction of the CNN+PTR model and stacking model. This framework could be extended to other material properties research.

2 Materials and methods

2.1 Feature engineering

The feature engineering refers to a series of processes performed on the raw data to generate input features that are suitable for the ML models [29,30]. It generally includes feature construction, feature transformation, feature selection, and so on. For the phase classification of HEAs, various features should be considered in the feature construction, regardless of their importance in determining the HEAs phase, such as composition features, and the physical parameter features proposed based on domain knowledge and playing important role in the HEAs phases formation. Due to the limitations of the existing domain knowledge, a variety of physical parameter features based on different elemental properties can be generated by compositional weighted averaging, taking maximum and minimum values, and other operations. The specific information on these three types of features is presented in the next section. However, these features are manually designed, which are limited by the human-defined calculation operations and creates barriers to feature collection and calculation, leading to limited prediction performance. Therefore, other features are extracted automatically by CNN from the PTR. The specific process is described in Section 2.3.

In the feature transformation, as the method adopted in Refs. [29,30], the constructed features could be transformed by mathematical variations, such as X_1^2 and X_1X_2 for different features X_1 and X_2, respectively. This method could mathematically change how the features influence the phase prediction in ML algorithms and characterize the material from high dimensionality [29,30]. However, the features adopted in this work contain sufficiently rich information, which has been extracted and processed in various ways. In addition, the feature dimension is too large after performing the transformation operation. Therefore, similar feature transformation operations are not used.

In the feature selection, it can be mainly divided into the filter method, embedded method, and wrapper method. The filter method, which is based on Pearson's correlation coefficient, Chi-square test, and so on, only considers simple correlation indexes and may remove potentially important features. The embedded method relies on some specific ML models. For the general wrapper methods, such as sequential backward selection, the result easily falls into the local optimum. However, wrapper methods based on optimization algorithms can ensure accuracy and efficiency. A representative optimization algorithm is the GA, a method to search for the optimal solution by simulating the natural evolutionary process. It can avoid the feature selection process from falling into the local optimum to a certain extent and find suitable feature combinations.

For the hardness prediction of HEAs, another type of feature engineering method is adopted. Not

only the composition information is represented in the PTR, but the corresponding phase information predicted by the HEAs phase classification model and hardness-related physical parameters are added in the blank positions of the PTR in different ways. Then, the CNN model is adopted to be trained in the PTR and used to predict the hardness of HEAs. It achieves the deep coupling of phase classification and hardness prediction of HEAs.

2.2 Dataset briefing

892 HEAs phase data collected from the literature [27,28] are adopted as the phase dataset, and their processing conditions are all as-cast. The concentrations of the constituent elements are referred to as the composition descriptors. 23 physical parameters commonly used for HEAs properties prediction are named as HEAs common descriptors [11,16−18,31], which include VEC, ΔS and so on. In addition, 132 elemental-property parameters are extracted from the Magpie feature set, which is proposed by WARD et al [32], and they are named as elemental-property descriptors. They are defined as the composition-weighted mean, mean absolute deviation, range, minimum, maximum and mode of 22 different physical properties of constituent elements, and the elemental-properties include the specific volume and magnetic moment (per atom) of the element at the 0 K ground state, and so on. The latter two types of descriptors are known as physical parameter descriptors and the detailed information of them is provided in Table S1 of the Supplementary Information (SI) and related paragraph next to the Table S1 of the SI, and the main sources of access come from literature [14,32].

The phase could be categorized in different ways depending on the different demands. In this work, two different types of phase classification schemes are adopted. One is the mutually exclusive phase classification, which involves four common labels: solid solutions (SS), intermetallic compounds (IM), SS+IM mixed phases and amorphous phases (AM) [17,18]. In this scheme, each alloy belongs to only one of the phase categories. The other is the mutually non-exclusive phase classification [16,28], which involves five labels: face centered cubic (FCC), body centered cubic (BCC), hexagonally close-packed (HCP), IM and amorphous (AM). In this scheme, each sample could be described as one label for single-phase alloy or a combination of multiple labels for multi-phase alloys, such as FCC+BCC. Thus, the five-label classification scheme is able to predict various phases, even those not present in the training set. For this phase dataset, the number of occurrences of the constituent elements is counted and shown in Fig. S1 of the SI.

2.3 CNN model with periodic table representation

In the construction of the CNN+PTR classification model, the mapping process of the PTR is shown in Fig. 1(a), which takes AlCoCuFeNi as an example. The PTR image is a 9×18 matrix, and the columns and rows of the PTR correspond to the groups and periods of the periodic table, respectively. The atomic percentages of the constituent elements are filled in the corresponding positions and zeros are filled in the remaining positions. The construction process of the CNN model is shown in Fig. 1(b). It can be seen that the CNN model consists of two parts. One is the combination of several convolutional layers and pooling layers, the other is a shallow neural network model. The former is a feature extractor that automatically extracts features from the PTR images and the latter is a classifier which predicts HEAs phases based on the extracted features.

The CNN+PTR model can be used to directly predict the phases of HEAs or to extract features. The extracted material descriptors are the output of the flatten layer and the dimension is 192. The CNN model is compressed and simplified from the visual geometry group (VGG) to suit this work [21,28]. The structure and hyper-parameters of different CNN models are determined through multiple experiments, and the details are provided in the Fig. S2 of the SI that is drawn by the drawing function of the TensorFlow package in Python software. The training process of the CNN+PTR model has randomness, so the prediction performance is estimated as the average of three times of ten-fold cross-validation and the intermediate values of the different results.

2.4 Feature selection of genetic algorithm

Genetic algorithm is a global search algorithm for solving optimization problems [33], and it has proved to be a powerful tool for the feature selection [34,35]. In order to select an appropriate feature combination for the phase classification

model, GA is adopted and the feature selection process is shown in Fig. 2. In GA, a binary string with the length of the total number of features is generated to represent the feature combination, where each bit corresponds to a feature. The binary string is encoded by adding either the value 1 or the value 0 to each bit, where the value 1 indicates the presence of the corresponding feature and vice versa. Such a string is defined as an individual, and a collection of individuals is called a population.

The feature selection process of GA consists of five parts. Starting with the random generation of the initial population, the other four steps are repeated cyclically until the number of iterations reaches a preset value. Specifically, in the fitness calculation step, the fitness value for each individual is evaluated as the accuracy of the ML model with the corresponding combination of features, as evaluated by the ten-fold cross-validation. In the selection step, a new population is selected from the input population by using the roulette wheel selection method, where individuals

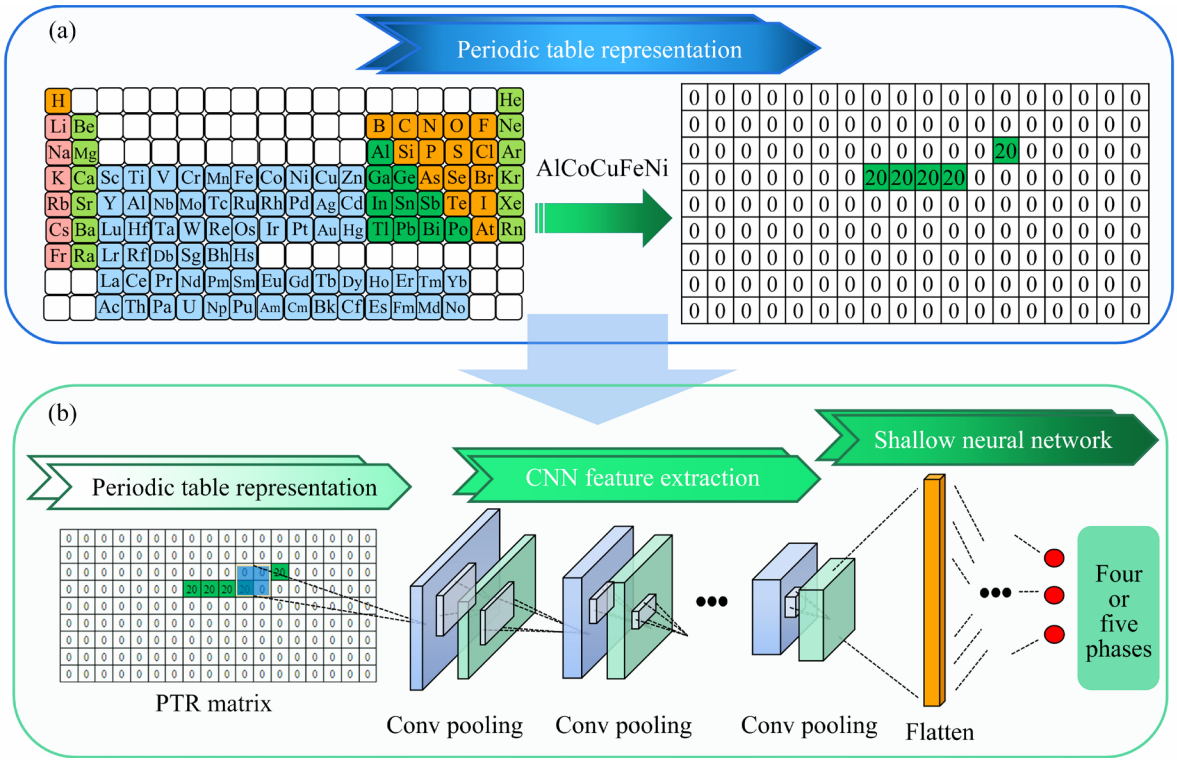

Fig. 1 Flow chart of CNN+PTR for HEAs phase classification: (a) Example of PTR; (b) PTR-based CNN phase classification model

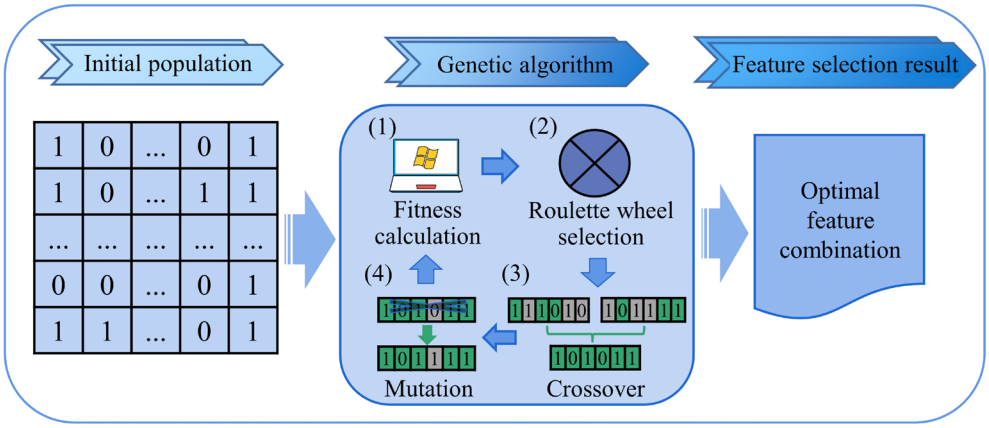

Fig. 2 Flow chart of feature selection based on GA consisting of initial population generation, genetic algorithm and feature selection result

with higher fitness values are more likely to be selected repeatedly. In the crossover step, a certain number of individuals are selected based on the crossover probability, and then a new individual is generated by randomly selecting two individuals from them and filling each position in a new individual with a value randomly selected from the corresponding values in these two individuals. A new population is formed by combining a certain number of newly generated individuals with the individuals not selected in the crossover step. In the mutation step, for each value in each individual, the value selected based on the mutation probability is replaced to the opposite value. The final feature selection result of the GA is the individual with the best accuracy that has occurred over the iterations of multiple GA runs. The hyper-parameters of the GA are determined through multiple experiments and experience, and they are listed in Table S2 of the SI.

2.5 Interpretability analysis method

The "black box" nature of the ML model hides specific mapping processing in the model and hinders the understanding of the material properties. Therefore, two interpretability analysis methods, feature importance of the Random Forest (RF) model and SHapley Additive exPlanations (SHAP) [36] are adopted in the phase classification model. Based on the SHAP results, the SHAP average absolute plot, the SHAP summary plot and the SHAP dependence plot could be shown. The importance value of each feature in the RF model represents the contribution of that feature in the model training, which is the sum of the reduction in impurity at each feature split during the training process. The impurity is estimated by the Gini index. SHAP is a post-hoc, model-independent interpretation method developed from game theory. It decomposes the prediction of each sample into the contribution of each feature [36]. The prediction of each sample could be considered as the sum of the SHAP values of all features:

$$\hat{y} = \hat{y}_0 + \sum_{i=1}^{n} \phi_i \tag{1}$$

where $\hat{y}$ is the model prediction result and $\hat{y}_0$ is a constant. The SHAP value for the ith feature (ϕ_i) is calculated as follows:

$$\phi_i = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|!(|F|-|S|-1)!}{|F|!} \cdot [f_{S\cup\{i\}}(X_{S\cup\{i\}}) - f_S(X_S)] \tag{2}$$

where F is the feature set involving all features, S is the feature subset of F, $S \cup \{i\}$ is the union of the feature subset S and the ith feature, $f_{S\cup\{i\}}(X_{S\cup\{i\}})$ is the prediction result of the ML model with the ith feature, and $f_S(X_S)$ is the prediction of the ML model trained without the ith feature. Equations (1) and (2) are applied to each sample to calculate the SHAP values of each feature. For a HEA phase, if the SHAP value of a feature is positive, the feature enhances the prediction of this phase; conversely, the feature with a negative SHAP value weakens the prediction of this phase. The higher the absolute SHAP value, the greater the importance of the feature.

3 Results and discussion

3.1 Process of HEAs phase classification and hardness prediction

The flow chart of this work is shown in Fig. 3. It can be seen that the whole work is divided into two parts: the primary part is the phase classification of HEAs, and the secondary part is the phase classification-assisted hardness prediction. In the HEAs phase classification, firstly, four types of descriptors, which contain a wide variety of rich information about the HEAs phase classification, are generated. Secondly, based on the selected ML classification algorithm and RF model, the features are selected by GA for the mutually exclusive and the non-mutually exclusive phase classifications. Thirdly, the RF models are constructed with the corresponding selected feature combinations for different phase classification schemes. Finally, for the mutually exclusive four-label phase classification, interpretability analyses are carried out to gain insight into the mapping mode from features to phase classification and the way that some important physical parameters influence the HEAs phase formation. Meanwhile, the extrapolation and generalization performance of the phase classification model is validated on the HEAs system of $(CoFeNiMn)_{1-x}Mo_x$. In the HEAs hardness prediction, in order to improve the prediction performance, the phase prediction probabilities obtained from the non-mutually exclusive five-label phase classification and some

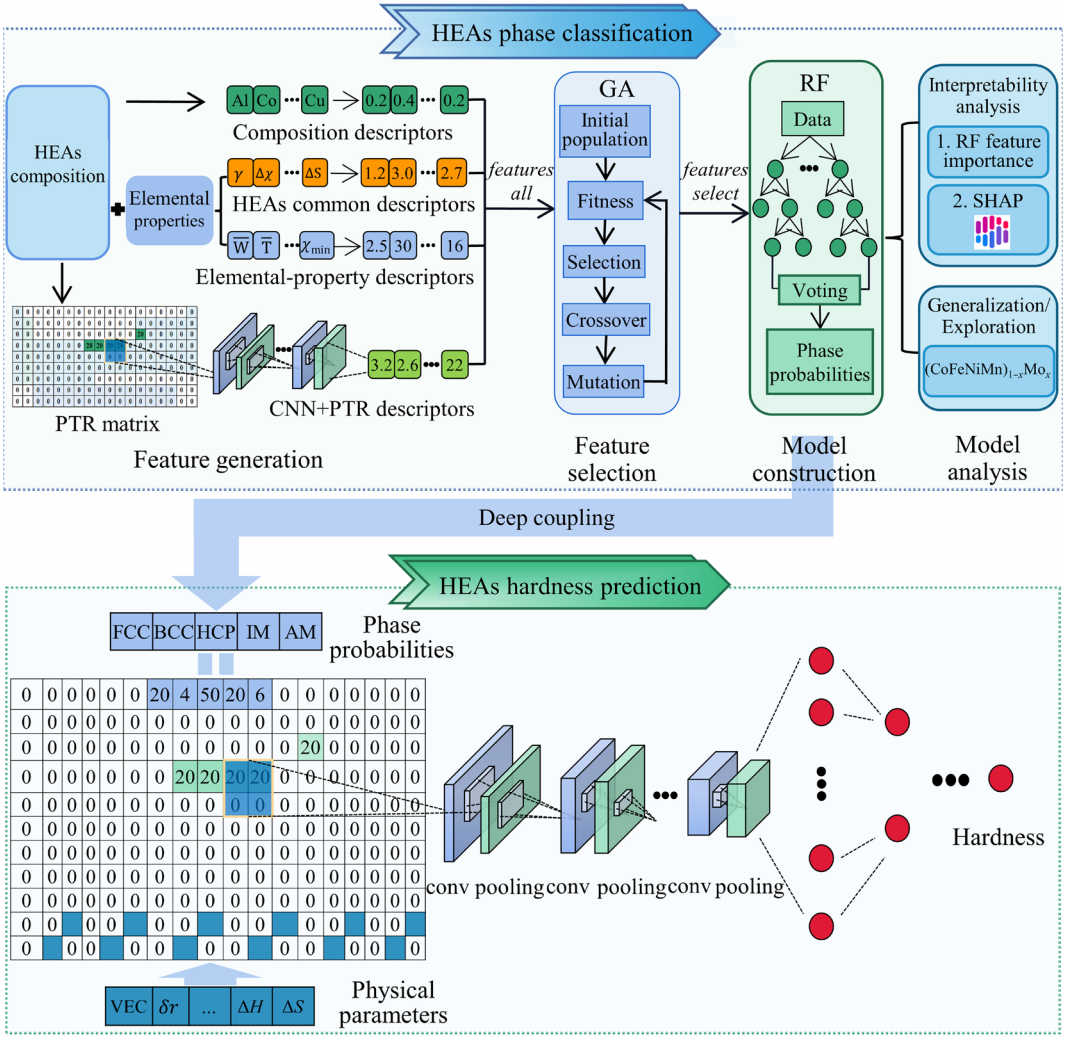

Fig. 3 Flow chart of this work consisting of phase classification and hardness prediction of HEAs

HEAs hardness-related physical parameters are added into the PTR of the HEAs hardness data.

3.2 Model construction and comparative analysis results

For the model selection, five ML models commonly used in HEAs phase classifications are considered. They are RF, K-nearest neighbor classification (KNN), logistic regression classification (LR), support vector machine classification (SVC) and decision tree classification (DTC) [14,17,18,34]. In order to ensure the reliability and generalization of model selection, four types of material descriptors, composition descriptors, 23 common HEAs physical parameter descriptors, 132 elemental-property parameter descriptors and 192 CNN+PTR extracted descriptors are adopted as input features, respectively. In addition, the CNN+PTR classification models are constructed as described in Section 2.3.

It must be emphasized that for the non-mutually exclusive five-label phase classification, a trained CNN+PTR model could predict the phase formation for HEAs, but other ML algorithms are only applicable to mutually exclusive classification, and they classify alloys into one of the categories present in the training set. Thus, it is necessary to construct five models for five binary classifications to determine whether five phases are present in HEAs respectively. The performance of non-mutually exclusive classification model is usually estimated by the balanced accuracy [16], which is the average proportion of correct predictions in each category. The calculation formula is shown as follows [16]:

$$\text{Balanced accuracy}=\frac{1}{k}\sum_{i=1}^{k}\frac{1}{m_i}\sum_{j=1}^{m_i}I\left(y_{ij}=\hat{y}_{ij}\right) \quad (3)$$

where k is the number of categories; m_i is the number of samples in category i; $I(\cdot)$ is indicator function, and the function value is 1 if the condition in the brackets is met, otherwise it is 0; y_{ij} and $\hat{y}_{ij}$ are the true label and the prediction result for the jth sample in the category i, respectively; the forms of the true label and the prediction result are set according to whether the sample belongs to category i or not. The prediction performance of these ML models is shown in Fig. 4. In this work, the prediction performance of the ML models is estimated by the average value of the ten-fold cross-validation.

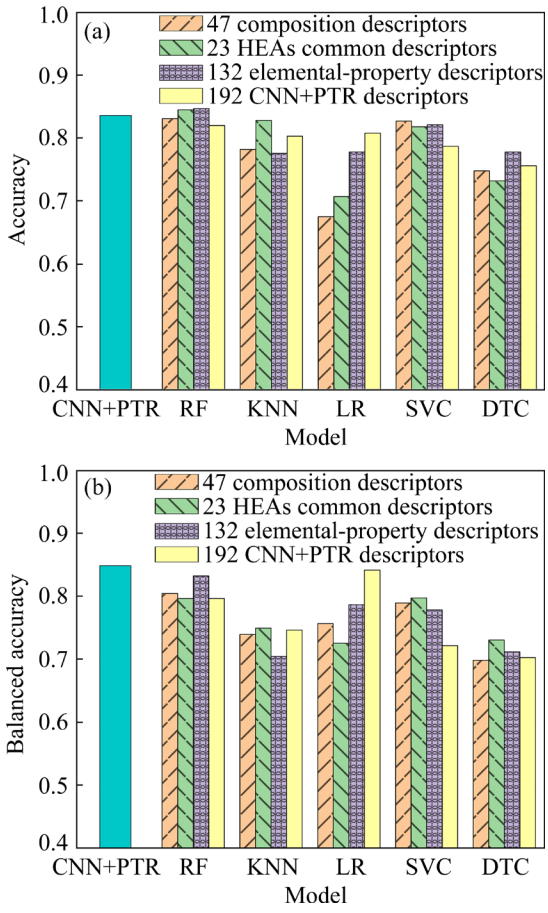

Fig. 4 Prediction performance of different models with different descriptors: (a) Results for four-label phase classification; (b) Results for five-label phase classification

It could be seen that for these two classification schemes, with the inputs of any of the four types of features, the RF model outperforms other common models, only except for the LR model with 192-dimensional CNN+PTR features in five-label phase classification. This demonstrates that the RF model has good prediction ability for any kind of features and phase classification task. Based on such results and our experience, the RF model could also show satisfactory performance in the following phase classification tasks. RF is an ensemble learning model which could reduce the high "bias" and high "variance" for a single ML model. RF introduces randomness into the model training, so it is less prone to overfitting, and may also have better prediction ability for data outside the existing dataset. In addition, compared to other models, the RF model also has the advantages of being noise-resistant, possessing a high training speed, and having simple hyper-parameters that generally do not require tuning. Therefore, in this work, based on these reasons, RF is chosen as the phase classification model and performed for further feature selection.

It is difficult to rank the role of different types of descriptors because they have different effects in different cases. The CNN+PTR model, which uses only composition information and the periodic table of elements as input, outperforms most of the other models, especially those based on composition descriptors. This is because the CNN+PTR model extracts phase-related physical information from the positions of the constituent elements in the periodic table. Therefore, the prediction performance of the models based on the CNN+PTR extracted descriptors is also good, especially when the LR model is used for the five-label phase classification. The elemental-property descriptors perform best in the RF model for these two classification schemes. Therefore, the addition of the elemental-property descriptors and CNN+PTR descriptors to the feature pool is expected for the development of high-quality HEAs phase classifiers.

3.3 Feature selection and comparative analysis results

In order to combine all the useful information, feature pool containing four types of descriptors is constructed. This feature pool could represent composition information, physical information from the physical properties of the constituent elements, and periodic table information. The combination of all of these features is referred to as features-all. Then, in order to select the optimal feature combination from features-all for two phase classifications, GA with RF model is utilized for feature selection. In order to show the change trends

of the feature number and the prediction performance during the optimization process of the GA, the changes in the feature number and the accuracy of the optimal individual found from the beginning to different iteration numbers are shown in Fig. S3 of the SI. It can be seen that in the different phase classification tasks, as the iteration process proceeds, the prediction performance gradually rises and the feature number in the corresponding feature combination generally decreases. This suggests that GA-based feature selection ensures both prediction performance and significant feature number reduction, which implies improving model generalization and avoiding over-fitting for RF model.

The selected feature combination for four-label phase classification is referred to as features-select-four-label, and that for five mutually exclusive binary phase classifications in non-mutually exclusive five-label phase classification is named as features-select-FCC, features-select-BCC, and so on, respectively. For convenience, any selected feature combination can be abbreviated as features-select if it does not refer specifically to a phase classification. The content of each features-select is listed in Tables S3 and S4 of the SI. The number of different types of descriptors in features-select and the prediction performance of RF models with features-select for four-label and five-label classifications are shown in Table 1. It should be noted that, because the balanced accuracy is estimated by the average proportion of correct predictions in each category, the value obtained for the five-label is lower than the accuracy for any of the five labels when they are predicted individually.

For the RF model with features-select-four-label, the accuracy value of the mutually exclusive four-label phase classification is 0.907. Based on the five RF models with features-select-BCC, features-select-FCC, and so on, the balance accuracy of the mutually non-exclusive five-label phase classification is 0.876, and the accuracy values of the five binary phase classifications are all above 0.92, and even above 0.98 for both HCP and AM phases. Each features-select contains more than one type of descriptor, and both the elemental-property descriptors and the CNN+PTR descriptors are included. Thus, it is necessary to combine multiple descriptors, especially elemental-property and CNN+PTR descriptors for feature selection.

In order to verify the accuracy of model selection, we selected the SVC model, which also generally shows good phase classification performance in Fig. 4, for the above feature selection process. The results are shown in Table S5 of the SI. The random grid search method is used for tuning hyper-parameters of SVC model. It can be seen that the prediction ability of the RF model is generally similar to that of the SVC model in phase classification tasks. However, the feature numbers in the selected features for the SVC model are all obviously larger than those for the RF model, so the feature dimensionality reduction effect of the feature selection with the RF model is more obvious than that with the SVC model. In summary, it is demonstrated again that the RF model should be selected to perform feature selection and HEAs phase classification.

In order to further explore the effects of the combination of different types of features and feature selection, the prediction performance of the RF models with different features is compared and analyzed. The prediction performance of the RF model with elemental-property descriptors, features-all and features-select is shown in Fig. 5. It

Table 1 Prediction performance for features-select and number of different type descriptors included in features-select

Classification task	Number of composition descriptor	Number of HEAs common descriptor	Number of elemental-property descriptor	Number of CNN+PTR descriptor	Accuracy
Four-label (SS, AM, IM, SS+IM)	1	4	13	10	0.907
Five-label (FCC)	3	1	7	12	0.969
Five-label (BCC)	2	0	13	5	0.954
Five-label (HCP)	0	0	7	6	0.982
Five-label (IM)	5	7	14	8	0.921
Five-label (AM)	3	1	6	12	0.996

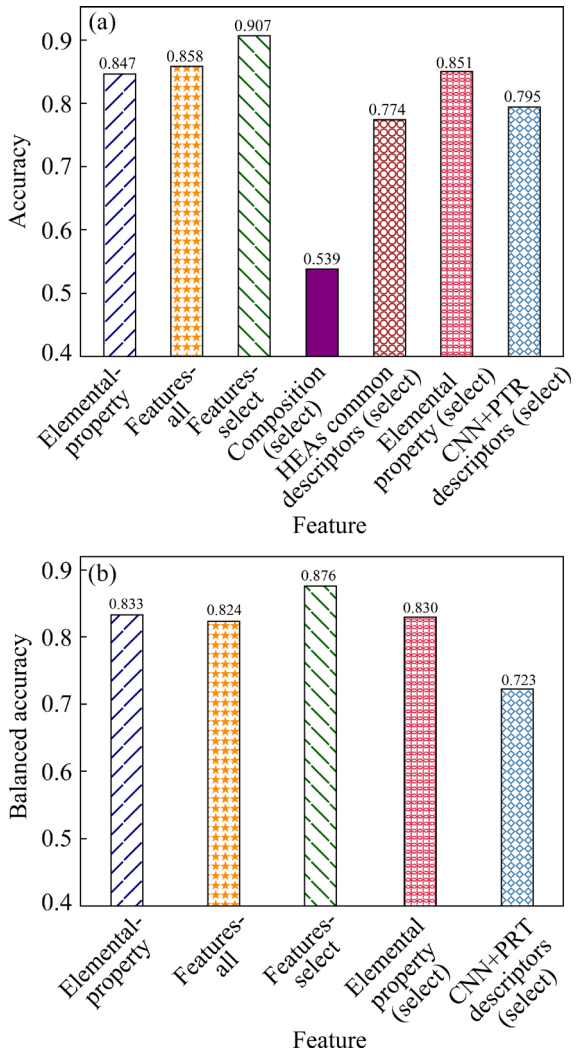

Fig. 5 Prediction performance of RF models with different features: (a) Results for four-label phase classification; (b) Results for five-label phase classification

could be seen that the prediction performance of the RF model with features-all is almost equivalent to that with elemental-property descriptors, and this is due to the high dimensionality of features-all and the information redundancy in it. The prediction performance of the RF model is significantly improved by the feature selection of the GA, and the effectiveness of the feature selection is demonstrated. For the sake of showing the role of different types of descriptors in features-select, the prediction performance of each type of descriptor in features-select is shown in Fig. 5. The prediction performance of features-select is significantly better than that of others. The conclusions of the model comparison in two classification schemes are consistent. In summary, for different HEAs phase classification schemes, the ML workflow that combines four types of features and appropriate feature selection of GA could show excellent prediction performance.

3.4 Interpretability analysis results

The interpretability analysis is used for the four-label phase classification because the study of these four phases, especially the formation of the SS, is meaningful for the research of HEAs and receives wide attention [2,3]. In order to analyze the specific ways in which material descriptors influence the formation of different HEAs phases, two interpretability methods, feature importance of the RF model and SHAP are adopted. Furthermore, four types of figures, the RF model feature importance plot, the SHAP average absolute plot, the SHAP summary plot and the SHAP dependence plot are shown.

3.4.1 Feature importance

Since the prediction performance of the RF model is evaluated by the ten-fold cross validation, the feature importance of the RF model and the SHAP values of different features for four phases are calculated as the average values of the ten models results for the dataset. The normalized feature importance of the RF model and the average absolute SHAP values are shown in Figs. 6(a) and (b), respectively. The greater the absolute value of the SHAP value, the greater the influence of the feature on the target. It could be seen that the ranking of the features is generally similar in these two figures. In particular, the top four features are exactly the same. This demonstrates the reliability of the feature importance. The distributions of the SHAP values of these features for four phases are shown in the SHAP summary plot in Fig. S4 of the SI. The CNN+PTR extracted descriptors have no clear physical meaning, so there is no further interpretative analysis for them.

From Fig. 6, it could be seen that the four most important features are radii gamma (γ), radii local mismatch (D_r), electronegativity local mismatch (D_χ) and mean GSvolume_pa ($\overline{V}_{K=0}$). The first three belong to the HEAs common physical parameter descriptors and the last one belongs to elemental-property descriptors. The corresponding formulae are as follows:

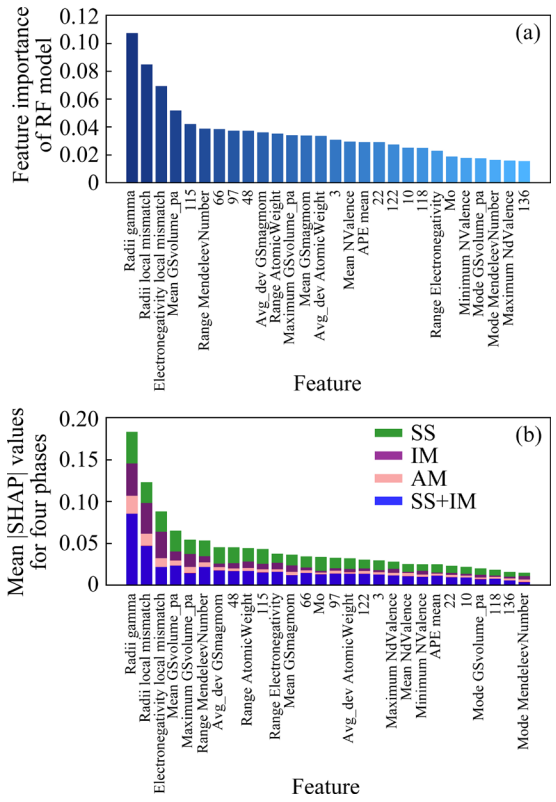

Fig. 6 Role of features in phase classification: (a) Feature importance of features-select in RF model; (b) Mean absolute SHAP values of features-select for four phases

$$\gamma=\left(1-\sqrt{\frac{\left(\overline{r}+r_{\min}\right)^2-\left(\overline{r}\right)^2}{\left(\overline{r}+r_{\min}\right)^2}}\right)\Bigg/\left(1-\sqrt{\frac{\left(\overline{r}+r_{\max}\right)^2-\left(\overline{r}\right)^2}{\left(\overline{r}+r_{\max}\right)^2}}\right) \tag{4}$$

$$D_r=\sum_{i=1}^{n}\sum_{j=1,i\neq j}^{n}C_iC_j\left|r_i-r_j\right| \tag{5}$$

$$D_\chi=\sum_{i=1}^{n}\sum_{j=1,i\neq j}^{n}C_iC_j\left|\chi_i-\chi_j\right| \tag{6}$$

$$\overline{V}_{K=0}=\frac{1}{n}\sum_{i=1}^{n}C_i\left(V_{K=0}\right)_i \tag{7}$$

where $\overline{r}$ is the mean value of the atomic radius, $\overline{r}=\frac{1}{n}\sum_{i=1}^{n}C_ir_i$; $r_{\min}$ and $r_{\max}$ are the minimum atomic radius and the maximum atomic radius of all elements, respectively; C_i, r_i, χ_i and $(V_{K=0})_i$ are the mole ratio, atomic radius, Pauling electronegativity and the specific volume at the 0 K ground state of the ith element, respectively. The parameters γ, D_r and D_χ were used in previous works on predicting the properties of HEAs [28,33]. The parameter $\overline{V}_{K=0}$ is a physical parameter first found in this paper to have a great influence on the phase classification of HEAs.

3.4.2 Analysis results of important features

The plots of the two most important features for different phases are shown in Fig. 7 and the plots of the third and fourth important features for different phases are shown in Fig. S5 of the SI. They show the SHAP values of different feature values for different phases, where (a) and (b) show the SHAP values of different features, and (a_1−a_4) and (b_1−b_4) represent the SHAP values of a given feature for four phases. In each figure, points with dark color indicate that the SHAP values are greater than 0, and the feature values have positive effect on the target phase prediction; points with light color are the opposite. It is found that for these features there are some boundaries between positive and negative SHAP values. The summary and analysis of them, especially for the SS phase, are meaningful for exploring the HEAs with the desired phases and for understanding the classification process of the model.

As shown in Fig. 7(a), the high values of the parameter γ inhibit the formation of the SS phase and favor the formation of the other phases, and vice versa. The transition threshold is 1.08 for the SS and SS+IM phases and 1.10 for the IM and AM phases, and these two values are close. The conclusions are similar to those obtained by parameter partitioning in the study by WANG et al [37] and the Hume–Rothery rules. As shown in Fig. 7(b), the trends of SHAP values for different phases with feature values of D_r [38] are like those of γ. For the SS, IM and SS+IM phases, the SHAP values of D_r change to the opposite sign in a transition range from 1.8 to 4.7. The transition range is firstly proposed in this work to describe a feature value range with both positive and negative SHAP values, where the SHAP values on either side of the range have opposite signs. The transition threshold is 4.7 for the AM phase. These two parameters represent the atomic size difference in HEAs from different perspectives. Large differences in atomic size would result in lattice distortions and be detrimental to the stability of solid solutions. The γ is the ratio between the solid angles of the smallest and largest atoms [37,39]. Since the atoms with the largest and smallest sizes

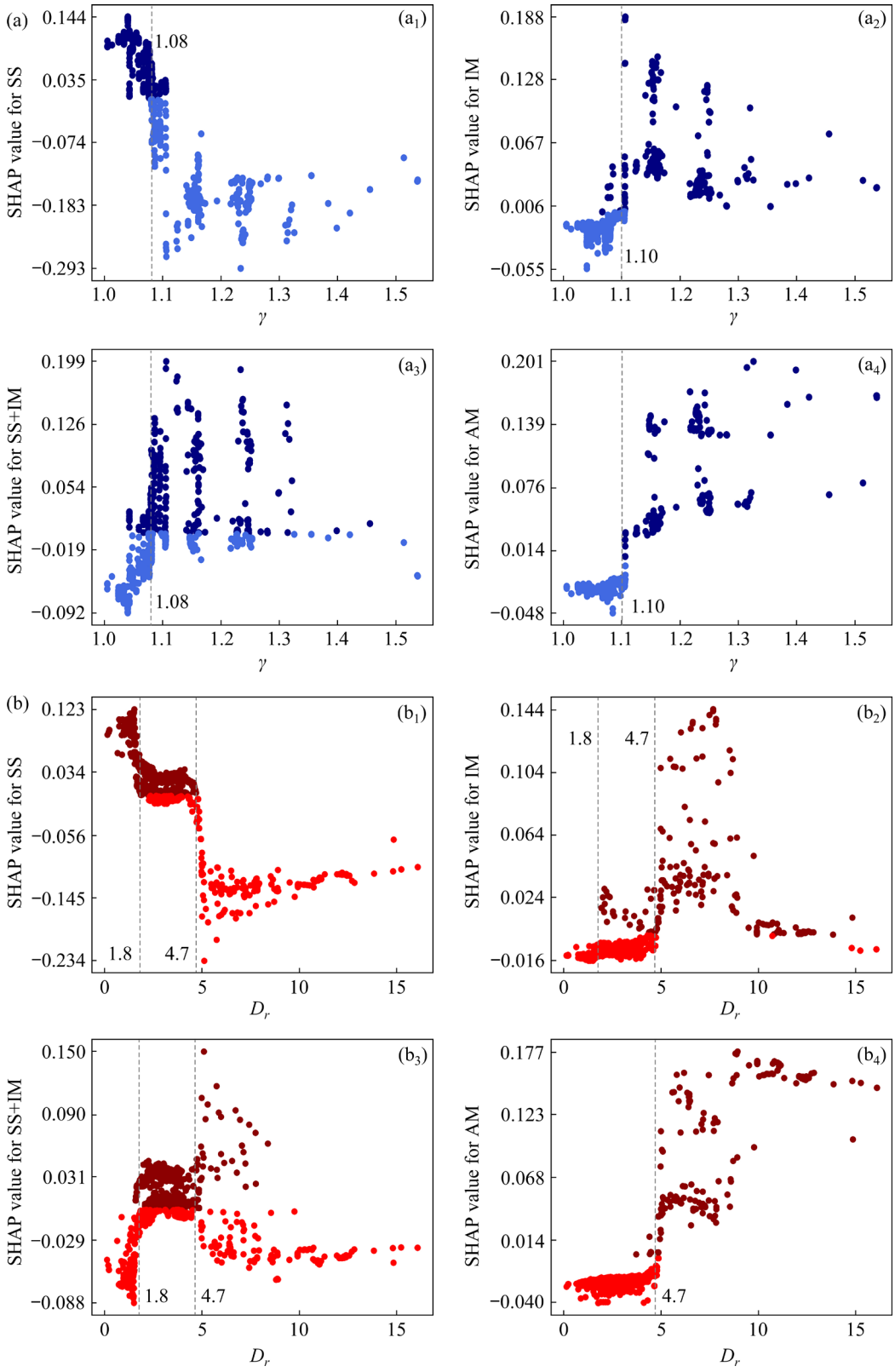

Fig. 7 Plots of two most important features of γ (a) and D_r (b) for different phase predictions

play a dominant role in determining lattice stability, as an indicator of revealing atomic stacking mismatches and topological instability [37,39], the parameter γ plays a more important role than the parameter D_r.

As shown in Fig. S5(a) of the SI, the high values of the parameter D_χ inhibit the formation of SS and SS+IM phases and favor the formation of

IM and AM phases, and vice versa. The transition threshold is 0.27. D_χ represents the difference in Pauling electronegativity within the HEAs, and the electronegativity is a chemical property that describes the tendency of an atom to attract electrons towards itself [40]. Thus, the greater the difference in electronegativity values between the constituent elements, the easier it is to form the IM phase, especially the Topological Closed Packed phase, in HEAs [40]. Meanwhile, the bonds formed between atoms with large differences in electronegativity values tend to be stronger, which may limit the maximum solid solubility.

As shown in Fig. S5(b), the relationship between the value range of parameter $\overline{V}_{K=0}$ and the HEAs phase prediction is relatively complex, and the detailed results are shown in Table S6 of the SI. In general, if a high-purity SS phase is required, it is recommended that the $\overline{V}_{K=0}$ is less than 11.3, and for the AM phase, $\overline{V}_{K=0}$ needs to be greater than 21.0. The parameter $\overline{V}_{K=0}$ is the composition-weighted mean of the elemental specific volume at the 0 K ground state. Specific volume is the ratio of the volume of a substance to its mass, and its value is affected by temperature [41]. When the temperature is 0 K and the element reaches the ground state, the specific volume represents the absolute minimum volume and reflects the arrangement and interaction of the atoms. The value of this elemental property is taken from the Open Quantum Materials Database (OQMD) [42].

The specific volume affects the structural stability of alloys and has been used to determine the glass phase formation [41,43]. Some structures are stable at small specific volumes, while others are the opposite. For example, the SS phase is more compact in structure than the other phases, and a small specific volume is beneficial to maintaining a compact structure and reducing the lattice distortion and strain energy, thus improving the stability of the SS phase. On the contrary, a higher specific volume would favor the formation of the glass phase of the AM phase [44]. LOUZGUINE-LUZGIN et al [44] have found that the structure of $Zr_{65}Ni_{10}Cu_5Al_{7.5}$-$Pd_{12.5}$ changes from completely glass phase to the SS-containing microstructure under different annealing conditions, with the specific volume generally decreasing. Therefore, with different specific volumes, the phase formation would also be different to ensure the stability of the structure.

In addition, based on the Cohen−Grest free-volume theory [45], it is demonstrated that the viscosity at the melting temperature is related to the volume change during crystallization [46]. The high-temperature viscosity and the free volume of a liquid have a significant effect on the critical cooling rate of vitrification, and the correlation between them is crucial for understanding the crystallization kinetics of bulk metallic glasses [46]. The specific volume of the alloy is strongly related to the parameter $\overline{V}_{K=0}$. Therefore, this parameter can provide unique and useful information for the researches of phase formation and other properties in HEAs.

3.5 Model validation and analysis results in $(CoFeNiMn)_{1-x}Mo_x$ HEAs

As an example of evaluating the model performance in practical applications, i.e., the generalization and exploration abilities of the phase classification model, some representative experimental data are additionally provided for model prediction. The content of element Mo is the only composition descriptor among the features-select for the four-label phase classification. Moreover, Mo is often considered an important element in the design of HEAs [47−49]. Therefore, the HEAs system $(CoFeNiMn)_{1-x}Mo_x$, which is not present in the phase dataset, is selected. These HEAs were prepared by vacuum arc-melting [49]. Based on X-ray diffraction measurements, it was found that the HEAs evolved from a single FCC matrix to an FCC + Laves phase, i.e., from SS to SS+IM phase, after the addition of Mo to CoFeNiMn [49]. The proportion of Laves phase gradually increases with increasing Mo content [50−52]. The detailed information is listed in Table S7 of the SI, and the source of access comes from Ref. [50].

The RF model with the features-select in four-label classification is used to predict the collected data, and the prediction results are shown in Fig. 8. It could be seen that as the Mo content increases, the prediction probability of SS decreases, that of SS+IM phases increases, and those of IM and AM phases are all close to 0. The relationship between SHAP values and Mo content for the four phases is shown in Fig. 9. The trends of the microstructure change are in good agreement with the experiment results and the interpretability

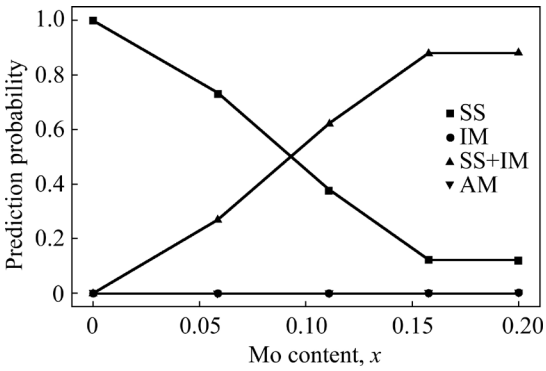

Fig. 8 Prediction probabilities of four phases for $(CoFeNiMn)_{1-x}Mo_x$

analysis results in Fig. 9. The phase at Mo content of 0.0588 is predicted to be SS, rather than the SS+IM obtained in experiment because the content of Laves phase is too low and is distributed in granular form on the FCC solid solution matrix [50].

Due to the large atomic radius, high modulus, and high electronegativity of Mo [53,54], the addition of Mo would result in lattice distortions and form intermetallic compound. Specifically, some of the Mo is incorporated into the FCC solid solution structure and induces lattice distortion, and the remainder forms the Fe_2Mo Laves phase with the Fe element [50]. As shown in Fig. 9, the transition threshold of Mo content is 0.06 for the SS and SS+IM phases and 0.03 for the IM and AM phases. Both thresholds are low for the lowest constituents of HEAs. Therefore, the presence of Mo is likely to induce the formation of the SS+IM phase in HEAs of $(CoFeNiMn)_{1-x}Mo_x$.

3.6 Deep coupling of phase classification with hardness prediction

An important reason for studying the phase classification of HEAs is that the phase has a great influence on the properties of HEAs. However, the relationship between them is often described qualitatively and is difficult to be used to design HEAs with specific properties. For HEAs hardness prediction, the collected HEAs hardness dataset contains 483 data, and their processing conditions are all as-cast [28]. In addition, 12 empirical physical parameters related to the hardness of HEAs are calculated and presented in Table S8 of the SI, and the sources of access come from author's

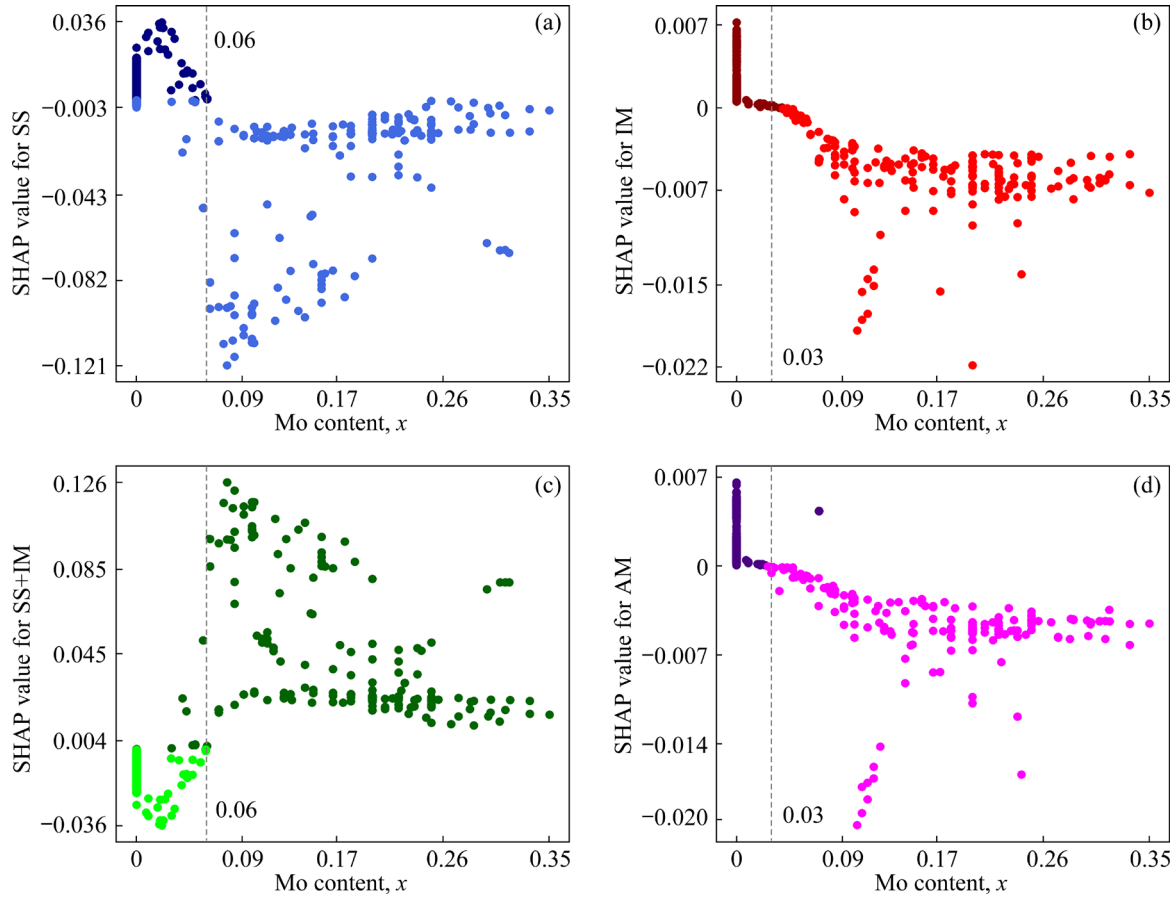

Fig. 9 Relationship between SHAP value and Mo content for four phases

experience and previous work [28]. In the process of constructing the CNN+PTR hardness prediction model, two additional rows are added below the PTR, and then the physical parameters are standardized by the *z*-score method and displayed in Rows 10 and 11 with a dispersed arrangement. Furthermore, the phase probabilities predicted by the existing four-label and five-label phase classification models are added to the PTR in the middle four or five positions of the first row. The schematic of the PTR mapping is shown in Fig. S6 of the SI. In this model, the predicted phase classification results, i.e., the microstructure information, are used to predict HEAs properties and to establish a quantitative relationship between them.

Several CNN+PTR models with different information are constructed, the hyper-parameters and other details of the model structure and training process are introduced in the Fig. S2 of the SI, and they are determined through multiple experiments. An ensemble learning strategy, the stacking method, is adopted to obtain more accurate and stable prediction results. Specifically, three CNN models are trained on the same training set and the phase prediction probabilities of the training set and the testing set are used as features for the new training set and the new testing set. Then, a support vector regression (SVR) model is trained on the new training set. The errors of the hardness prediction models are evaluated by the mean of the root mean square error (RMSE) in three times ten-fold cross-validation. The results are shown in Table 2, where "true" indicates that the corresponding information is added to the PTR, and "None" indicates the opposite.

Table 2 Prediction errors (RMSE) of different CNN+PTR and stacking models

Phase information	Physical parameter	CNN+PTR error	Stacking error
None	None	67.75	64.27
Four-label phases	None	67.84	63.16
Five-label phases	None	64.11	61.18
None	True	61.73	58.73
Four-label phases	True	61.45	58.13
Five-label phases	True	60.02	56.69

It could be seen from Table 2 that the stacking method significantly improves the accuracy of the hardness prediction. In addition, the stacking method constructs the final model based on the randomness of CNN+PTR to improve the stability of hardness prediction. The reason is that ensemble learning could reduce the high "bias" and high "variance" and mitigate the statistical, computational and representation problem for a single ML model. Therefore, the stacking model is adopted as the final hardness prediction model. In order to determine whether the model is statistically improved by the addition of different information or the implementation of the stacking method, the Diebold−Mariano (DM) test [55] is used to compare the prediction performance of different models from a statistical point of view. The details and results are given in Table S9 of the SI and related paragraph next to Table S9. The results can be summarized and analyzed as follows.

In the cases where no physical parameters are added, the addition of the five-label phase classification probabilities obviously improves the prediction accuracy of the stacking model, but the four-label phase classification results do not. This is because, unlike the mutually exclusive four-label phase classification, each phase in the non-mutually exclusive five-label phase classification can be predicted individually, and various combinations of these phases can be represented. Furthermore, these five phases are particularly related to the hardness of HEAs. For example, the FCC phase generally results in relatively low hardness, and the BCC and IM (Laves, σ and η) phases generally result in relatively high hardness, and the mixed phase alloys have uncertain properties [41].

In the cases where physical parameters are added to the PTR, the prediction performance of the stacking models is all significantly improved. These 12 physical parameters represent different physical properties based on the domain knowledge. Therefore, the addition of 12 physical parameters helps predict the HEAs hardness. However, the effect of adding the five-label phase information is relatively small in these cases. This may be because most of the physical parameters are related to the formation of HEAs phases and could represent the phase information of HEAs [43], such as the parameter γ mentioned above and the valence electron concentration (VEC). For HEAs, the BCC phase is stable at a low VEC value and the FCC phase is stable at a high VEC value [44−46].

3.7 Challenges and outlook

Data availability is a major challenge for the phase classification and hardness prediction of HEAs. Despite the increasing research on HEAs, there is still limited relevant data reported, especially for some properties such as hardness. This limits the exploration and generalization capabilities of ML models. In future research, as much relevant data as possible should be collected, or data augmentation could be used.

For the phase classification, various types of features were considered in this work, especially CNN+PTR features, which enriched the information of the features and led to better accuracy of the models. These phase classification models can be used to determine the microstructure of unknown HEAs, and different ranges of physical parameters summarized above are instructive for designing HEAs with the desired microstructure. However, the phase classification models can be further enhanced and the derived phase formation laws are limited by the classification method and classification accuracy of the models. Thus, in future research, the variety of features could be expanded; more detailed phase classification schemes could be further explored; the interpretability analysis of the model could be further deepened and optimized to explore more useful factors affecting the phase formation of HEAs.

In the hardness prediction, for the best model, the RMSE is 56.69, the R^2 (coefficient of determination) is 0.92 and the MAPE (mean absolute percentage error) is 8.43%. It could be used to predict the hardness of unknown HEAs and to design HEAs with high hardness. However, the accuracy of the model needs to be further improved, mainly due to insufficient data. In other perspectives, the hardness prediction accuracy is significantly improved by adding the phase classification probability and physical parameters to the PTR. Therefore, in future research, the coupling between phase classification and hardness prediction can be further deepened. More hardness relevant information, such as more physical parameters, can also be incorporated into the hardness prediction model. The way used in this work could also be extended to other material research tasks.

In the integrated computing materials and engineering (ICME), machine learning, physical modeling methods, such as phase field (PF), first principles, molecular dynamic simulations, and corresponding experiments could be coupled together to accelerate materials design and manufacturing [56]. The “integrated phase field method” (IPFM) has great potential and is becoming one of the most important and promising core components of ICME [56]. With the advancement of PF simulation and computer computing power, a promising application area is the combination of PF simulation, image data construction and ML to optimize the microstructure and design materials with desired properties [57]. As summarized in Ref. [58], the research directions of ML-PF worthy exploring in the future include the prediction of material properties from their microstructure [59,60], reverse design for material with specific property by optimizing alloy microstructure or processing condition [61,62], and acceleration of PF simulation [63]. In this work, based on the collected experiment data, the phase is predicted by the ML model, and the phase prediction probabilities are adopted to improve the hardness prediction. Guided by the above introduction, more data from theory calculation could be adopted. Based on the microstructure evolution and alloy properties explored by PF and so on [57,58,60], the deeper structural information could be coupled into the mechanical performance prediction, and the design of HEAs with high properties could be achieved accurately and efficiently.

4 Conclusions

(1) Based on the ML flow, high-performance ML models were constructed for the mutually exclusive four-label and the non-mutually exclusive five-label HEAs phase classifications.

(2) The interpretability analyses in the four-label phase classification demonstrated important descriptors of γ, D_r, D_χ and $\overline{V}_{K=0}$, and further indicated the influence of different descriptor value ranges on different phase forms. The $\overline{V}_{K=0}$ which has never been raised and emphasized for HEAs phase classification was found, and the concept of transition range was proposed. The results could guide the exploration of HEAs with desired phase.

(3) The generalization and exploration abilities of the model were successfully validated in $(CoFeNiMn)_{1-x}Mo_x$ HEAs and the influence of Mo content on the phase formation was revealed.

(4) In order to improve the hardness prediction, the phase prediction results and the physical parameters were added to the PTR, and the stacking method was employed. The accuracy of HEAs hardness prediction by CNN+PTR model could be significantly improved by ensemble learning and information addition of structure and physical parameters which are mostly related to the phase formation.

CRediT authorship contribution statement

Shuai LI: Conceptualization, Investigation, Software, Visualization, Writing − Original draft; **Jia YANG:** Methodology, Investigation; Software; **Shu LI:** Supervision, Project administration, Resources, Funding acquisition, Investigation, Writing − Review & editing; **Dong-rong LIU:** Supervision, Project administration; **Ming-yu ZHANG:** Data curation, Visualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The work was supported by the National Natural Science Foundation of China (Nos. 51671075, 51971086), and the Natural Science Foundation of Heilongjiang Province, China (No. LH2022E081).

Supplementary Information

Supplementary Information in this paper can be found at: http://tnmsc.csu.edu.cn/download/09-p1855-2023-1165-Supplementary_Information.pdf.

References

[1] YEH J W. Recent progress in high-entropy alloys [J]. Annales de Chimie Science Des Matériaux, 2006, 31: 633−648. https:// doi.org/10.3166/acsm.31.633-648.

[2] MIRACLE D B, SENKOV O N. A critical review of high entropy alloys and related concepts [J]. Acta Materialia, 2017, 122: 448−511. https://doi.org/10.1016/j.actamat.2016.08.081.

[3] TSAI M H. Physical properties of high entropy alloys [J]. Entropy, 2013, 15: 5338−5345. https://doi.org/10.3390/e15125338.

[4] LI Ya-song, LIAO Wei-bing, CHEN Huai-cai, BRECHTL J, SONG Wen-li, YIN Wen, HE Zhan-bing, LIAW P K, ZHANG Yong. A low-density high-entropy dual-phase alloy with hierarchical structure and exceptional specific yield strength [J]. Science China Materials, 2023, 66: 780−792. https://doi. org/10.1007/ s40843-022-2178-x.

[5] FENG R, LIAW P K, GAO M C, WIDOM M. First-principles prediction of high-entropy-alloy stability [J]. NPJ Computational Materials, 2017, 3: 50. https://doi.org/10.1038/ s41524-017-0049-4.

[6] AGRAWAL A, CHOUDHARY A. Perspective: Materials informatics and big data: Realization of the “fourth paradigm” of science in materials science [J]. APL Materials, 2016, 4: 053208. https://doi.org/10.1063/1.4946894.

[7] BECKER C A, WARREN J A. Progress in materials data availability and application: A review [J]. IEEE signal Processing Magazine, 2022, 39(1): 104−108. https://doi.org/10.1109/MSP.2021.3121563.

[8] SHAN Yun-xiao, LI Shu, LI Fu-xiang, CUI Yu-xin, CHEN Ming-hua. Dual-level clustering ensemble algorithm with three consensus strategies [J]. Scientific Reports, 2023, 13: 22617. https://doi.org/10.1038/s41598-023-49947-9.

[9] SHAN Yun-xiao, LI Shu, LI Fu-xiang, CUI Yu-xin, LI Shuai, CHEN Ming-hua, HE Xun-jun. Fuzzy self-consistent clustering ensemble [J]. Applied Soft Computing, 2024, 151: 111151. https://doi.org/10.1016/j.asoc.2023.111151.

[10] SHAN Yun-xiao, LI Shu, LI Fu-xiang, CUI Yu-xin, LI Shuai, ZHOU Ming, LI Xiang. A density peaks clustering algorithm with sparse search and K-d tree [J]. IEEE Access, 2024, 10: 74883−74901. https://doi.org/10.1109/ACCESS.2022.3190958.

[11] WANG Shen, LI Da, XIONG Jun. Prediction of elastic properties of face-centered cubic high-entropy alloys by machine learning [J]. Transactions of Nonferrous Metals Society of China, 2023, 33(2): 518−530. https://doi.org/10.1016/S1003-6326(22)66124-7.

[12] WANG Xin, LI Shu, LIU Feng. Modeling for free dendrite growth based on physically-informed machine learning method. Scripta Materialia. 2024, 242: 1155918. https://doi.org/10.1016/j.scriptamat.2023.115918.

[13] CHEN De-chuang, Li Shu, TAO Ting-peng, LI Shuai, LIU Dong-rong, LIU Xin, CHEN Ming-hua. A machine learning framework for predicting physical properties in configuration space of gate alloys [J]. Materials Today Communications, 2023, 37: 107526. https://doi.org/10.1016/j.mtcomm.2023.107526.

[14] YANG Zhi-yuan, LI Shu, LI Shuai, YANG Jia, LIU Dong-rong. A two-step data augmentation method based on generative adversarial network for hardness prediction of high entropy alloy [J]. Computational Materials Science, 2023, 220: 112064. https://doi.org/10.1016/j.commatsci.2023.112064.

[15] WU Qing-feng, WANG Zhi-jun, HU Xiao-bing, ZHENG Tao, YANG Zhong-sheng, HE Feng, LI Jun-jie, WANG Jin-cheng. Uncovering the eutectics design by machine learning in the Al−Co−Cr−Fe−Ni high entropy system [J]. Acta Materialia, 2020, 182: 278−286. https://doi.org/10.1016/j.actamat.2019.

10.043.

[16] KLIMENKO D, STEPANOV N, RYLTSEV R, ZHEREBTSOV S. Phase prediction in high-entropy alloys with multi-label artificial neural network [J]. Intermetallics, 2022, 151: 107722. https://doi.org/10.1016/j.intermet.2022.107722.

[17] ZHANG Hui-ran, HU Rui, LIU Xi, LI Sheng-zhou, ZHANG Guang-jie, QIAN Quan, DING Guang-tai, DAI Dong-bo. An end-to-end machine learning framework exploring phase formation for high entropy alloys [J]. Transactions of Nonferrous Metals Society of China, 2023, 33(7): 2110−2120. https://doi.org/10.1016/S1003-6326(23)66247-8.

[18] HOU Shuai, SUN Meng-yue, BAI Mei-juan, LIN Dong, LI Yu-jiao, LIU Wei-wei. A hybrid prediction frame for HEAs based on empirical knowledge and machine learning [J]. Acta Materialia, 2022, 228: 117742. https://doi.org/10.1016/j.actamat.2022.117742.

[19] ISAYEV O, OSES C, TOHER C, GOSSETT E, CURTAROLO S, TROPSHA A. Universal fragment descriptors for predicting properties of inorganic crystals [J]. Nature Communications, 2017, 8: 15679. https://doi.org/10.1038/ncomms15679.

[20] GOSCINSKI A, FRAUX G, IMBALZANO G, CERIOTTI M. The role of feature space in atomistic learning [J]. Machine Learning: Science and Technology, 2021, 2(2): 025028. https://doi.org/10.1088/2632-2153/abdaf7.

[21] FENG Shuo, FU Hua-dong, ZHOU Hui-yu, WU Yuan, LU Zhao-ping, DONG Hong-biao. A general and transferable deep learning framework for predicting phase formation in materials [J]. NPJ Computational Materials, 2021, 7: 10. https://doi.org/10.1038/s41524-020-00488-z.

[22] GOODFELLOW I, BENGIO Y, COURVILLE A. Deep learning [M]. Cambridge, MA: The MIT Press, 2016.

[23] LECUN Y, BENGIO Y, HINTON G. Deep learning [J]. Nature, 2015, 521: 436−444. https://doi.org/10.1038/nature14539.

[24] AGBOZO R, JIN W Y. Quantitative metallographic analysis of GCr15 microstructure using mask R-CNN [J]. Journal of the Korean Society for Precision Engineering, 2020, 37: 361−369. https://doi.org/10.7736/JKSPE.019.144.

[25] HU Jing, YANG Song-ran, MAO Jun, SHI Chao-jie, WANG Guang-chuan, LIU Yi-jing, PU Xue-mei. Exploring a general convolutional neural network-based prediction model for critical casting diameter of metallic glasses [J]. Journal of Alloys and Compounds, 2023, 947: 169479. https://doi.org/10. 1016/j.jallcom.2023.169479.

[26] ZHENG Xiao-long, ZHENG Peng, ZHANG Rui-zhi. Machine learning material properties from the periodic table using convolutional neural networks [J]. Chemical Science, 2018, 9: 8426. https://doi.org/10.1039/C8SC02648C.

[27] GUO Qing-wei, XU Xiao-tao, PEI Xiao-long, DUAN Zhi-qiang, LIAW P K, HOU Hua, ZHAO Yu-hong. Predict the phase formation of high-entropy alloys by compositions [J]. Journal of Materials Research and Technology, 2023, 22: 3331−3339. http://dx.doi.org/10.1016/j.jmrt.2022.12.143.

[28] LI Shuai, LI Shu, LIU Dong-rong, YANG Jia, ZHANG Ming-yu. Hardness prediction of high entropy alloys with period table representation of composition, processing, structure and physical parameters [J]. Journal of Alloys and Compounds, 2023, 967: 171735. https://doi.org/10.1016/j.jallcom.2023.171735.

[29] QI J, HOYOS D I, POON S J. Machine learning-based classification, interpretation, and prediction of high-entropy-alloy intermetallic phases [J]. High Entropy Alloys & Materials, 2023, 1: 312−326. https://doi.org/10.1007/s44210-023-00017-9.

[30] DAI Dong-bo, XU Tao, WEI Xiao, DING Guang-tai, XU Yan, ZHANG Jin-cang, ZHANG Hui-ran. Using machine learning and feature engineering to characterize limited material datasets of high-entropy alloys [J]. Computational Materials Science, 2020, 175: 109618. https://doi.org/10.1016/j. commatsci.2020.109618.

[31] WEN Cheng, ZHANG Yan, WANG Chang-xin, XUE De-zhen, BAI Yang, ANTONOV S, DAI Lan-hong, LOOKMAN T, SU Yan-jing. Machine learning assisted design of high entropy alloys with desired property [J]. Acta Materialia, 2019, 170: 109−117. https://doi.org/10.1016/j.actamat.2019.03.010.

[32] WARD L, AGRAWAL A, CHOUDHARY A, WOLVERTON C. A general-purpose machine learning framework for predicting properties of inorganic materials [J]. NPJ Computational Materials, 2016, 2: 16028. https://doi.org/10.1038/npjcompumats.2016.28.

[33] GOLDBERG D E. Genetic algorithms in search, optimization, and machine learning [M]. Reading: Addison-Westly, 1989.

[34] LI Shuai, LI Shu, LIU Dong-rong, ZOU Rui, YANG Zhi-yuan. Hardness prediction of high entropy alloys with machine learning and material descriptors selection by improved genetic algorithm [J]. Computational Materials Science, 2022, 205: 111185. https://doi.org/10.1016/j.commatsci.2022. 111185.

[35] ZHANG Yan, WEN Cheng, WANG Chang-xin, ANTONOV S, XUE De-zhen, BAI Yang, SU Yan-jing. Phase prediction in high entropy alloys with a rational selection of materials descriptors and machine learning models [J]. Acta Materialia, 2020, 185: 528−539. https://doi.org/10.1016/j.actamat.2019.11.067.

[36] XIONG Jie, SHI San-qiang, ZHANG Tong-ti. Machine learning of phases and mechanical properties in complex concentrated alloys [J]. Journal of Materials Science & Technology, 2021, 87: 133−142. https://doi.org/10.1016/j.jmst.2021.01.054.

[37] WANG Zhi-jun, HUANG Yun-hao, YANG Yong, WANG Jin-cheng, LIU C T. Atomic-size effect and solid solubility of multicomponent alloys [J]. Scripta Materialia, 2015, 94: 28−31. https://doi.org/10.1016/j.scriptamat.2014.09.010.

[38] WANG Zhi-peng, FANG Qi-hong, LI Jia, LIU Bin, LIU Yong. Effect of lattice distortion on solid solution strengthening of BCC high-entropy alloys [J]. Journal of Materials Science & Technology, 2018, 34: 349−354. https://doi.org/10.1016/j.jmst. 2017.07.013.

[39] GUO S. Phase selection rules for cast high entropy alloys: An overview [J]. Materials Science and Technology, 2015, 31: 1223−1230. https://doi.org/10.1179/1743284715Y.0000000018.

[40] DONG Yong, LU Yi-ping, JIANG Li, WANG Tong-min, LI Ting-ju. Effects of electro-negativity on the stability of topologically close-packed phase in high entropy alloys [J]. Intermetallics, 2014, 52: 105−109. https://doi.org/10.1016/j.intermet.2014.04.001.

[41] RYU C W, KANG D H, JEON S, LEE G W, PARK ES. Accurate quantification of glass-forming ability by measuring effective volume relaxation of supercooled melt [J]. APL Materials, 2017, 5(10): 106103. https://doi.org/10.1063/ 1.4999357.

[42] SAAL J E, KIRKLIN S, AYKOL M, MEREDIG B, WOLVERTON C. Materials design and discovery with high-throughput density functional theory: The open quantum materials database (OQMD) [J]. JOM, 2013, 65: 1501−1509. https://doi.org/10.1007/s11837-013-0755-4.

[43] LUCKABAUER M, KÜHN U, ECKERT J, SPRENGEL W. Specific volume study of a bulk metallic glass far below its calorimetrically determined glass transition temperature [J]. Physical Review B, 2014, 89(17): 173113. https://doi.org/10.1103/PhysRevB.89.174113.

[44] LOUZGUINE-LUZGIN D V, FUKUHARA M, INOUE A. Specific volume and elastic properties of glassy, icosahedral quasicrystalline and crystalline phases in Zr−Ni−Cu−Al−Pd alloy [J]. Acta Materialia, 2007, 55(3): 1009−1015. https://doi. org/10.1016/j.actamat.2006.09.013.

[45] COHEN M H, GREST G S. Liquid-glass transition, a free-volume approach [J]. Physical Review B, 1979, 20: 1077−1098. https://doi.org/10.1103/PhysRevB.20.1077.

[46] MUKHERJEE S, SCHROERS J, ZHOU Z, JOHNSON W L, RHIM W K. Viscosity and specific volume of bulk metallic glass-forming alloys and their correlation with glass forming ability [J]. Acta Materialia, 2004, 52(12): 3689−3695. https://doi.org/10.1016/j.actamat.2004.04.023.

[47] ZHU J M, ZHANG H F, FU H M, WANG A M, LI H, HU Z Q. Microstructures and compressive properties of multi-component AlCoCrCuFeNiMo$_x$ alloys [J]. Journal of Alloys and Compounds, 2010, 497: 52−56. https://doi.org/10.1016/j.jallcom.2010.03.074.

[48] ZHU J M, FU H M, ZHANG H F, WANG A M, LI H, HU Z Q. Microstructures and compressive properties of multicomponent AlCoCrFeNiMo$_x$ alloys [J]. Materials Science and Engineering: A, 2010, 527(26): 6975−6979. https://doi.org/10.1016/j.msea.2010.07.028.

[49] LIU W H, LU Z P, HE J Y, LUAN J H, WANG Z J, LIU B, LIU Y, CHEN M W, LIU C T. Ductile CoCrFeNiMo$_x$ high entropy alloys strengthened by hard intermetallic phases [J] Acta Materialia, 2016, 116: 332−342. https://doi.org/10.1016/j.actamat.2016.06.063.

[50] LIU Yong-qin, ZHU Man, YAO Li-juan, JIAN Zeng-yun. Evolution of microstructure and mechanical properties of the CoFeNiMnMo$_x$ high-entropy alloys [J]. Crystals, 2022, 12(8): 1124. https://doi.org/10.3390/cryst12081124.

[51] NA S M, YOO J H, LAMBERT P K, JONES N J. Room-temperature ferromagnetic transitions and the temperature dependence of magnetic behaviors in FeCoNiCr-based high-entropy alloys [J]. AIP Advances, 2018, 8: 056412. https:// doi.org/10.1063/1.5007073.

[52] YURCHENKO N, STEPANOV N, SALISHCHEV G. Laves- phase formation criterion for high-entropy alloys [J]. Materials Science and Technology, 2017, 33: 17−22. https://doi.org/10. 1080/02670836.2016.1153277.

[53] FARRARO R, MCLELLAN R B. Temperature dependence of the Young's modulus and shear modulus of pure nickel, platinum, and molybdenum [J]. Metallurgical Transactions A, 1977, 8: 1563−1565. https://doi.org/10.1007/BF02644859.

[54] TAKEUCHI A, INOUE A. Classification of bulk metallic glasses by atomic size difference, heat of mixing and period of constituent elements and its application to characterization of the main alloying element [J]. Materials Transactions, 2005, 46: 2817−2829. https://doi.org/10.2320/matertrans.46.2817.

[55] DIEBOLD F X, MARIANO R S. Comparing predictive accuracy [J]. Journal of Business & Economic Statistics, 1995, 13(3): 253−263. https://doi.org/10.1080/07350015.1995.10524599.

[56] ZHAO Yu-hong. Editorial: Phase field method and integrated computing materials engineering [J]. Frontiers in Materials, 2023, 10: 1145833. https://doi.org/10.3389/fmats.2023.1145833.

[57] CHEN Long-qing, ZHAO Yu-hong. From classical thermodynamics to phase-field method [J]. Progress in Materials Science, 2022, 124: 100868. https://doi.org/10.1016/j.pmatsci. 2021.100868.

[58] ZHAO Yu-hong. Understanding and design of metallic alloys guided by phase-field simulations [J]. NPJ Computational Materials, 2023, 9: 94. https://doi.org/10.1038/s41524-023-01038-z.

[59] VILALTA P C, SHEIKHOLESLAMI S, SALEME RUIZ K, YEE X C, KOSLOWSKI M. Machine learning for predicting the critical yield stress of high entropy alloys [J]. Journal of Engineering Materials and Technology, 2021, 143: 021005. https://doi.org/10.1115/1.4048873.

[60] ZHAO Yu-hong, LIU Ke-xin, HOU Hua, CHEN Long-qing. Role of interfacial energy anisotropy in dendrite orientation in Al−Zn alloys: A phase field study [J]. Materials & Design, 2022, 216: 110555. https://doi.org/10.1016/j.matdes.2022.110555.

[61] LIU Dong, PEI Jia-qi, HOU Hua, NIU Xiao-feng, ZHAO Yu-hong. Optimizing solidification dendrites and process parameters for laser powder bed fusion additive manufacturing of GH3536 superalloy by finite volume and phase-field method [J]. Journal of Materials Research and Technology, 2023, 27: 3323−3338. https://doi.org/10.1016/j.jmrt.2023.10.188.

[62] ZHU Yu-quan, XU Tao, WEI Qing-hua, MAI Jia-wei, YANG Hong-xin, ZHANG Hui-ran, SHIMADA T, KITAMURA T, ZHANG Tong-yi. Linear-superelastic Ti−Nb nanocomposite alloys with ultralow modulus via high-throughput phase-field design and machine learning [J]. NPJ Computational Materials, 2021, 7: 205. https://doi.org/10.1038/s41524-021-00674-7.

[63] YANG Kai-qi, CAO Yi-fan, ZHANG You-tian, FAN Shao-xun, TANG Ming, ABERG D, SADIGH B, ZHOU Fei. Self-supervised learning and prediction of microstructure evolution with convolutional recurrent neural networks [J]. Patterns, 2021, 2(5): 100243. https://doi.org/10.1016/j.patter.2021.100243.

基于成分、常见物理、元素性质描述符和元素周期表表示的高熵合金相分类

李 帅[1]，杨 佳[2]，李 述[3]，刘东戎[1]，张洺玉[3]

1. 哈尔滨理工大学 材料科学与化学工程学院，哈尔滨 150080；
2. 哈尔滨理工大学 计算机科学与技术学院，哈尔滨 150080；
3. 哈尔滨理工大学 电气与电子工程学院 工程电介质及其应用教育部重点实验室，哈尔滨 150080

摘 要：相分类对于高熵合金的设计具有明显的指导意义。为了实现互斥和非互斥相分类，收集了成分描述符、常用物理参数描述符、元素属性描述符以及通过卷积神经网络从元素周期表表示(PTR)中提取的描述符。在具有丰富信息的特征中进行适当选择有助于相分类。基于随机森林，四分类的准确率和五分类的平衡准确率分别提高到 0.907 和 0.876。通过可解释性分析总结了 4 个最重要特征的作用，并发现一个新重要特征。通过对$(CoFeNiMn)_{1-x}Mo_x$进行相预测，证明了模型外推能力和 Mo 元素的影响。相信息有助于硬度预测，分类结果与硬度数据的 PTR 相耦合，预测误差(均方根误差)降低到 56.69。

关键词：高熵合金；相分类；特征工程；周期表表示法；卷积神经网络；硬度预测

(Edited by Wei-ping CHEN)