International Journal of Engineering Innovation & Research
Volume 10, Issue 2, ISSN: 2277 – 5668

Reliability and Hardness Prediction of High-Entropy Alloy Based on Machine Learning

Qian Lin [1*] and Haifeng Wu [2]
[1] Qinghai University for Nationalities, Xining, Xining, China.
[2] Chengdu Ganide Technology, Chengdu, China.
*Corresponding author email id: linqian@tju.edu.cn

Date of publication (dd/mm/yyyy): 22/04/2021

Abstract **– This article provides a method for using machine learning to predict the reliability and hardness of Al-Co-Cr-Cu-Fe-Ni system high-entropy alloys, and searches for several high-hardness and high-entropy alloys. This paper uses the most representative machine learning feature screening method to perform feature screening on 20 physical features that are highly related to the hardness and reliability prediction of Al-Co-Cr-Cu-Fe-Ni system high-entropy alloys to find the optimal feature combination and the time-consuming and accuracy of each method was evaluated and compared horizontally. In the end, it is found that genetic algorithm has the best effect on feature selection. Based on support vector regression, a highly accurate high-entropy alloy hardness and reliability prediction model is trained using the selected feature combination, and the prediction accuracy can reach 50.468672. Based on this model, this paper conducts corresponding hardness prediction and alloy search in the vast Al-Co-Cr-Cu-Fe-Ni system high-entropy alloy composition space, and successfully searches for several high-hardness high-entropy alloys with the highest specific composition. The predicted hardness value is 791.5322HV. It means high reliability for application.**

Keywords **– Reliability Prediction, Hardness, Machine Learning, High-Entropy Alloy, Support Vector Regression.**

I. INTRODUCTION

In recent years, with the advent of the era of artificial intelligence and big data, data-driven science based on machine learning big data has become the fourth material search science paradigm after the three eras of relay experiment, theory, and calculation in the field of scientific discovery, and has received extensive attention and recognition in the field of materials science [1].

Using machine learning to predict material performance often requires the following steps: data collection, feature engineering, model selection and training, and performance prediction. XUE et al [2] used an exhaustive method for feature selection when studying the hardness and reliability of high-entropy alloys, but it was too time-consuming and laborious. Zhang et al. used genetic algorithm to search for the best feature combination among 75 features [3]. Another machine learning paper published by Zhang et al using Pearson correlation coefficient, univariate feature selection and feature stability for feature selection [4]. However, the studies so far have not been able to provide a horizontal comparison of the comprehensive application effects of these representative feature selection methods, nor have they proposed how to use convenient and effective methods to find effective methods for the prediction of high-entropy alloy hardness feature combination.

HEAs have performing properties [5-6], and there are many studies can help to study it [7-8], an important thing about the study of HEAs is to find the best features combination. This paper selects the most representative feature selection methods: Pearson correlation coefficient, univariate feature selection, stability selection, sequence forward selection method, sequence backward selection method and genetic algorithm, and screen them separately the results are compared and analyzed. First, the Pearson correlation coefficient can be calculated separately for each pair of features in all the features to evaluate the linear correlation between the

features, and the features with relatively poor predictive performance can be filtered out of the several features with high correlation. After that, several feature screening methods were used to screen out the optimal feature combination, and the time consumption of each method was compared with the feature combination selected by each method. The error in the hardness prediction of high-entropy alloys is used to comprehensively analyze the advantages and disadvantages of each algorithm, to obtain the most excellent feature selection method and the feature combination with the best prediction performance, and bring it into the pre-selected machine learning model to train a model of alloy reliability prediction, the reliability prediction of then alloy model is finally applied to the vast virtual composition space of high-entropy alloys for hardness prediction, and the specific high-entropy alloy composition with higher hardness and reliability is searched out.

II. METHOD

A. *Establish Feature Data Set and Machine Learning Model Selection*

The data set used in this article comes from the 155 component data set given in the literature [2], including 1 ternary alloy, 22 quaternary alloys, 95 pentad alloys, and 38 six-element alloys.

According to the literature [2], the support vector regression of Gaussian kernel has better prediction accuracy and stability. When only the molar ratio of each component in the alloy is used as the eigenvector of the sample. The error of the ten-fold cross-validation is shown in Table 1:

Table 1. Error values of ten-fold cross-validation using components as features.

1	2	3	4	5	6	7	8	9	10
55.14408	55.50087	55.86885	55.34160	53.50900	56.56432	56.39552	56.54928	56.55817	53.72977

The average of the error of ten times of ten-fold cross-validation is 55.516150.

B. *Feature Selection*

It can be seen from the literature [2] that in the process of using physical characteristics to predict the reliability and hardness of high-entropy alloys, the combination of the three characteristics can be considered. First of all, for all feature data sets, this article randomly selects 200 features combinations containing different numbers of features, and then selects the corresponding feature data for training and uses ten-fold cross-validation to verify the accuracy. The accuracy results are as shown in Fig. 1.

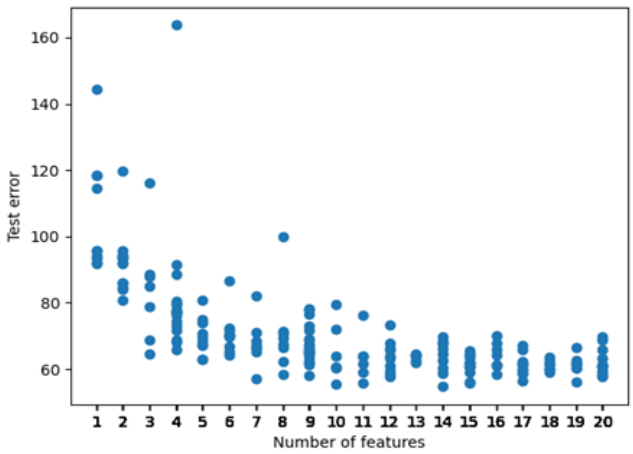

Fig. 1. 200 features combinations containing verify accuracy.

It can be seen from the figure that the feature combination containing the three features can guarantee a certain accuracy value, but also can ensure that the number of feature combinations is small [9].

III. RESULTS AND ANALYSIS

A. *Feature Set of Exhaustive Method*

This paper selects the feature set containing only three features based on the idea in literature [2], the top 10 feature combinations with prediction accuracy are listed in the table 2:

Table 2. The top 10 feature combinations with prediction accuracy searched by exhaustive method.

Rank	Features	Error
1	γ parameter, electron concentration, work function	50.468672
2	Valence electron concentration, work function, shear modulus	53.240014
3	Electron concentration, local size mismatch, work function	53.564123
4	Valence electron concentration, cohesive energy, work function	54.393227
5	Valence electron concentration, Nabarro coefficient, work function	55.109845
6	Atomic radius difference, electron concentration, work function	55.247944
7	γ parameter, work function, shear modulus	55.875620
8	γ parameter, Nabarro coefficient, work function	55.988801
9	Valence electron concentration, configuration entropy, cohesive energy	56.430437
10	Valence electron concentration, γ parameter, work function	57.362392

B. *Pearson's Correlation Coefficient, Univariate Feature Selection and Stability Selection*

If the absolute value of Pearson's correlation coefficient is greater than 0.95, these two features are considered highly correlated [2]. Therefore, this paper will remove the highly correlated features (energy term ($\boldsymbol{A}$), lattice distortion energy ($\boldsymbol{\mu}$), shear modulus ($\boldsymbol{G}$), shear modulus difference ($\boldsymbol{\delta G}$), electronegativity difference ($\boldsymbol{\Delta\chi}$)).

However, the Pearson correlation coefficient describes the linear dependence relationship between variables. If there is a non-linear relationship between the variables, the Pearson correlation coefficient results are very poor. Therefore, further use of univariate feature selection and stability selection is required. The evaluation is carried out and the result is shown in Fig. 2.

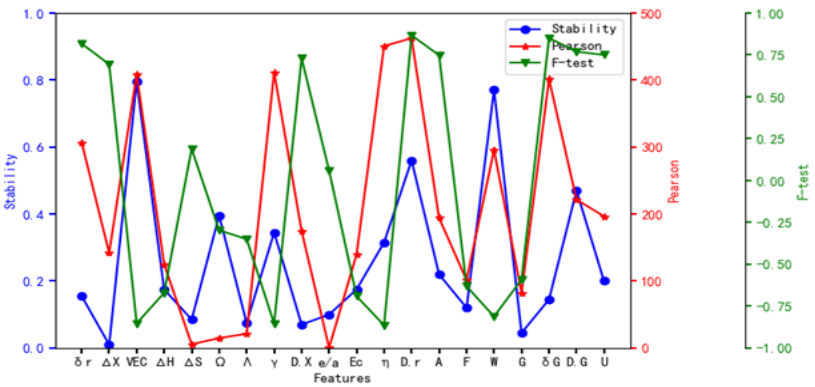

Fig. 2. Pearson correlation coefficient, univariate feature selection, stability selection.

From the Fig. 2 above, the three characteristics selected by the Pearson correlation coefficient that are most relevant to the hardness value are: local size mismatch (Dr), modulus mismatch ($I\!I$), Υ parameter (Υ). This feature combination is substituted into the model and the test error is 83.20397. The three features selected by the F-test method that are most relevant to the hardness value are: local size mismatch (Dr), shear modulus difference (δG), atomic radius difference (δr). This feature combination is substituted into the model and the test error is 81.13132. The three corresponding features that are most important in the reliability and hardness prediction problem selected by stability are: valence electron concentration (VEC), work function (W), and local size mismatch ($D.r$). The corresponding error is 61.88044. The feature combinations selected by these three feature screening methods are lower in the ranking of the listed feature combination accuracy values.

a. *Forward/Backward Sequence Selection*

Use Sequence Forward Selection (SFS) among 15 features, use ten-fold cross-validation to calculate the error, and use root mean square error (RMSE) for error analysis, and finally the selected feature combination containing three features is: atomic radius difference, electron concentration, work function. The error corresponding to this feature combination is the lowest, reaching 55.24794, ranking sixth in the overall ranking of the three features.

Use sequence backward selection in the same way finally the selected feature combination containing the three features is: valence electron concentration, nabarro coefficient, work function. The error of this feature combination is the lowest, reaching 55.10984, ranking fifth in the overall ranking of the three features.

By comparing, the error difference between the two is 0.1381, the error is relatively close. The ranking order in the overall ranking is in sequence, and the ranking is relatively high, all among the top 6, and we also tried to perform SFS and SBS feature screening methods in 20 features, and found that the results obtained are the same as those obtained in 15 features. Therefore, we believe that SFS and SBS can save time and will not affect the prediction results after using Pearson's correlation coefficient to screen out relevant features. But this method is a heuristic search, the previous screening result too much affects the next screening result, and it is difficult to consider the multiple possible situations of feature combination. They are greedy algorithm, so it is easy to fall into a local optimum.

b. *Genetic Algorithm*

The genetic algorithm in this paper determines that the population number is 100 and the number of iterations is 200, but a stable result can be obtained after dozens of iterations of the population. The mutation probability is set to 0.01, and the genetic algorithm is executed multiple times to obtain different iterative results, but each screening result is in the top 6 of the feature ranking, and the first place is also found. Regardless of whether it is selected among 15 features or selected among 20 features, the results of each are within the top six, and in each case, the optimal feature combination has been found, and it takes less time. This shows that the genetic algorithm can save the time of feature selection while ensuring accuracy.

c. *Summary of Feature Screening Methods*

This article introduces several representative feature screening methods. The characteristics and results of various feature screening methods are shown in the following table:

Table 3. Comparison of advantages and disadvantages of each feature selection method.

	Filter Feature Combinations (Best Results)	Screening Test Error of Feature Combination	Feature Selection Time-Consuming	Advantage	Disadvantage
Pearson correlation coefficient	Local size mismatch, modulus mismatch, γ parameter	83.20397	less	Less time, easy to understand and implement	The accuracy is not high. Not matched with the model
Univariate feature selection	Local size mismatch, difference in shear modulus, difference in atomic radius	81.13132	less	Less time, easy to understand and implement	The accuracy is not high. Not matched with the model
Stability selection	Valence electron concentration, work function, local size mismatch	61.88044	less	Less time, easy to understand and implement	Low accuracy
Sequence antecedent selection algorithm	Atomic radius difference, electron concentration, work function	55.24794	less	High accuracy, considering the combination	Easy to fall into local optimum
Sequence post item selection algorithm	Valence electron concentration, Nabarro coefficient, work function	55.10984	less	High accuracy, considering the combination	Easy to fall into local optimum
Genetic algorithm	γ parameter, electron concentration, work function	50.46867	general	High accuracy, easy to find the global best	The result is not always

Pearson's correlation coefficient, univariate feature selection, and stability selection methods focus too much on the correlation between a single feature and the target variable, or the importance of prediction, and ignore the affect of the combination of several features [4]. It takes less time, but the accuracy is often not high, and it is difficult to perform accurate feature screening. The forward/backward sequence selection is dependent on the previous result. Such iterations are often difficult to consider the global impact. It is a greedy algorithm and is easy to fall into the global optimal and it takes more time than the former. For comparison, the genetic algorithm, it is a little time-consuming, but the feature combination sought is also ranked in the top few feature combination rankings. It is generally easy to find the best and get a better feature combination, but the initial population of the genetic algorithm is randomly generated, and the result of the final iteration is often uncertain. This paper chooses genetic algorithm as the optimal feature selection method, the selected optimal feature combination γ parameter, electron concentration, and work function are the physical feature combinations used in the subsequent steps of this article.

IV. High Reliability and Hardness Alloy Search

The high-entropy alloy search space of Al-Co-Cr-Cu-Fe-Ni system used in this paper comes from the high-entropy alloy search space 1895147 given in literature [2]. The 614143 high-entropy alloy composition data sets with predicted hardness values higher than 750HV searched out from the specific composition of alloys. Searching in such data sets is conducive to better searching for excellence the specific composition of the high-entropy alloy.

The specific composition of the high-entropy alloy with higher hardness found in the prediction search process is as follows:

Table 4. The specific composition and predicted hardness of the six high-entropy alloys with the highest hardness.

	Al	Co	Cr	Cu	Fe	Ni	Hardness
1	0.41	0.2	0.18	0	0.16	0.05	791.532194
2	0.41	0.16	0.18	0	0.17	0.08	791.5160014
3	0.41	0.16	0.19	0	0.16	0.08	791.5055031
4	0.41	0.2	0.19	0	0.15	0.05	791.4941856
5	0.41	0.17	0.18	0	0.17	0.07	791.4628587
6	0.41	0.19	0.18	0	0.16	0.06	791.4172123

The component proportion of each element in the high-entropy alloy shown in Table 4 is its corresponding mole ratio. The high-entropy alloy with the highest predicted hardness is: $Al_{0.41}Co_{0.2}Cr_{0.18}Fe_{0.16}Ni_{0.05}$, and the predicted hardness is 791.532194 HV.

V. CONCLUSIONS

In this paper, various feature screening methods are summarized and compared. Finally, through the comprehensive comparison of prediction accuracy and timeliness, the conclusion that the genetic algorithm feature selection method is the optimal feature screening method in the feature screening step of high entropy alloy reliability and hardness prediction process is obtained, and the prediction is obtained by repeated feature screening of genetic algorithm. The best performance combination of prediction features, gamma parameter, electron concentration, work function, compared with the 20 physical features in reference [2] which uses exhaustive method to select the optimal feature combination, this paper first determines the number of features in the feature combination, then selects the features with high correlation by Pearson correlation coefficient, and then selects the features by genetic algorithm, which saves more computing resources and time, and performs feature combination screening first. Then, by using machine learning models such as Gaussian kernel support vector regression, the process of high entropy alloy reliability and hardness prediction and high entropy alloy search in virtual space is efficient and stable. In this paper, a complete and general implementation framework of machine learning prediction of material properties is constructed, and a search method for specific components of high-performance materials in the vast virtual space is realized. Several kinds of high entropy alloys with high hardness were successfully predicted and recommended, and their corresponding reliability and hardness were predicted.

ACKNOWLEDGEMENT

The authors gratefully acknowledge financial support from the West Light Youth Talent Program of the Chinese Academy of Sciences (1_14) and the Natural Science Foundation of Qinghai Province (2021-ZJ-910).

REFERENCES

[1] Lauri H., Amber G., Adam S.F., and Patric R*,Data-driven materials science: status, challenges, and perspectives. Adv. Sci, 2019, 1900808, 1-23.
[2] Cheng W, Yan Z, Changxin W, Dezhen X,Machine learning assisted design of high entropy alloys with desired property . Acta Mater,

2019,170 10-117.
[3] Yan Z., Cheng W., Changxin W., Stoichko A., Dezhen X., Yang B., Yanjing S., Phase prediction in high entropy alloys with a rational selection of materials descriptors and machine learning models. Acta Mater, 2020, 185 528–539.
[4] Lei Z., Hongmei C., Xiaoma T., Hongguo C., Jingneng L., Yifang O., Qing P., Yong D., Machine learning reveals the importance of the formation enthalpy and atom-size difference in forming phases of high entropy alloys. Mater. Des, 2020, S0264-1275(20) 30369-5.
[5] Yeh J.W., Chen, Lin S.J., Gan J.Y., Chin T.S., Shun T.T., Tsau C.H., Chang S.Y., Nanostructured high-entropy alloys with multiple principal elements: novel alloy design concepts and outcomes. Adv. Eng. Mater, 2004, 6 299–303.
[6] Tsai M.H., Yeh J.W., High-entropy alloys: a critical review. Mater. Res. Lett., 2014, 2 10-123.
[7] Varvenne C., Luque A., Curtin W.A., theory of strengthening in fcc high entropy alloys. Acta Mater, 2016, 118 164-176.
[8] Wang W.Y., Shang S.L., Wang Y., Han F.B., Darling K.A., Wu Y.D., Xie X., Senkov O.N., Li J.S., Hui X.D., Dahmen K.A., Liaw P.K., Kecskes L.J., Liu Z.K., Atomic and electronic basis for the serrations of refractory high-entropy alloys. NPJ Comput. Mater, 2017, 32-10.
[9] Geman E., Bienenstock R., Doursat, Neural networks and the bias/variance dilemma. Neural Comput, 1992, 4 1-58.

AUTHOR'S PROFILE

First Author
Qian Lin, received the B.S. degree in Electronic Information Science and Technology from Qinghai Nationalities University, Xining, China, in 2004, the M.S. degree in Electronic Information Technology from Qinghai Normal University, Xining, China, in 2010, and the Ph. D. degree in Circuits and Systems from Tianjin University, Tianjin, China, in 2017. She is currently professor in Qinghai Nationalities University. Her research interests include the modeling, testing, simulation of circuit reliability, electronic technology application.

Second Author
Haifeng Wu, native of Baotou, Inner Mongolia, PhD, senior engineer, Chengdu Jiana Haiwei Science and Technology Co., Ltd., 29th Research Institute of China Electronic Science and Technology Corporation, main research interests are RF microwave circuit design, device modeling, circuit reliability analysis. email id: abgott@126.com