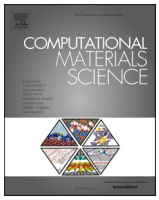
Computational Materials Science 205 (2022) 111185

Contents lists available at ScienceDirect

Computational Materials Science

journal homepage: www.elsevier.com/locate/commatsci



Editor's Choice

Hardness prediction of high entropy alloys with machine learning and material descriptors selection by improved genetic algorithm

Shuai Li [a], Shu Li [b,*], Dongrong Liu [a,**], Rui Zou [c], Zhiyuan Yang [b]

[a] School of Materials Science and Engineering, Harbin University of Science and Technology, Harbin, Heilongjiang 150080, People's Republic of China
[b] School of Electrical and Electronic Engineering, Harbin University of Science and Technology, Harbin, Heilongjiang 150080, People's Republic of China
[c] School of Computer Science and Technology, Harbin University of Science and Technology, Harbin, Heilongjiang 150080, People's Republic of China

ARTICLE INFO

Keywords:
Genetic algorithm
Feature selection
High entropy alloy
Hardness prediction
Machine learning
Ensemble learning

ABSTRACT

With the coming of the age of artificial intelligence and big data, machine learning (ML) has been showing powerful potentials for properties prediction of materials. For achieving satisfying prediction performance, rational feature selection plays a key role along with a suitable ML model itself. In the present work, the traditional genetic algorithm (GA) has been further improved to serve as a feature selection method for the hardness prediction problem of high entropy alloys (HEAs). The concepts of feature importance and gene manipulation were introduced into the improved GA to make it more comprehensible. Comparative analysis demonstrated that the improved GA is superior to the traditional GA in the aspects of accuracy, stability and efficiency obviously. A comparison with other typical feature selection methods was also made. In addition, ML model selection was discussed with the composition feature or the optimal physical feature combination selected by the improved GA. Finally, in order to elevate the prediction ability of ML model, the stacking method as an ensemble learning strategy was proposed in Al-Co-Cr-Cu-Fe-Ni HEAs hardness prediction. It was shown that the prediction errors are successfully lowered. This ML framework could be regarded as a method with general applicability to select suitable ML model and material descriptors, for designing various materials with excellent properties and complex composition.

1. Introduction

It is difficult to design materials by traditional experiments or traditionally theoretical calculation [1,2]. When the material properties were explored by experiments according to trial and error [3–5], it would cost a lot of time and raw materials, and even has high requirements for equipment [6]. As for complex theoretical calculation [7,8], such as density functional theory (DFT) [9,10] and phase field simulation [11,12], although there are some works focusing on the high-throughput question [13–17], there remains bottleneck restriction for elevating computational efficiency [1]. Thus, it is not satisfactory in addressing complex search problems of materials that involve massive compositional spaces. In contrast, machine learning (ML) algorithm could effectively infer the relationship between material descriptors and target property through constructing a surrogate model without explicit programming. Recently, ML has been extensively used to study properties of materials [18–24], for example, properties prediction in high entropy alloys (HEAs) [18,19], crystal structures prediction of zeolites [20] and binary compounds [21] as well as compositional design of high-performance copper alloys [22] and super alloys [23,24]. ML has been showing powerful potentials in material design and its research is remarkably meaningful.

For ML model to be used successfully, the rational selections of ML model and feature combination are the relatively most important factors determining prediction ability. As a preliminary screening method for feature combination, the Pearson correlation coefficient (PCC) is commonly calculated to remove some features among feature groups with highly linear correlation [25–29]. For this treatment, however, there is also possibility that some key features are removed. The enumeration method could avoid this problem and is simple to implement, but it is very inefficient especially with a huge feature space. In order to increase searching efficiency, Zhang et al. used sequential backward selection method (SBS) to find the best feature subset to explore copper alloys with good ultimate tensile strength (UTS) and

* Corresponding author.
** Corresponding author.
E-mail addresses: lishu@hrbust.edu.cn (S. Li), dongrongliu@hrbust.edu.cn (D. Liu).

https://doi.org/10.1016/j.commatsci.2022.111185
Received 23 September 2021; Received in revised form 30 December 2021; Accepted 2 January 2022
Available online 18 January 2022

electric conductivity (EC) simultaneously [30]. SBS can accelerate the optimization procedure, but it is a greedy algorithm, which also possibly ignores certain important feature combination. The recursive feature elimination (RFE) [29,31] and sequential forward selection method (SFS) are the same as SBS [32]. Moreover, Ghiringhelli et al utilized LASSO (Least absolute shrinkage and selection operator regression) to select the optimal feature subset from available 10,000 material descriptors for predicting the crystal structure of semiconductors [33]. These ML models, including LASSO, linear regression, GBDT (Gradient Boosting Decision Tree) and random forest, could effectively find suitable feature subset directly by themselves [34]. However, there is another limitation that these models, as feature selection methods, could only provide an unchangeable feature combination which cannot be further optimized to match various object ML models. In addition, dimensionality reduction algorithms such as principal component analysis (PCA) and fisher algorithm can compress the feature space into a small set, but they are lacking of interpretability [35,36].

For the sake of finding the global optimal feature subset accurately and quickly, Zhang et al have tried the genetic algorithm (GA) for feature selection from 70 related features and effectively classified the phases of HEAs [37]. GA is one of the most effective methods for feature selection. It simulates the biological evolution mechanism of "natural selection, survival of the fittest", so it could find the global optimal result by updating the individuals in population iteratively without trying each possible feature combination [38]. However, GA is a stochastic global optimization algorithm, and thus the result of GA is strongly depended on the quality of initial population which is normally generated randomly, as well as the iteration times for reaching steady state. There are lots of improved GAs proposed for various specific problems [39–42], in which some GAs [34,37,43–46], such as svmGA [43], rfGA [44] and GARS [34] are designed to focus on feature selection. To the author's knowledge, however, these methods [34,37,43–46] are mainly aimed to classify datasets.

Furthermore, HEAs have been demonstrated to be a novel type of alloys with various excellent physical and mechanical properties, such as high hardness [47,48], good wear resistance [49], great magnetic performance [50], exceptional fracture toughness at low temperature [51] etc. Due to the complex composition structure for HEAs, ML is highly suitable for making composition design. In order to find appropriate input features for ML model, compositions [18,52–54], physical features defined according to the domain knowledge and simple parameter models [55,56], as well as physical features selected by the feature selection methods mentioned above have been adopted respectively. Therefore, it is strongly recommended to develop a highly effective way for feature selection of HEAs to achieve better excellent performance. In this work, an improved GA for feature selection in ML models was proposed to predict the hardness of Al-Co-Cr-Cu-Fe-Ni HEAs. Comparisons of the improved GA newly developed with traditional GA and other representative feature selection methods have also been made. Finally, one of the ensemble strategies, the stacking method was used to further enhance the prediction ability of ML model. This framework could be regarded as a general method to select suitable ML model and material descriptors, for designing various materials with high property and complex composition.

2. Method

2.1. Reduce the number of feature combinations

In order to increase searching efficiency, it is reasonable to make a preliminary screening for feature combinations, especially with the huge number of available feature combinations. Firstly, a certain number (a) of feature subsets with different random numbers of features are sampled. The value of a should be large enough to make these feature subsets more representative. Then prediction errors are evaluated based on ML model with the feature subsets selected, respectively. Secondly, the optimal feature number (b) in feature subset is determined according to the variation trend of prediction errors with the feature numbers in feature subsets. Here, the value of b should be moderate for balancing the prediction ability and the complexity of model. In the following, the feature number in feature subsets is fixed as b to search the optimal feature combination from the whole possible feature combination space.

2.2. The common feature selection procedure of GA

For GA to be used as a feature selection method in ML, the feature combination should be coded as a binary string with either 1 or 0 for each bit. The value 1 represents that the corresponding feature is selected, vice versa. Such a string is known as an individual and a bit in each string is called a gene. The set of individuals defines a population. The procedure of GA could be divided into the five parts as shown in Figs. 1 and 2, including the creation of initial population, fitness calculation, selection, crossover and mutation. Beginning with the initial population creation, the other four steps repeat circularly until the iteration times reach a preset value (i). The final result of GA is the individual with the minimum prediction error found in the whole iteration process.

2.3. Traditional GA

As a comparison with the present GAs improved in the next section, the traditional GA, marked by GA-0, is specified as follows in the present study, which is also shown in Fig. 1. The initial population with n individuals is generated randomly, under the condition that the number of genes with value 1 is fixed as b for any individuals. The fitness values of individuals are calculated by the reciprocal of the prediction errors which are evaluated based on ML model. In the selection step, the population is modified to a new one by the roulette wheel selection. Note that there is a higher possibility for these individuals with relatively high fitness value to be selected repeatedly. According to the crossover probability (c), a certain number of individuals are further chosen, for which each pairs of individuals (f and m) are exchanged between the two individuals, i.e. the half genes with value 1 in f and 0 in m as well as the half genes with value 0 in f and 1 in m are exchanged simultaneously. For mutation, any two genes with values 1 and 0 are changed to their opposite values, respectively, in each individuals selected by mutation probability (p).

2.4. Improved GAs

As mentioned above, there are very few of the GAs designed to focus on feature selection in ML modeling for regression problem. For this problem, the core objective is to find the global optimal feature subset accurately, quickly and steadily. In genetic algorithms, a feature subset is regarded as an individual from population and the importance of an individual is characterized by the fitness value for selection. Similarly, the concept of *feature importance* is introduced into the present work, which could be assumed as the fitness value of a feature (a gene in individuals) since an individual is made up a certain number of genes. In this section, the traditional GA-0 is further improved by *genes manipulation* based on the feature importance to elevate accuracy, stability and searching efficiency simultaneously.

The first version of the improved GA is specified as follows, which is labeled as GA-1 and shown in Fig. 2. For initial population creation, a certain number (d) of binary strings (individuals) are firstly generated randomly. Then, as fitness values of individuals, the reciprocals of prediction errors of the top e strings with prediction error minimum are calculated. The feature importance is obtained by summing these gene values (0 or 1), weighted by the fitness value of corresponding individual, with the same position traversing every individual in population (top e individuals with prediction error minimum). Moreover it is repeated to select b features, which determine one individual, by the

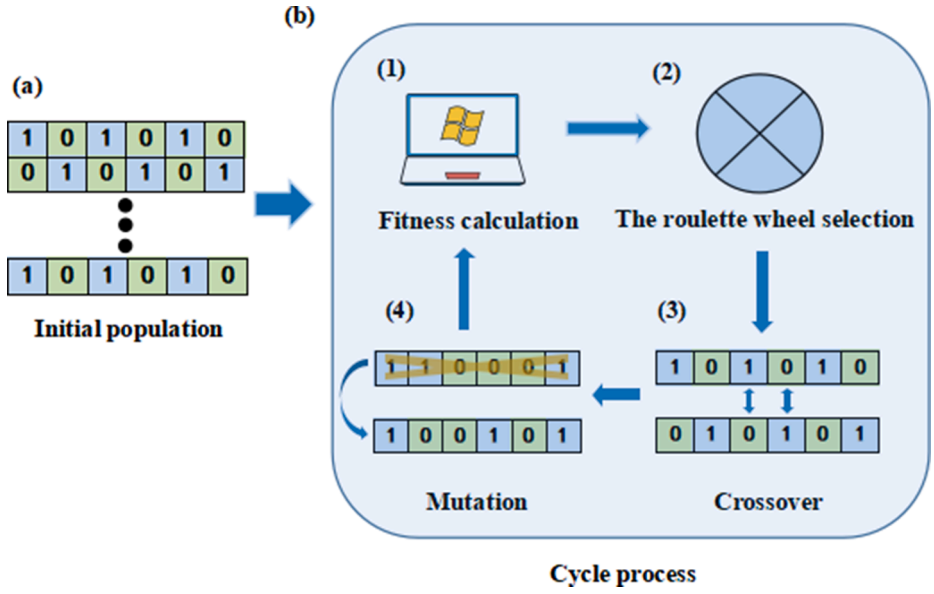


Fig. 1. The whole procedure of GA-0, (a) the creation of initial population, (b) the cycle process between four steps, including the fitness calculation, the roulette wheel selection, the crossover and the mutation.

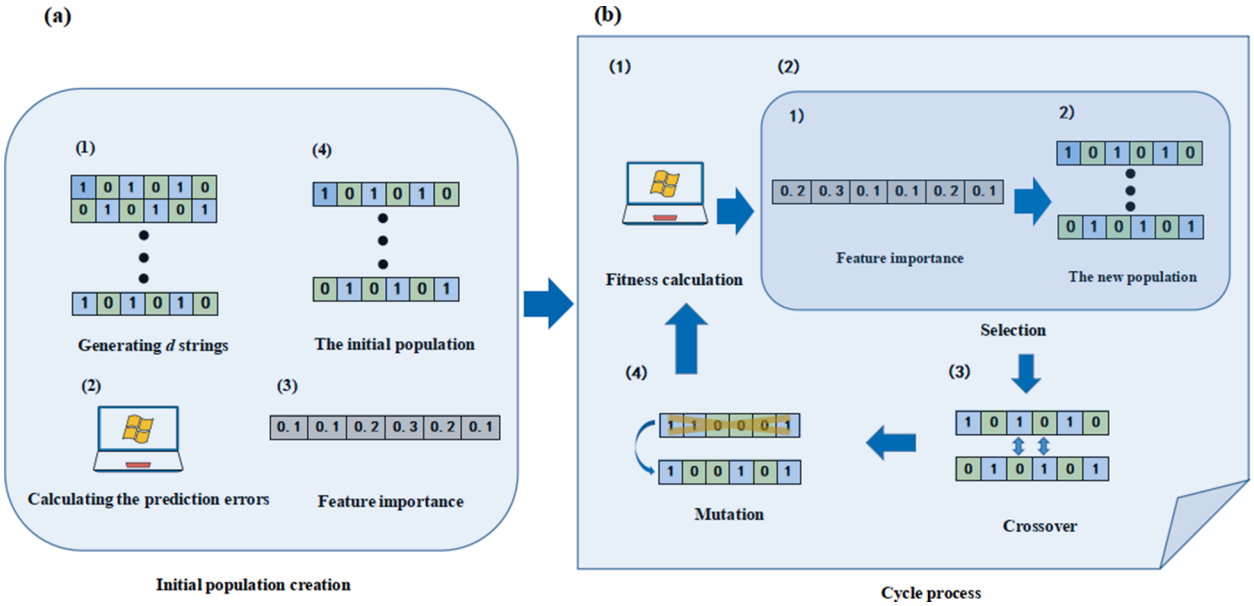


Fig. 2. The whole procedure of GA-1, (a) the initial population creation by feature importance, (b) the cycle process among four steps, including the fitness calculation, the selection based on feature importance, the crossover and the mutation.

roulette wheel selection method based on the feature importance, until the initial population with n individuals is constructed. In the selection step, the methods of calculating feature importance and updating new population with n individuals are the same as what is adopted in creation of initial population. Furthermore, the treatment for crossover in GA-0 is maintained in the present algorithm. Finally, for mutation step, in each individuals selected by mutation probability p, a gene is chosen from all of these genes with value 0 by the roulette wheel selection method and the feature importance, while a gene is chosen from all of these genes with value 1 by the same method, but with the reciprocal of the feature importance. Then these two genes are exchanged for mutation. Up till now, the improved GA-1 has been described.

There are three different versions of the improved GAs listed in Table 1, including GA-1, GA-2, and GA-3, to make a comparative analysis on the influence derived from these different parts improved in GAs.

Table 1
The differences among the three versions of improved GAs.

Improved GAs	GA-1	GA-2	GA-3
Initial population	Improved	Traditional	Improved
Fitness	Traditional	Traditional	Traditional
Selection	Improved	Improved	Traditional
Crossover	Traditional	Traditional	Traditional
Mutation	Improved	Improved	Improved

For describing the improved GAs conveniently, the concept of *genes manipulation* is introduced and defined by the two successive steps, i.e. calculating the feature importance for the present population and updating the population by the roulette wheel selection method based on the feature importance. In order to illustrate the effect of improving

the initial population quality, different from GA-1, GA-2 uses the traditional random method to generate initial population as shown in GA-0. For showing the importance of the genes manipulation, based on GA-1, GA-3 adopts the traditional selection operator in GA-0 to remove the genes manipulation. As indicated in Table 1, the treatments for other parts in GA-2, and GA-3 are the same as that described in GA-1.

3. Results and discussion

3.1. HEAs dataset and physical features

The 205 Al-Co-Cr-Cu-Fe-Ni HEAs data collected from literatures [18,57] were used as original data set. As the entire feature pool, 20 physical features [18] were adopted, among which 11 features are related to phase formation and indirectly influence the HEAs hardness, while other 9 features are bound up with the mechanical properties of HEAs. The former feature group includes the difference in atomic radii (δr) between elements, the difference in electronegativity ($\Delta\chi$) between elements, the valence electron concentration (VEC), the mixing enthalpy (ΔH), the configuration entropy (ΔS), the local electronegativity mismatch (D.χ) between elements, the cohesive energy (E_c), the number of itinerant electrons (e/a) and the parameters Ω, Λ, γ. It is worthy to mention that, the value of feature e/a is controversial for transition metal elements. Thus two definitions for e/a were adopted in the present work, marked by e_1/a and e_2/a. The calculation of e_1/a does not consider d-electrons [18], while e_2/a is evaluated by setting the value as 0 for transition metal element based on the assumption that the ability of the transition metal elements to hold electrons is comparable to the ability to release electrons. The latter group includes the modulus mismatch (η), the local size mismatch (D.r), the energy term in the strengthening model (A), the Peierls-Nabarro factor (F), the sixth square of the electron work function (ω), the shear modulus (G), the local modulus mismatch (δG), the difference in shear modulus (D.G) and the lattice distortion energy (μ).

3.2. ML model selection

For successfully finding the relatively better ML model to predict the hardness of HEAs, several well-known models are taken into account, including linear regression (LR), polynomial regression (PR), LASSO regression, ridge regression (RR), support vector regression with a linear kernel (SVR-L), a radial basis function kernel (SVR-R) and back propagation neural network (BPNN). Tree-based models, such as regression tree and random forest, are not included in the present work due to lack of extrapolating ability, which would further lead to failure in exploring unknown materials with higher performance than original data. In order to compare different feature selection methods, a ML model should be specified in advance. Here, the compositions of HEAs are employed as input feature values to specify a suitable ML model rapidly and simply [18]. For different ML models, the ten rounds mean value of the root mean square error (RMSE) which evaluated by ten-fold cross validation are calculated and shown in Fig. 3. This method of calculating predictive errors is also adopted in the present work for all of the following sections. In Fig. 3, it can be seen that SVR-R is a relatively more suitable ML model. Thus, SVR-R is adopted in the next section to demonstrate the superiority of the feature selection method proposed by the present work.

3.3. Feature selection

3.3.1. The optimal number of features

As mentioned in Section 2.1, in order to make a preliminary screening of feature combinations, a feature combinations with different random numbers of features should be sampled randomly. Here $a = 500$ is set and the prediction errors calculated with SVR-R are shown in Fig. 4. With increasing features number, the mean error gradually

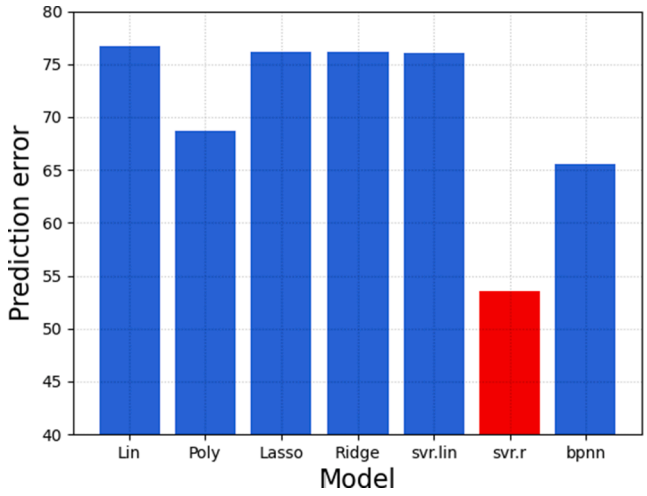


Fig. 3. The prediction errors of different ML models which are evaluated by the composition feature. The SVR-R model has the relatively lowest prediction error.

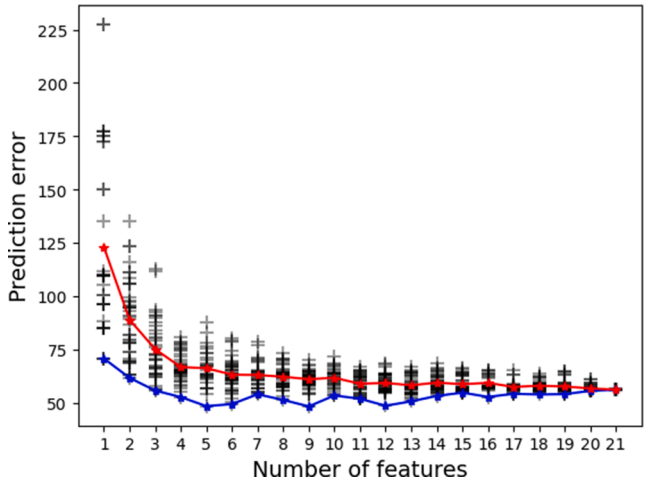


Fig. 4. The prediction error of ML model as a function of the number of features. Every spot represents a feature subset which is generated randomly; the red line and the blue line indicate the mean and the lowest prediction errors for the spots with a same given number of features, respectively.

decreases and converges, and the trend is slow down from the number of 4 significantly. In addition, there is little difference between the lowest error of 4 and the lowest error in Fig. 4. The value of optimal feature number (b) in the feature subset should be moderate to balance the prediction ability and the complexity of the ML model, so on the premise of ensuring relatively low mean and lowest prediction errors, the lower number of features would be better. Thus, according to the experience and comprehensive analysis, it is very possible to find excellent feature combinations in feature subsets with 4 features, and the feature number b is fixed as 4.

3.3.2. The results of GAs

In the present calculation, the individual number n in a population is taken to be 100, and the crossover probability c, the mutation probability p as well as the total iteration times i are set to be 0.2, 0.01 and 50, respectively. Moreover, the number d of binary strings randomly generated is 200, and the number e of binary strings selected subsequently is 50, for the initial population creation of GA-1 and GA-3. For other case, these parameter values could be adjusted to satisfy actual situation.

In order to estimate the ability of finding good feature subset for different GAs, the concept of *accuracy*, defined as the possibility of finding the feature combination with the first level prediction errors in any GA run, is introduced. Here, since the differences of prediction errors are not obvious between adjacent feature subsets, and the prediction error for a certain feature combination has a certain degree of randomness, these prediction errors considered are divided as five levels from the minimum 47.9 with interval 1. On the one hand, in order to regard the feature combinations with similar prediction errors as the same level, the interval value is obtained by the 2% of the lowest prediction error. ($47.9 * 2\% = 0.958 \approx 1$). In addition, the mean value of ten times ten-fold cross validation is re-calculated 60 times, independently. The prediction errors obey the normal distribution approximately. Moreover, according to the calculation of the normal distribution, the probability of the prediction errors located in the range of 0.5 above and below the mean prediction error is up to 0.78. Thus, for the sake of decreasing the influence of fluctuation in prediction error calculation, the interval of prediction errors is fixed as 1. These GAs are run 50 times respectively and the results are shown in Fig. 5. Although there are some results of GA are the local optimal result, the improved GA enhances the performance of GA to find the feature combination with the first level obviously. The probabilities of finding the feature combination with the first level of prediction errors for different GAs are listed in Table 2. It could be concluded that, the improved GA-1 possesses a higher degree of accuracy than the traditional GA-0 remarkably. In addition, other improved GAs are also better than the traditional GA-0.

Since GA is a stochastic optimization algorithm, the results for each run are usually different. According to statistics, the information entropy is defined by $\mathrm{Ent}(X) = -\sum_{j=1}^{N} p(x_j) log_2\left(p(x_j)\right)$ to describe the discrete degree of a discrete random variable. Here, $N = 5$ indicating five levels of prediction error and $p(x_j)$ is the probability of feature combination x_j with prediction error level j. The lower the value of information entropy $\mathrm{Ent}(X)$ is, the more concentrated and stable the results are for multiple runs with GAs. The calculated values of information entropy for GA-0, GA-1, GA-2 and GA-3 by 50 runs are 1.776, 1.359, 1.466 and 1.406 respectively. Therefore, improved GA-1 is obviously more stable than other GAs. In Fig. 5, it can also be seen that the results with GA-1 is clearly concentrated on level 1.

Table 2
The probabilities of finding the feature combination with the first level of prediction errors for different GAs.

GA	GA-0	GA-1	GA-2	GA-3
Probability	20%	68%	52%	28%

In order to compare *efficiencies* of different GAs, these relationships between prediction error and iteration number are shown in Fig. 6. Considering the algorithmic stability, each GA is run 50 times independently and five results among them are shown here randomly. It is clearly indicated that prediction errors of GAs decrease from relatively high values to relatively low values with increasing the iteration number and finally reach their stable values. It could be seen that, due to the GA has a certain degree of randomness in each step, it is possible to avoid local optimal result to be the final result in each run. The minimum iteration number which results in a stable prediction error for GA runs could be used to estimate the *convergence speed* of GAs. The lower the minimum is, the higher the convergence speed is. These mean minimum values with 50 times independent runs for GA-0, GA-1, GA-2 and GA-3 are 25.92, 25.14, 26.12 and 25.34, respectively. It can be seen that the difference of convergence speed between GAs is not obvious.

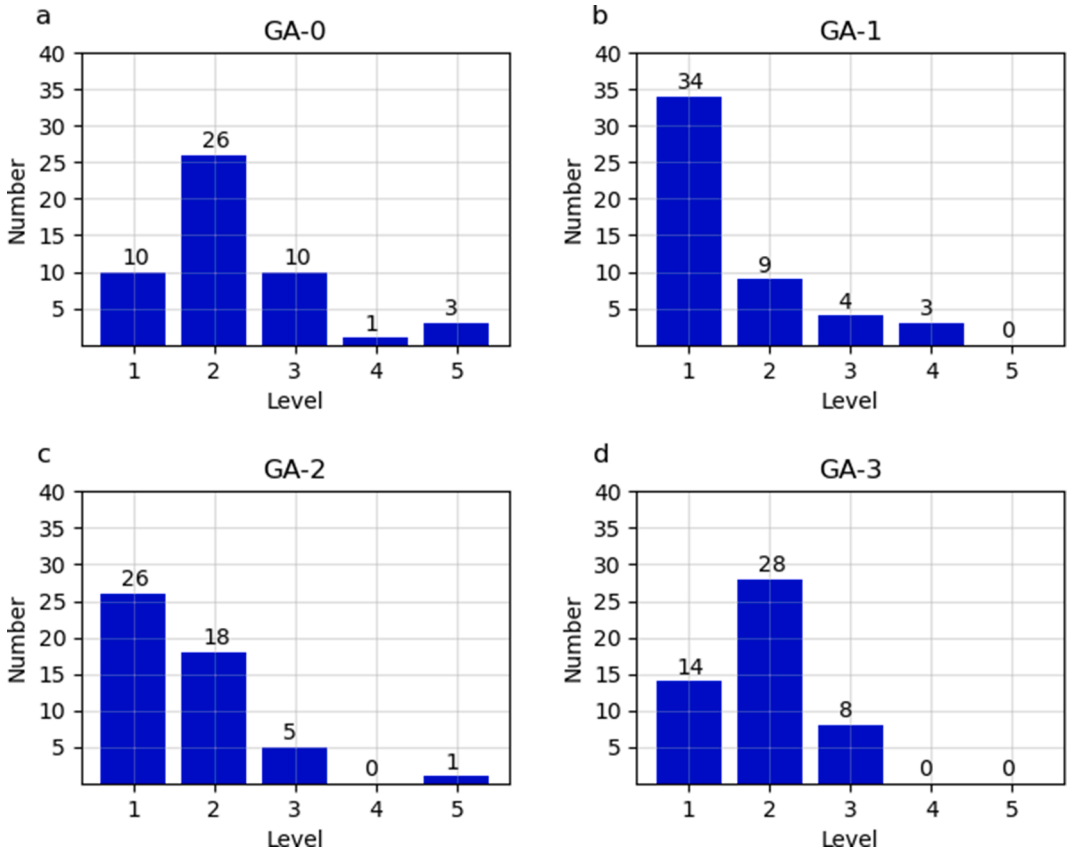


Fig. 5. The feature selection results by 50 times GAs runs. (a) - (d) indicate the feature subset numbers located in the top five levels of predication errors calculated with GA-0, GA-1, GA-2 and GA-3 respectively. The first level is in the range of 47.9–48.9, and the level 2 includes these prediction errors in 48.9–49.9 with interval 1 and so on.

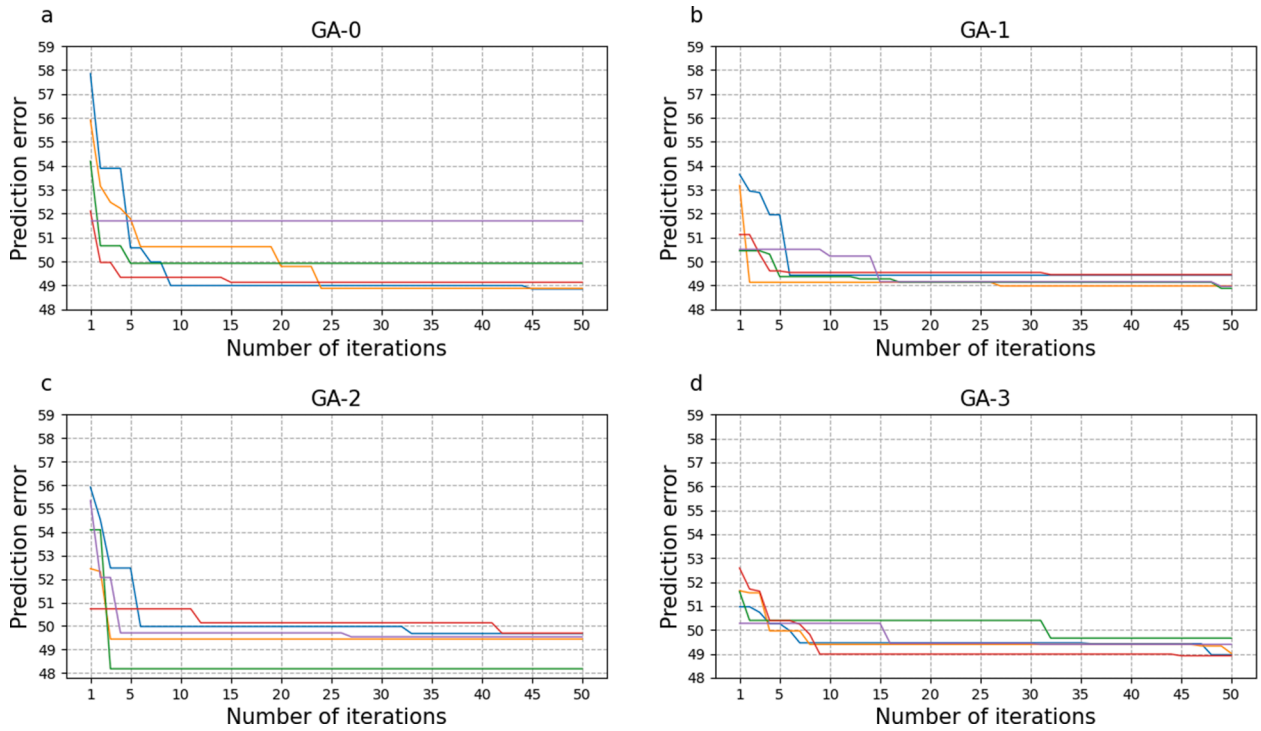


Fig. 6. The optimization processes of several GAs for five independent runs. (a) - (d) indicate prediction errors as functions of the number of iterations for five times independent runs of GA-0, GA-1, GA-2 and GA-3 respectively. During iterations, the minimum is served as the prediction error given by this iteration.

However, this does not mean that computational efficiencies to obtain final results are also similar for these four GAs. Due to the randomness in GA, the user could run GA several times to ensure that the global optimal results are obtained. Under the condition that the algorithm is relatively more stable, for obtaining the feature subset with level 1 the required independent run times would be relatively less and vice versa. Thus, in reality, the algorithmic stability should also be taken into account. Setting the reliability to be 90%, combined with the values of probability listed in Table 2, the minimum running times for GA-0, GA-1, GA-2 and GA-3 are 11, 2, 4 and 7 respectively. It is indicated that the final computational efficiency for GA-1 is remarkably higher than others.

Therefore, the improved GA-1 proposed by the present work is remarkably superior to the traditional GA-0 in all of these aspects of the accuracy, stability and efficiency. This is contributed by the improvement for operations of the initial population creation and selection (gene manipulation) in GA. Compared with GA-1, the GA-2 removes the modification of initial population creation, and the GA-3 does not take the gene manipulation. Thus, both the initial population creation and gene manipulation are important in elevating the three qualities of GAs. Furthermore, the active effect of gene manipulation is more obvious than that of the improvement of initial population creation for accuracy and efficiency, while their influence is comparable for stability.

3.3.3. *The comparison with other feature selection methods*

In order to make a comparison with other typical and common feature selection methods, the exhaustive search method, SFS, SBS and PCC methods are applied to the present feature selection calculation for HEAs hardness prediction. Based on the exhaustive method [18], prediction errors given by all possible feature combinations with 4 features are calculated, which are listed in the supplementary materials, in detail. There is no doubt that the satisfying result can be obtained by the exhaustion method. However, its computational efficiency is extremely poor, especially the feature number is large enough. There is a good agreement between the results given by the exhaustion method and GA-1 for optimal feature combinations. In addition, SFS is a method starting with an empty feature set and stepwise adding one feature to minimize the prediction error of the existing feature subset, while SBS is a method starting with a full feature set and stepwise removing one feature to minimize the prediction error of the existing feature subset [30]. The results of SFS and SBS are [*VEC*, γ, e_1/a, ω] and [ΔS, D.r, ω, G], which belong to the prediction error level 5 and level 4, respectively. This is due to the fact that they are greedy algorithm, which possibly falls into the local optimal result [32].

The PCC method is often used as a preliminary screening method for feature selection by removing the irrelevant and redundant features [18,25–29]. In Fig. 7, the PCC values for different feature pairs and the prediction errors calculated by SVR-R with different single features are shown. For any feature pair, the larger the absolute value of PCC is, the stronger the correlation is. Here, feature pairs with PCC absolute values larger than 0.95 are considered as highly correlated. For the present HEAs data, there are three groups of features with strong correlation. In each group, the features with relatively larger prediction errors are removed. Thus these features, E_C, A, μ and D.χ, are finally deleted, as shown in Fig. 7 (b). In addition, if removing the features with lower correlation coefficient absolute values with hardness values, the result are same as above, as shown in Fig. 7 (c). However, according to the feature selection results based on exhaustive method, the feature subsets [*VEC*, ΔS, γ, μ] in prediction error level 1 and [*VEC*, ΔS, D.r, μ], [*VEC*, ΔS, E_C, D.G] as well as [*VEC*, ΔS, F, μ] in level 2 include the features E_C and μ. It should be noted that the total count of the feature subsets is only 11 in level 1 and 2. Therefore, for PCC method there is a certain possibility that some potentially important feature subsets are neglected.

3.4. *Feature importance*

For GAs proposed by the present work, the concept of feature importance has been introduced, which is calculated by summing these gene values (0 or 1), weighted by the fitness value of corresponding individual, with the same position traversing every individual in population. As a key factor, feature importance has a substantial influence on the algorithmic performance. Taking into account the population consisting of the feature subsets in prediction errors level 1 and 2 given by

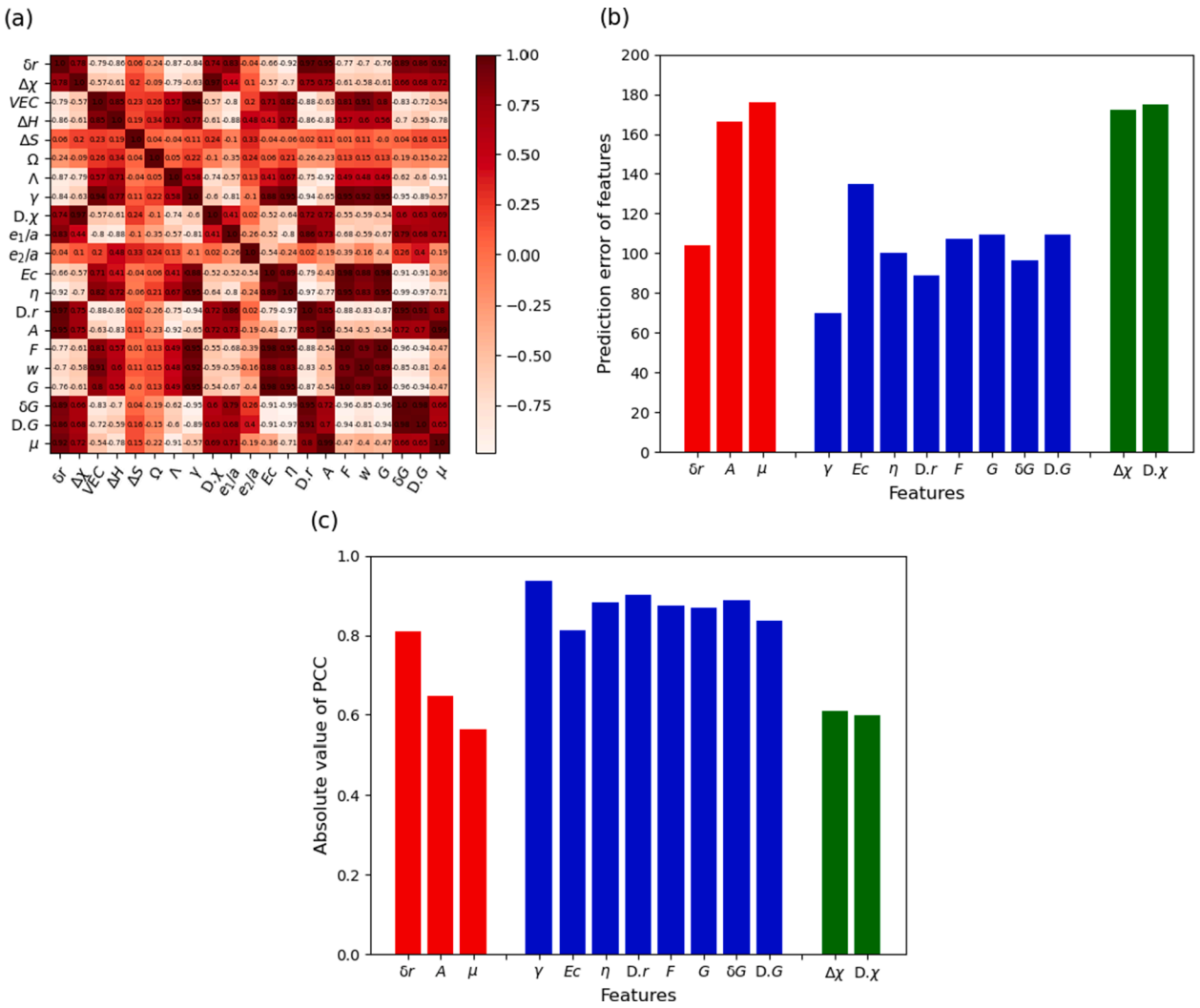


Fig. 7. The results of PCC method. (a) The PCC values of different feature pairs; (b) the prediction errors calculated by different single features; (c) the PCC absolute values between different single features and hardness values. In the figures (a) and (b), each color represents a group of features with strong correlation.

exhaustive method, the feature importances are calculated, which are shown in Fig. 8. It is incicated that both the *VEC* and ΔS have the highest feature importance obviously. In addition, these two features are included in all of the feature combinations in level 1 and 2. Thus, for HEAs hadeness prediction, they are significant features determined from the present definition of the feature importance.

The valence electron concentration *VEC* is defined by [58–60]:

$$VEC = \sum_{j=1}^{N} C_j VEC_j \tag{1}$$

where N is the total number of alloying element kinds, C_j and VEC_j are the atomic percentage and the valence electron number of the jth element, respectively. According to *Hume-Rothery* rule, the valencies of elements could affect the phase of alloys. Furthermore, there are a lot of works demonstrating that the *VEC* has a substantial influence on the phase formation of HEAs [58,61–65]. Guo et al used *VEC* to separate the compositions for BCC and FCC phases in as-cast $Al_xCoCrCuFeNi$ and $Al_xCoCrFeNi_2$ alloys ($0 \leqslant x \leqslant 2$) [61]. M. H. Tsai et al applied the *VEC* to predict whether the σ phase forms in annealed alloys which contains Cr and Fe along with Al, Co, Mn, Ni, Ti and/or V [66]. So *VEC* may influence the phase formation and thus affect the hardness or other properties of HEAs indirectly.

The configuration entropy ΔS is defined as follows [58–60]:

$$\Delta S = -R\sum_{j=1}^{N} C_j \ln C_j \tag{2}$$

where R is the gas constant. Configuration entropy is an important parameter which could be used to define HEAs. Some studies suggest that HEAs are the alloys with $\Delta S > 1.61R$ [58,61]. According to the high entropy effect, with the increasing of ΔS, HEAs may favor solid solution phases over competing intermetallic compound [58]. Thus the configuration entropy may also influence the the hardness of HEAs by affecting the phase formation.

3.5. The validation of ML model selection

In reality, for performance prediction of materials by ML models, a specific ML model should be selected in advance. In the present work (Section 3.2), SVR-R is determined by composition features. In Section 3.3, by adopting SVR-R combined with the improved GA-1 as feature selectin method, it is demonstrated that the GA-1 proposed is remarkably superior to the traditional GA-0 in all of these aspects of the accuracy, stability and efficiency. In this section, the prediction errors are also calculated by other ML models with GA-1 and the results are shown in Fig. 9. These ML models include LR, PR, LASSO, RR, SVR-L and BPNN. It could be seen that SVR-R is still with the relatively lowest prediction

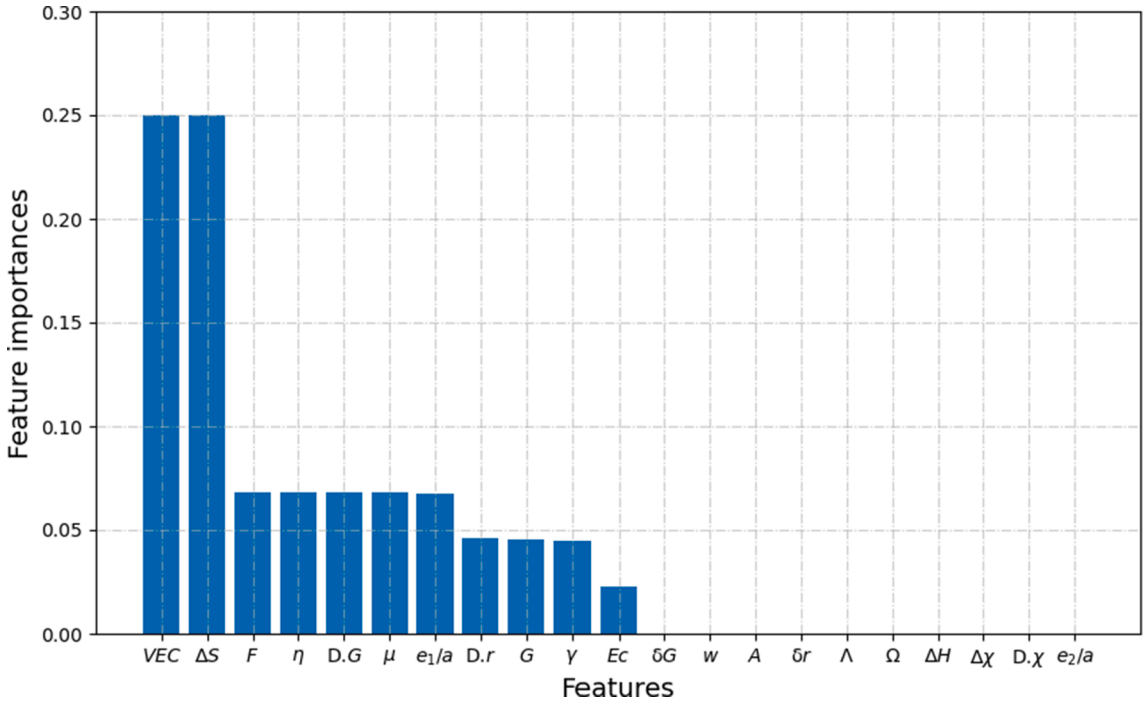


Fig. 8. The feature importance value of each feature, this results are calculated by adopting the population consisting of these feature subsets in prediction errors level 1 and 2.

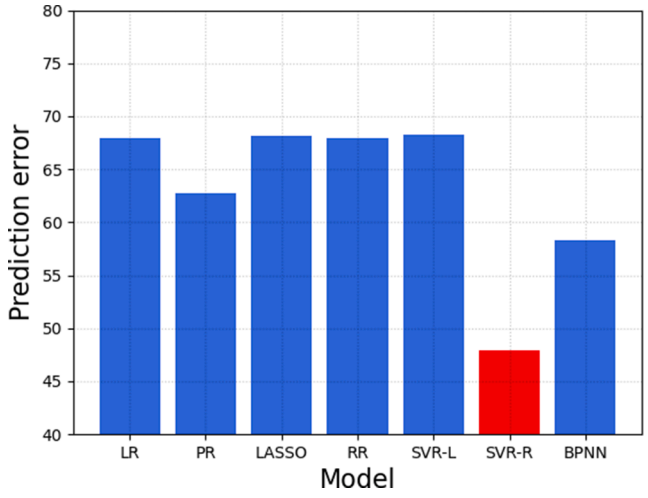


Fig. 9. The prediction errors of different ML models based on the optimal physical feature subset selected by GA-1, SVR-R has the lowest prediction error obviously.

error obviously. Furthermore, the prediction errors of all of these ML models shown in Fig. 9 with the optimal physical feature subsets are lower than that calculated by the composition features. Therefore, the GA-1 proposed by the present work is a feature selection method with remarkable general applicability.

3.6. The HEAs hardness prediction.

Applying to Al-Co-Cr-Cu-Fe-Ni HEAs, the testing results for the SVR-R model combined with improved GA-1 are given in Fig. 10. The optimal physical feature subset (*VEC*, ΔS, γ, e_1/a) selected by GA-1 is adopted. The data is divided into training and testing sets randomly and in Fig. 10 (a) the averaged prediction errors are calculated by 50 times for different split ratios. In Fig. 10 (b), the comparison between the hardness prediction values and the hardness measurement values is made, which is obtained by ten-fold cross validation. It is indicated that the relatively good prediction performance is achieved.

In order to further elevate the prediction ability of ML model, the stacking method as an ensemble learning strategy is proposed and illustrated in Fig. 11. First of all, based on the training data defined by composition or optimal physical feature subset, SVR-R model is trained, which is marked by model *C* or model *P*, respectively. Then adopting these two models, hardness values of training data are independently predicted. The two hardness prediction values are supplied into the optimal physical feature subset to produce the new feature combination including 6 features. By the refreshed training data, the final SVR-R model is constructed, which is labeled as [*P*, *P*-p, *C*-p] in Fig. 12. As a comparison, other two stacking models are also taken into account, by integrating the optimal physical feature subset and the hardness value as one of features predicted from the model *C* or model *P*. They are labeled as [*P*, *C*-p] and [*P*, *P*-p], respectively. In Fig. 12, the prediction errors calculated by model *P*, model *C* as well as the three ensemble learning models with different stacking types mentioned above are shown. It could be seen that the prediction errors have been further lowered by stacking method. Especially, compared with other models the model with stacking type [*P*, *P*-p, *C*-p] is the relatively most powerful one.

4. Conclusions

In this work, the improved genetic algorithm marked by GA-1 has been developed to serve as feature selection method for ML model construction. Applying to Al-Co-Cr-Cu-Fe-Ni HEAs for hardness prediction, a comparison with the traditional GA and other typical feature selection methods was made thoroughly. The main conclusions are summarized as follows:

1. It is demonstrated that the improved GA-1 is superior to the traditional GA in all of these aspects of accuracy, stability and efficiency obviously.
2. The reasonability for GA-1 was further clarified by introducing the concepts of feature importance and gene manipulation. In addition,

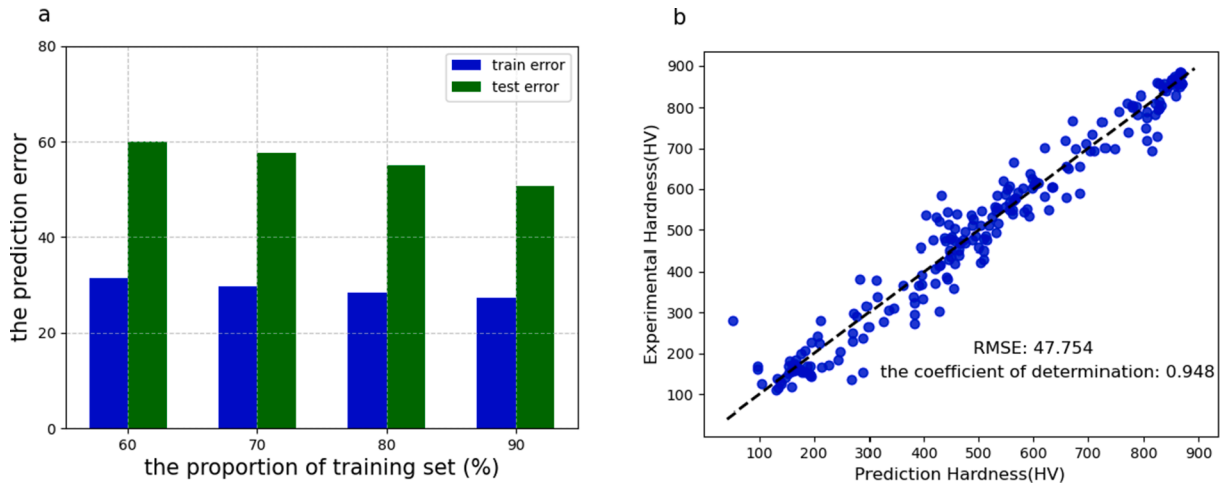


Fig. 10. The results of the SVR-R model with the optimal feature subset for HEAs hardness prediction. (a) the mean training errors and testing errors for different data split ratios, (b) the relationship between hardness prediction values and hardness measurement values obtained by ten-fold cross validation.

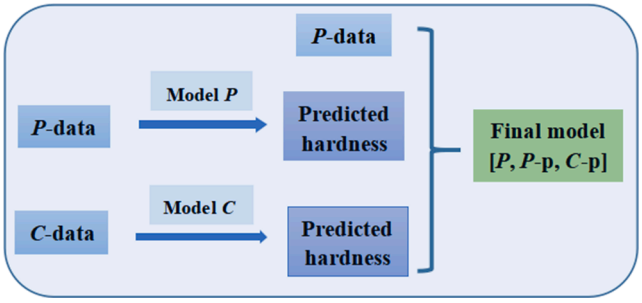


Fig. 11. The flow chart of stacking method. *P*-data represents the data defined by optimal physical feature subset and *C*-data is the data given by composition feature. The sign [*P*, *P*-p, *C*-p] represents the stacking type which is trained in feature combination including the optimal physical feature subset and two sets of hardness prediction values, marked by *P*-p and *C*-p respectively, predicted from the model *P* and model *C*.

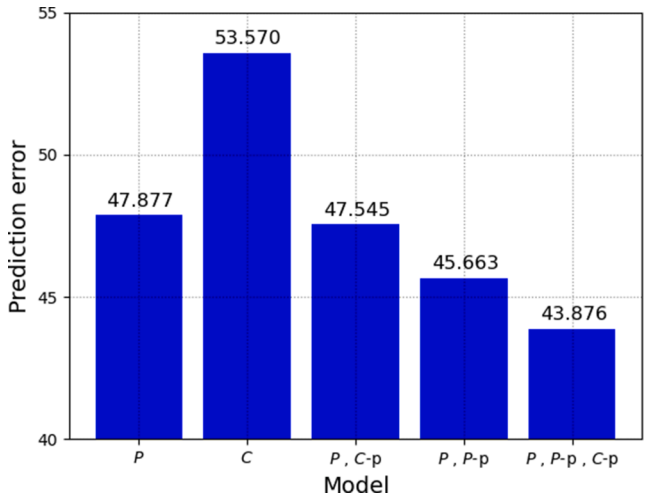


Fig. 12. The prediction errors of the model *P*, *C* and different stacking types, these results are calculated by ten-fold cross validation.

the ML model selection method was also validated with both the composition feature and the optimal physical feature combination selected by the improved GA.

3. The stacking method as an ensemble learning strategy is proposed in HEAs hardness prediction, to further elevate the prediction ability of ML model. By this method, prediction performance is successfully enhanced.

CRediT authorship contribution statement

Shuai Li: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Shu Li:** Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Writing – original draft, Writing – review & editing. **Dongrong Liu:** Conceptualization, Investigation, Methodology, Software, Supervision, Validation, Project administration, Supervision, Writing – review & editing. **Rui Zou:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft. **Zhiyuan Yang:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft.

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.

Acknowledgements

The authors acknowledge the support of National Natural Science Foundation of China (Nos. 51671075 and 51971086), Fundamental Research Foundation for Universities of Heilongjiang Province (No. LGYC2018JC004), project funded by China Postdoctoral Science Foundation (No. 2016M590970), fund of the State Key Laboratory of Solidification Processing in NWPU (No. SKLSP201606) and Heilongjiang Postdoctoral Fund for scientific research initiation (No. LBH-Q16118).

Data availability

The raw data required and processed data to reproduce these findings are available to download from supplement materials.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.commatsci.2022.111185.

References

[1] K.T. Butler, D.W. Davies, H. Cartwright, O. Isayev, A. Walsh, Machine learning for molecular and materials science, Nature. 559 (7715) (2018) 547–555, https://doi.org/10.1038/s41586-018-0337-2.

[2] L. Himanen, A. Geurts, A.S. Foster, P. Rinke, Data-driven materials science: status, challenges and perspectives, Adv. Sci. 6 (21) (2019) 1900808, https://doi.org/10.1002/advs.201900808.

[3] Z.G. Zhu, K.H. Ma, Q. Wang, C.H. Shek, Compositional dependence of phase formation and mechanical properties in three CoCrFeNi-(Mn/Al/Cu) high entropy alloys, Intermetallics. 79 (2016) 1–11, https://doi.org/10.1016/j.intermet.2016.09.003.

[4] W.R. Wang, W.L. Wang, S.C. Wang, Y.C. Tsai, C.H. Lai, J.W. Yeh, Effects of Al addition on the microstructure and mechanical property of AlxCoCrFeNi high-entropy alloys, Intermetallics. 26 (2012) 44–51, https://doi.org/10.1016/j.intermet.2012.03.005.

[5] Y.-F. Kao, T.-J. Chen, S.-K. Chen, J.-W. Yeh, Microstructure and mechanical property of as-cast, -homogenized, and -deformed AlxCoCrFeNi ($0 \le x \le 2$) high-entropy alloys, J. Alloys Compd. 488 (1) (2009) 57–64, https://doi.org/10.1016/j.jallcom.2009.08.090.

[6] J. Schmidt, M.R.G. Marques, S. Botti, M.A.L. Marques, Recent advances and applications of machine learning in solid-state materials science, npj Comput Mater. 5 (2019) 83, https://doi.org/10.1038/s41524-019-0221-0.

[7] C. Zhang, F. Zhang, S. Chen, W. Cao, Computational Thermodynamics Aided High-Entropy Alloy Design, JOM. 64 (7) (2012) 839–845, https://doi.org/10.1007/s11837-012-0365-6.

[8] T. Kostiuchenko, F. Körmann, J. Neugebauer, A. Shapeev, Impact of lattice relaxations on phase transitions in a high-entropy alloy studied by machine-learning potentials, npj Comput Mater. 5 (2019) 55, https://doi.org/10.1038/s41524-019-0195-y.

[9] P. Hohenberg, W. Kohn, Inhomogeneous electron gas, Phys. Rev. 136 (1964) B864–B871.

[10] W. Kohn, L.J. Sham, Self-consistent equations including exchange and correlation effects, Phys. Rev. 140 (1965) A1133–A1138.

[11] A. Badillo, C. Beckermann, Phase-field simulation of the columnar-to-equiaxed transition in alloy solidification, Acta Mater. 54 (8) (2006) 2015–2026, https://doi.org/10.1016/j.actamat.2005.12.025.

[12] R. Tönhardt, G. Amberg, Phase-field simulation of dendritic growth in a shear flow, J. Cryst. Growth. 194 (3-4) (1998) 406–425, https://doi.org/10.1016/S0022-0248(98)00687-3.

[13] K.F. Garrity, K. Choudhary, Database of Wannier tight-binding Hamiltonians using high-throughput density functional theory, Sci Data. 8 (2021) 1–10, https://doi.org/10.1038/s41597-021-00885-z.

[14] X.N. Mao, L. Wang, Y.F. Xu, P.J. Wang, Y.Y. Li, J.J. Zhao, Computational high-throughput screening of alloy nanoclusters for electrocatalytic hydrogen evolution, npj Comput Mater. 7 (2021) 46, https://doi.org/10.1038/s41524-021-00514-8.

[15] H. Ma, Y. Jiao, W. Guo, X. Liu, Y. Li, X.-D. Wen, Predicting Crystal Morphology Using a Geometric Descriptor: A Comparative Study of Elemental Crystals with High-Throughput DFT Calculations, J. Phys. Chem. C. 124 (29) (2020) 15920–15927, https://doi.org/10.1021/acs.jpcc.0c0353710.1021/acs.jpcc.0c03537.s001.

[16] A. Jain, G. Hautier, C.J. Moore, S. Ping Ong, C.C. Fischer, T. Mueller, K.A. Persson, G. Ceder, A high-throughput infrastructure for density functional theory calculations, Comp Mater Sci. 50 (8) (2011) 2295–2310, https://doi.org/10.1016/j.commatsci.2011.02.023.

[17] L. Qiao, Y. Liu, J. Zhu, A focused review on machine learning aided high-throughput methods in high entropy alloy, J Alloy. Compd. 877 (2021) 160295, https://doi.org/10.1016/j.jallcom.2021.160295.

[18] C. Wen, Y. Zhang, C.X. Wang, D.Z. Xue, Y. Bai, S. Antonov, L.H. Dai, T. Lookman, Y.J. Su, Machine learning assisted design of high entropy alloys with desired property, Acta Mater. 170 (2019) 109–117, https://doi.org/10.1016/j.actamat.2019.03.010.

[19] S.K. Dewangan, S. Samal, V. Kumar, Microstructure exploration and an artificial neural network approach for hardness prediction in AlCrFeMnNiWx High-Entropy Alloys, J. Alloys Compd. 823 (2020) 153766, https://doi.org/10.1016/j.jallcom.2020.153766.

[20] M. Tatlier, Artificial neural network methods for the prediction of framework crystal structures of zeolites from XRD data, Neural Comput. Appl. 20 (3) (2011) 365–371, https://doi.org/10.1007/s00521-010-0386-4.

[21] A.O. Oliynyk, L.A. Adutwum, J.J. Harynuk, A. Mar, Classifying crystal structures of binary compounds AB through cluster resolution feature selection and support vector machine analysis, Chem. Mater. 28 (18) (2016) 6672–6681, https://doi.org/10.1021/acs.chemmater.6b0290510.1021/acs.chemmater.6b02905.s00110.1021/acs.chemmater.6b02905.s002.

[22] C.S. Wang, H.D. Fu, L. Jiang, D.Z. Xue, J.X. Xie, A property-oriented design strategy for high performance copper alloys via machine learning, npj Comput Mater. 5 (2019) 87, https://doi.org/10.1038/s41524-019-0227-7.

[23] P. Liu, H.Y. Huang, S. Antonov, C. Wen, D.Z. Xue, H.W. Chen, L.F. Li, Q. Feng, T. Omori, Y.J. Su, Machine learning assisted design of γ′-strengthened Co-base superalloys with multi-performance optimization, npj Comput Mater. 6 (2020) 62, https://doi.org/10.1038/s41524-020-0334-5.

[24] X.B. Hu, J.C. Wang, Y.Y. Wang, J.J. Li, Z.J. Wang, Y.Y. Dang, Y.F. Gu, Two-way design of alloys for advanced ultra supercritical plants based on machine learning, Comp Mater Sci. 155 (2018) 331–339, https://doi.org/10.1016/j.commatsci.2018.09.003.

[25] C. Wen, C. Wang, Y. Zhang, S. Antonov, D. Xue, T. Lookman, Y. Su, Modeling solid solution strengthening in high entropy alloys using machine learning, Acta Mater. 212 (2021) 116917, https://doi.org/10.1016/j.actamat.2021.116917.

[26] Y.H. Wang, Y.F. Tian, T. Kirk, O. Laris, J.H. Ross Jr, R.D. Noebe, V. Keylin, R. Arróyavead, Accelerated design of Fe-based soft magnetic materials using machine learning and stochastic optimization, Acta Mater. 194 (2020) 144–155, https://doi.org/10.1016/j.actamat.2020.05.006.

[27] X.D. Liu, X. Li, Q.F. He, D.D. Liang, Z.Q. Zhou, J. Ma, Y. Yang, J. Shen, Machine Learning-based Glass Formation Prediction in Multicomponent Alloys, Acta Mater. 201 (2020) 182–190, https://doi.org/10.1016/j.actamat.2020.09.081.

[28] A. Rahnama, S. Clark, S. Sridhar, Machine learning for predicting occurrence of interphase precipitation in HSLA steels, Comp Mater Sci. 154 (2018) 169–177, https://doi.org/10.1016/j.commatsci.2018.07.055.

[29] D. Dai, T. Xu, X. Wei, G. Ding, Y. Xu, J. Zhang, H. Zhang, Using machine learning and feature engineering to characterize limited material datasets of high-entropy alloys, Comp Mater Sci. 175 (2020) 109618, https://doi.org/10.1016/j.commatsci.2020.109618.

[30] H.T. Zhang, H.D. Fu, X.Q. He, C.S. Wang, L. Jiang, L.Q. Chen, J.X. Xie, Dramatically Enhanced Combination of Ultimate Tensile Strength and Electric Conductivity of Alloys via Machine Learning Screening, Acta Mater. 200 (2020) 803–810, https://doi.org/10.1016/j.actamat.2020.09.068.

[31] R. Machaka, Machine learning-based prediction of phases in high-entropy alloys, Comp Mater Sci. 188 (2021) 110244, https://doi.org/10.1016/j.commatsci.2020.110244.

[32] J.A. Pérez-Benitez, L.R. Padovese, Feature selection and neural network for analysis of microstructural changes in magnetic materials, Expert Syst. Appl. 38 (8) (2011) 10547–10553, https://doi.org/10.1016/j.eswa.2011.02.088.

[33] L.M. Ghiringhelli, J. Vybiral, S.V. Levchenko, C. Draxl, M. Scheffler, Big data of materials science: critical role of the descriptor, Phys. Rev. Lett. 114 (2015), 105503.

[34] M. Chiesa, G. Maioli, G.I. Colombo, L. Piacentini, GARS: Genetic Algorithm for the identification of a Robust Subset of features in high-dimensional datasets, BMC Bioinformatics 21 (1) (2020), https://doi.org/10.1186/s12859-020-3400-6.

[35] A. Chowdhury, E. Kautz, B. Yener, D. Lewis, Image driven machine learning methods for microstructure recognition, Comp Mater Sci. 123 (2016) 176–187, https://doi.org/10.1016/j.commatsci.2016.05.034.

[36] S.R. Broderick, J.R. Nowers, B. Narasimhan, K. Rajan, Tracking chemical processing pathways in combinatorial polymer libraries via data mining, J. Combina tor. Chem. 12 (2) (2010) 270–277, https://doi.org/10.1021/cc900145d.

[37] Y. Zhang, C. Wen, C.X. Wang, S. Antonov, D.Z. Xue, Y. Bai, Y.J. Su, Phase prediction in high entropy alloys with a rational selection of materials descriptors and machine learning models, Acta Mater. 185 (2020) 528–539, https://doi.org/10.1016/j.actamat.2019.11.067.

[38] D.E. Goldberg, Genetic Algorithms in Search, Optimization and Machine Learning, Addison-Wesley Professional, Boston, 1989.

[39] A.M. Maia, Y. Ghamri-Doudane, D. Vieira, M. Franklin de Castro, An improved multi-objective genetic algorithm with heuristic initialization for service placement and load distribution in edge computing, Comput Netw. 194 (2021) 108146, https://doi.org/10.1016/j.comnet.2021.108146.

[40] D.H. Kim, A. Abraham, J.H. Cho, A hybrid genetic algorithm and bacterial foraging approach for global optimization, Inform Sciences. 177 (18) (2007) 3918–3937, https://doi.org/10.1016/j.ins.2007.04.002.

[41] A.G. Bakirtzis, P.N. Biskas, C.E. Zoumas, V. Petridis, Optimal Power Flow by Enhanced Genetic Algorithm, IEEE T Power Syst. 17 (2002) 229–236, https://doi.org/10.1109/MPER.2002.4311997.

[42] U. Aickelin, K.A. Dowsland, An Indirect Genetic Algorithm for a Nurse Scheduling Problem, Comput Oper Res. 31 (5) (2004) 761–778, https://doi.org/10.1016/S0305-0548(03)00034-0.

[43] A. Khazaee, A. Ebrahimzadeh, Classification of electrocardiogram signals with support vector machines and genetic algorithms using power spectral features, Biomed Signal Process Control. 5 (4) (2010) 252–263, https://doi.org/10.1016/j.bspc.2010.07.006.

[44] L. Scrucca, GA: A Package for Genetic Algorithms in R, Journal of Statistical Software, Foundation for Open Access Statistics, J Stat Softw. 53 (2013) 1–37.

[45] M.S. Mohamad, SAFAAI Deris, R.MD. Illias, A hybrid of genetic algorithm and support vector machine for features selection and classification of gene expression microarray, Int. J. Comput. Int. Sys. 05 (01) (2005) 91–107, https://doi.org/10.1142/S1469026805001465.

[46] M. Kuhn, Building predictive models in R using the caret package, J. Stat. Softw. 28 (2008) 1–26.

[47] K.M. Youssef, A.J. Zaddach, C. Niu, D.L. Irving, C.C. Koch, A novel low density, high hardness, high-entropy alloy with close-packed single-phase nanocrystalline structures, Mater. Res. Lett. 3 (2) (2015) 95–99, https://doi.org/10.1080/21663831.2014.985855.

[48] Y. Deng, C.C. Tasan, K.G. Pradeep, H. Springer, A. Kostka, D. Raabe, Design of a twinning-induced plasticity high entropy alloy, Acta Mater. 94 (2015) 124–133, https://doi.org/10.1016/j.actamat.2015.04.014.

[49] Y.Y. Chen, U.T. Hong, H.C. Shih, J.W. Yeh, T. Duval, Electrochemical kinetics of the high entropy alloys in aqueous environments - a comparison with type 304

stainless steel, Corros. Sci. 47 (11) (2005) 2679–2699, https://doi.org/10.1016/j.corsci.2004.09.026.
[50] Y. Zhang, T.T. Zuo, Y.Q. Cheng, P.K. Liaw, High-entropy alloys with high saturation magnetization, electrical resistivity, and malleability, Sci. Rep. 3 (2013) 1–7, https://doi.org/10.1038/srep01455.
[51] B. Gludovatz, A. Hohenwarter, D. Catoor, E.H. Chang, E.P. George, R.O. Ritchie, A fracture-resistant high-entropy alloy for cryogenic applications, Science 345 (2014) 1153–1158, https://doi.org/10.1002/chin.201447007.
[52] Q. Wu, Z. Wang, X. Hu, T. Zheng, Z. Yang, F. He, J. Li, J. Wang, Uncovering the eutectics design by machine learning in the Al–Co–Cr–Fe–Ni high entropy system, Acta Mater. 182 (2020) 278–286, https://doi.org/10.1016/j.actamat.2019.10.043.
[53] K. Kaufmann, K.S. Vecchio, Searching for high entropy alloys: A machine learning approach, Acta Mater. 198 (2020) 178–222, https://doi.org/10.1016/j.actamat.2020.07.065.
[54] L. Qiao, Z. Lai, Y. Liu, A. Bao, J. Zhu, Modelling and prediction of hardness in multi-component alloys: A combined machine learning, first principles and experimental study, J. Alloy Compd. 853 (2021) 156959, https://doi.org/10.1016/j.jallcom.2020.156959.
[55] W.J. Huang, P. Martin, H.L.L. Zhuang, Machine-learning phase prediction of high-entropy alloys, Acta Mater. 169 (2019) 225–236, https://doi.org/10.1016/j.actamat.2019.03.012.
[56] N. Islam, W.J. Huang, H.L.L. Zhuang, Machine learning for phase selection in multi-principal element alloys, Comp Mater Sci. 150 (2018) 230–235, https://doi.org/10.1016/j.commatsci.2018.04.003.
[57] C.K.H. Borg, C. Frey, J. Moh, T.M. Pollock, S. Gorsse, D.B. Miracle, O.N. Senkov, B. Meredig, J.E. Saal, Expanded dataset of mechanical properties and observed phases of multi-principal element alloys, Sci Data. 7 (2020) 430, https://doi.org/10.1038/s41597-020-00768-9.
[58] D.B. Miracle, O.N. Senkov, A critical review of high entropy alloys and related concepts, Acta Mater. 122 (2017) 448–511, https://doi.org/10.1016/j.actamat.2016.08.081.
[59] J.H. Li, M.H. Tsai, Theories for predicting simple solid solution high-entropy alloys: Classification, accuracy, and important factors impacting accuracy, Scr. Mater. 188 (2020) 80–87, https://doi.org/10.1016/j.scriptamat.2020.06.064.
[60] S. Guo, Phase selection rules for cast high entropy alloys: an overview, Mater. Sci. Technol. 31 (10) (2015) 1223–1230, https://doi.org/10.1179/1743284715Y.0000000018.
[61] S. Guo, C. Ng, J. Lu, C.T. Liu, Effect of valence electron concentration on stability of fcc or bcc phase in high entropy alloys, J. Appl. Phys. 109 (10) (2011) 103505, https://doi.org/10.1063/1.3587228.
[62] S. Guo, C.T. Liu, Phase stability in high entropy alloys: Formation of solid solution phase or amorphous phase, Prog. Nat. Sci. 21 (6) (2011) 433–446, https://doi.org/10.1016/S1002-0071(12)60080-X.
[63] S. Guo, C.T. Liu, Phase selection rules for complex multi-component alloys with equiatomic or close-to-equiatomic compositions, Chin. J. Nat. 35 (2013) 85–96.
[64] X. Yang, Y. Zhang, Prediction of high-entropy stabilized solid-solution in multi-component alloys, Mater. Chem. Phys. 132 (2-3) (2012) 233–238, https://doi.org/10.1016/j.matchemphys.2011.11.021.
[65] Y. Zhang, Y.J. Zhou, J.P. Lin, G.L. Chen, P.K. Liaw, Solid-solution phase fomation rules for multi-component alloys, Adv. Eng. Mater. 10 (2008) 534–538, https://doi.org/10.1002/adem.200700240.
[66] M.-H. Tsai, K.-Y. Tsai, C.-W. Tsai, C. Lee, C.-C. Juan, J.-W. Yeh, Criterion for sigma phase formation in Cr- and V-containing high-entropy alloys, Mater. Res. Lett. 1 (4) (2013) 207–212, https://doi.org/10.1080/21663831.2013.831382.