Big Data and Information Analytics

https://www.aimspress.com/journal/BDIA

BDIA, 9: 92–105.
DOI:10.3934/bdia.2025005
Received: 06 August 2025
Revised: 18 September 2025
Accepted: 20 September 2025
Published: 14 October 2025

Research article

Evading the strength-ductility trade-off dilemma in high-entropy alloys by multi-objective machine learning

Guanying Wei[1,2], **Jesper Byggmästar**[2], **Junzhi Cui**[3] **and Jingli Ren**[1,*]

[1] School of Mathematics and Statistics, Zhengzhou University, Zhengzhou 450001, China
[2] Department of Physics, University of Helsinki, P. O. Box 43, Helsinki 00014, Finland
[3] Institutes of Computational Mathematics, Chinese Academy of Sciences, Beijing 100190, China

* **Correspondence:** Email: renjl@zzu.edu.cn.

Abstract: The strength-ductility trade-off has limited the potential of many structural materials, especially in high-entropy alloys (HEAs). Here, we study the yield strength and ductility of HEAs with an experimental dataset consisting of 144 samples using multi-objective machine learning. First, we construct a feature pool including 20 features from the phase and mechanical properties of these HEAs, and utilize feature engineering to screen and rank the features. Then, the multi-objective random forest and MultiTaskLasso algorithms are chosen to train and predict the yield strength and ductility. The differences between multi-task Lasso and Lasso are compared. Moreover, through the interpretable feature analysis method—shapley additive explanation, the influence of the feature value on the mechanical properties of HEAs is analyzed.

Keywords: multi-objective; machine learning; high-entropy alloys; strength-ductility

1. Introduction

Understanding the super high strength-ductility of metal materials is of a great significance to the service safety, service life, and energy saving of metal materials. However, in most materials, the properties of strength and ductility are mutually exclusive [1, 2]. High-entropy alloys (HEAs) are an emerging class of materials with an array of highly desirable properties, including wear resistance [3], corrosion resistance [4], excellent structural stability [5, 6], and so on. It is beneficial to design HEAs with high strength and ductility.

Researchers design the HEAs with high strength and ductility according to change the compositions and experimental conditions [7,8]. Through changing the element Mn to Pd in the CrMnFeCoNi Cantor alloy, deformation mechanisms in the CrPdFeCoNi alloy are promoted by pronounced fluctuations in the composition and an increase in stacking-fault energy, which leads to a higher yield strength without

compromising the strain hardening and tensile ductility [9]. The evolution of the microstructure and compressive mechanical properties of $(CoCrFeNi)_x(Co_{0.26}Cr_{0.07}Fe_{0.16}Ni_{0.31}Hf_{0.4})$ with x variation was systematically investigated, where the change of microstructure from hyper-eutectic to hypo-eutectic produced a better combination of compressive strength and ductility than conventional $CoCrFeNiHf_x$ HEAs [10]. Moreover, Pan et al. showed that cyclic torsion on a HEA enhances the strength without degrading the ductility. Cyclic torsion creates a gradient of dislocations and low-angle grain boundaries from the surface to the interior that organize into tiny stacking faults and twin when straining begins. These structures allow for an improved ductility while simultaneously helping to harden the alloy [11]. The challenge is how to design the two properties concurrently with as few experiments or calculations as possible.

In the field of material science, state-of-the-art machine learning (ML) methods have been utilized to accelearate the alloy design [12–15]. Xiong et al. applied ML models to classify the different phase structures and to predict the hardness and ultimate tensile strength of complex concentrated alloys [16]. Wen et al. combined ML and the design of experiments method to accelerate the optimization process [17]. Ren et al. created win-wins from strength-ductility trade-off in multi-principal element alloys using ML [18]. However, these models [16–18] were all single target ML models, that is, a target needed a set of corresponding feature quantities to describe. Multiple targets need multiple sets of different feature quantities to describe, while multi-objective ML only uses a set of feature variables to describe multiple targets.

In this paper, MultiTaskLasso and a multi-objective random forest (RF) regression are utilized to predict the yield strength and ductility of HEAs. Feature engineering screens the features in the feature pool collected by the phase and mechanical parameters, and the features after screening are used to predict the two mechanical properties of the HEAs at the same time. Moreover, the linear expressions between these two mechanical properties and features are given by linear regression. Combining the shapley additive explanation (SHAP) with the multi-objective random forest, we analyze the effect of features on the yield strength and ductility of HEAs.

2. Methods

2.1. Data selection and analysis

The yield strength and ductility data of 144 HEAs fabricated by vacuum arc melting-cast under compression at room temperature were obtained from the report of Gorsse et al. [19]; for more information, see the sample data in Supplementary materials. The representation of features is important for an ML algorithm towards a specific material property [20]. The phase structures of the HEAs determine the mechanical properties [16]. We construct a feature pool including 20 features, 11 of which are phase parameters and the other 9 relate to the mechanical properties. The 11 phase parameters are as follows: 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 configurational entropy (ΔS); the Ω parameter [21] (which is related to the entropy, enthalpy and the melting point, see the equation in Supplementary materials), the Λ parameter [21] (which is related to an atom's configuration on a lattice and its radius; for more information, see the equation in Supplementary materials); the γ parameter [21] (the solid angles of atomic packing for the elements with the largest and smallest atomic sizes); the local electronegativity

mismatch (D.χ) between elements; the number of itinerant electrons (e/a); and the cohesive energy (Ec), all of which potentially relate to the phase formation [21]. The 9 mechanical features include the following: the modulus mismatch (η) [22]; the local size mismatch (D.r); the energy term in the strengthening model (A) [23]; the Peierls-Nabarro factor (F) [24]; sixth square of the work function (w) [25]; the shear modulus (G); the local shear modulus mismatch (D.G); the difference in shear modulus (δG); and the lattice distortion energy (μ) [25, 26]. A detailed list of these features is provided in the Supplementary materials, Table S1. The datasets were exectued using the Jupyter Notebook as a DataFrame [27]. Before training, these features were normalized by the maximum and the minimum so that their values fell between 0 and 1. This eliminates the order-of-magnitude effect between features.

$$X^* = \frac{X - X_{\min}}{X_{\max} - X_{\min}}, \tag{2.1}$$

where X^* is the value of feature X after normalization, and $X_{\min}$ and $X_{\max}$ are the maximum and minimum values of feature X, respectively.

2.2. Machine learning algorithms and evaluation method

ML algorithms that can be used for multi-objective prediction, such as MultiTaskLasso [28] and multi-objective RF [29], are applied to predict the yield strength and ductility of HEAs. These models were chosen because MultiTaskLasso enhances the interpretability through joint feature selection, while multi-objective RF algorithms are robust to nonlinear relationships, which make them well-suited for capturing the strength-ductility behavior of HEAs. In addition, linear regression algorithms are utilized to regress the features and mechanical properties. These algorithms are found within the open-source algorithm package in Scikit-learn [30]; for more information, see the Supplementary materials for the algorithm principle. The algorithm parameters choose the default values.

K-fold cross-validation (CV) randomly divides the data set of N samples into K parts, of which $K-1$ is used as the training set and the remaining 1 is used as the verification set. The trained model is used to predict the remaining sample set and obtain the predicted value. In this way, K predicted values are obtained after K cycles, which gives the prediction error of the verification set of all data. Due to a limited sample size, we used the 10-fold CV method to produce more training data. After the 10-fold CV is completed, the 10 models of CV are used to predict the new samples, and the mean value of the predicted values of the 10 models is taken as the output value. In this paper, the Pearson correlation coefficient R is chosen as the prediction performance evaluation index, which is defined as follows:

$$R = \frac{\sum_{i=1}^{N}(y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^{N}(y_i - \bar{y})^2 \sum_{i=1}^{N}(\hat{y}_i - \bar{\hat{y}})^2}}, \tag{2.2}$$

where y_i and $\hat{y}_i$ represent the experimental value and the predicted value of each data respectively, and $\bar{y}$ and $\bar{\hat{y}}$ represent the mean value of the experimental value and the mean value of the predicted value of all data respectively.

3. Results and discussion

3.1. Feature selection and modeling based on multi-objective random forest

To reduce the computation time and improve the model robustness by removing the irrelevant and redundant features, we employ a hybrid method that combines a correlation analysis and a wrapper method to perform the feature selection. The heat map in Figure 1 depicts the Pearson correlation coefficients map between the different features.

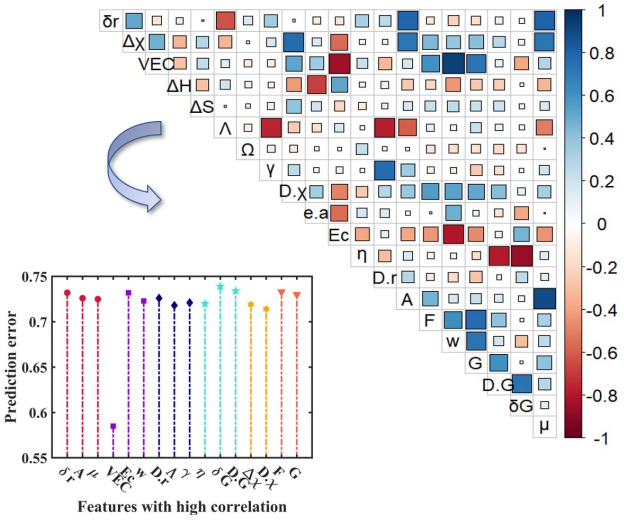

Figure 1. Feature selection to identify the most important features from the feature pool. The top right of the figure is a Pearson correlation map of the initial twenty features. A larger square with a darker color indicates a high level of correlation. The bottom left of the figure is the influence of high correlation features on the model. Each color represents the features that are highly correlated with each other, and among each color, the one with the lower correlation is retained.

Those with a correlation coefficient greater than 0.75 are considered as highly correlated. For these highly correlated features, we only retain one of them to reduce redundant information. To choose the feature to be retained, we remove one of these highly correlated features to compare the test errors of the model, and then rank the importance of each feature. We split our original data into a training set including 115 samples (80%) and a testing set including 29 samples (20%); then, we build the multi-objective RF model on the training set with all features except the one in the set with a high correlation removed. The correlation between the predicted values and the experimental values is calculated through the evaluation index R, and the correlations are ranked, as shown in Figure 1. The smaller R is, the more important this feature is to the model; deleting it will reduce the prediction ability of the model, so this feature should be retained. Accordingly, the highly correlated features δr, A, Ec, w, D.r, γ, δG, D.G, $\Delta\chi$, and F are removed.

We further reduce the feature space by considering all possible subsets of these features to identify the subset giving rise to the lowest model error. The errors $1-R$ with 10-fold CV of the multi-objective

RF model based on different subsets of features are plotted in Figure 2. According to Figure 2, as the number of features increases, the error of the model $1 - R$ initially decreases. Then, the error of the model increases as the number of features increases, which may be due to overfitting [31]. The best performance of the model is given by a 7-tuple feature, where the features are VEC, ΔH, ΔS, Λ, Ω, D.χ, and e/a. We call the subset composed of these 7 features as the RF-Subset.

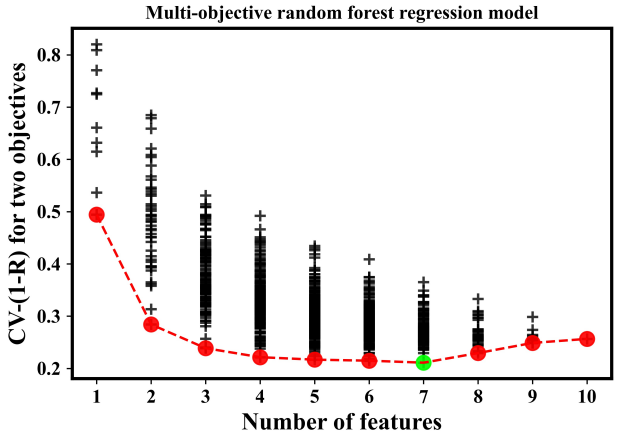

Figure 2. The 10-fold cross-validation error of each possible multi-objective RF model containing a subset of processing features. The red frontier tracks the best model for a given number of features. The green frontier gives the best number of features to train the multi-objective RF model.

Based on the RF-Subset, we use a multi-objective RF model combined with 10-fold CV for training. The 10 models of CV are used to predict the corresponding folded test set, and the predicted values of the two target performances (i.e., strength and ductility) are given. The 10 times predicted values of the two performances are averaged as the final predicted value. Figure 3(a),(b) are the comparison charts of the model predicted value and the experimental measurement value of the two target performances, where the evaluation indices R are 0.9516 and 0.9432, respectively.

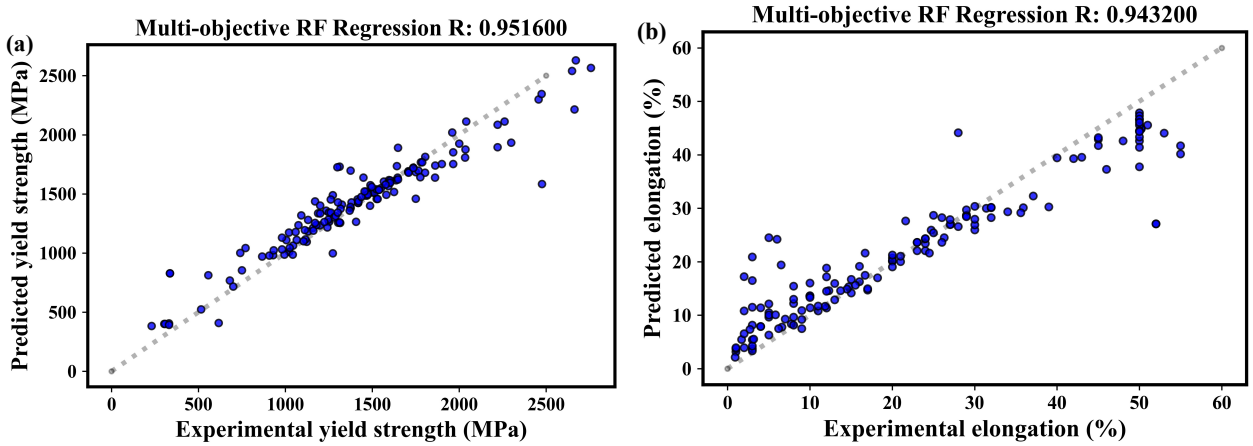

Figure 3. The comparison between the experimental values and predicted values of the multi-objective random forest (RF) models based on RF-Subset: (a) yield strength, (b) ductility.

3.2. Feature selection and modeling based on MultiTaskLasso

A multiTaskLasso regression and a single-objective Lasso regression are two linear regression methods with penalty items, which can be used to filter features by adjusting the penalty item parameter α. The parameter α determines the weight of the penalty item in the loss function. As α increases, each feature coefficient will become smaller. When α increases to a certain value, all coefficients will become 0. The feature whose coefficient finally becomes 0 is considered to be the most important feature [32], and so on, until we obtain the feature importance ranking. The curve in which the characteristic coefficient changes with α is called the regression path. The regression path diagram generally uses $-\log_{10}(\alpha)$ as the horizontal coordinate. Additionally, we also normalize the values of the yield strength and ductility in the regression path because the values of these two objectives are not in the same order of magnitude. Figure 4(a),(b) show the change of characteristic coefficients which correspond to the yield strength and ductility with α in the MultiTaskLasso regression, respectively. In a multi-objective regression, the two mechanical properties are described by the same set of characteristic quantities, and the order of characteristic importance is shown in Table 1. Figure 4(c),(d) show the change of characteristic coefficients which correspond to the yield strength and ductility with α in the single-objective Lasso regression, respectively. Table 2 shows the single-objective Lasso feature ranking results which correspond to the two mechanical properties. It can be seen that the features that play a leading role in the two negative correlation targets are not exactly the same. This is the main difference between single-objective and multi-objective ML.

Among the top five important features in Table 2, ΔH, η, and ΔS are all important for the yield strength and ductility; alternatively, VEC is more important for the yield strength and e/a is more important for the ductility. The feature importance ranking of the MultiTaskLasso regression in Table 1 integrates the feature importance ranking of two single objectives.

Table 1. Features ranking based on the MultiTaskLasso regression.

Features ranking	1	2	3	4	5	6	7	8	9	10
Features name	ΔH	η	VEC	ΔS	e/a	Λ	G	D.χ	μ	Ω

Therefore, it is meaningful to use the MultiTaskLasso regression to train and predict the yield strength and ductility at the same time. Additionally, we reduce the feature space by considering all possible subsets of these features to identify the subset that produces the lowest model error. The feature number is 9 when the model has the smallest evaluation error, which can be seen within the figure in the Supplementary materials, where the features are VEC, ΔH, ΔS, Λ, Ω, D.χ, e/a, η, and G.

Table 2. Features ranking based on the Lasso regression.

Features ranking	1	2	3	4	5	6	7	8	9	10
Yield strength	ΔH	VEC	η	Λ	ΔS	G	e/a	Ω	μ	D.χ
Ductility	ΔH	η	ΔS	e/a	D.χ	μ	Ω	Λ	VEC	G

Based on the features after screening, we use the MultiTaskLasso regression combined with 10-fold CV for training. The method which utilizes 10-fold CV is same as the multi-objective RF. The evaluation indices R between the experimental value and the model predicted value of the yield strength and ductility are 0.6963 and 0.6897, respectively.

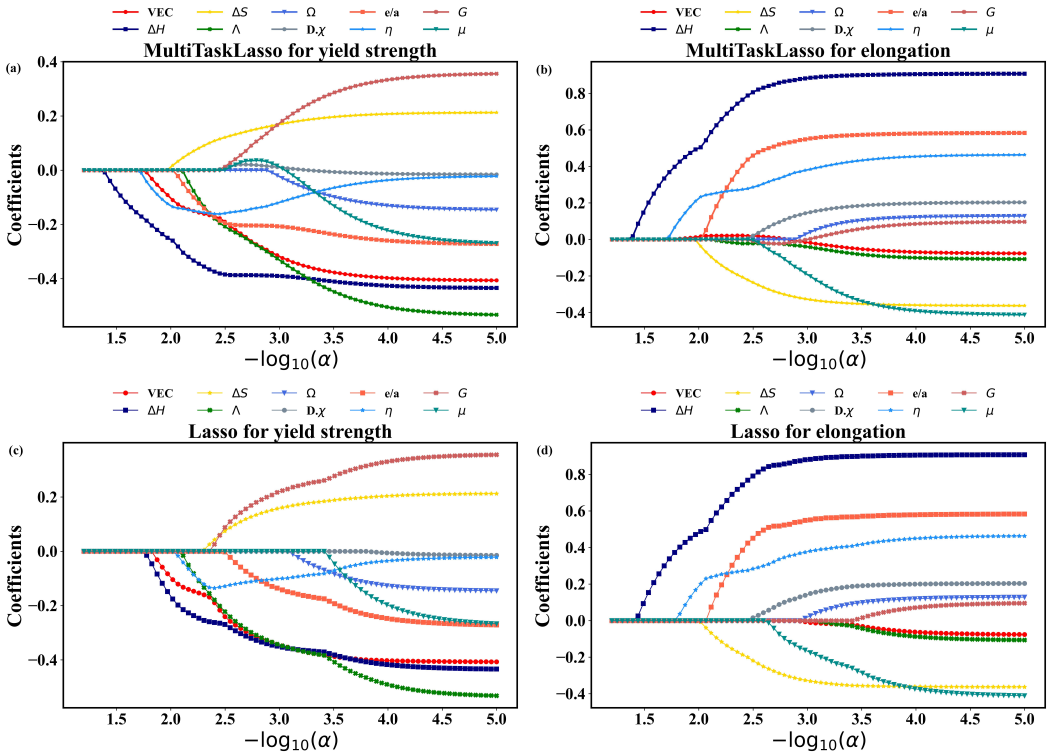

Figure 4. MultiTaskLasso and Lasso regressions for the yield strength and ductility, and the coefficient value of the corresponding characteristics changes with α: (a) yield strength under the MultiTaskLasso regression, (b) ductility under the MultiTaskLasso regression, (c) yield strength under the Lasso regression, (d) ductility under the Lasso regression.

It is obvious that the prediction result by a multi-objective RF is better than the MultiTaskLasso. Additionally, we choose the multi-objective RF to study the influence of features after screening on the yield strength and ductility.

3.3. Interpretability analysis of features

Multi-objective RF is an integrated learning algorithm based on a decision tree. Because multi-objective RF is a nonlinear black box algorithm, it is difficult to directly explore the impact of a feature on the target variables. SHAP is combined with the multi-objective RF algorithm to analyze the influence of features on the mechanical properties of HEAs through the SHAP value, so as to improve the interpretability of the multi-objective RF model. The SHAP model is an additive explanatory model inspired by the SHAP value. For each prediction sample, the model generates a prediction value, and the SHAP value is the value assigned to each feature in the sample. Assuming that the *i*th sample is x_{i}, the *j*th feature of the *i*th sample is $x_{\mathrm{i,j}}$, the predicted value of the model for the *i*th sample is y_{i}, and the baseline of the whole model (usually the mean value of the target variables of all samples) is y_{base}, then the SHAP value is calculated using the following equation:

$$y_i = y_{\mathrm{base}} + f(x_{i,1}) + f(x_{i,2}) + \cdots + f(x_{i,k}), \tag{3.1}$$

where $f(x_{i,j})$ is the SHAP value of $x_{\mathrm{i,j}}$. Intuitively, $f(x_{i,1})$ is the contribution value of the first feature on the ith sample to the final predicted value y_{i}. If $f(x_{i,1}) > 0$, then it indicates that the first feature improves the predicted value and has a positive effect. If $f(x_{i,1}) < 0$, then it shows that the first feature reduces the predicted value and has a negative effect.

After two kinds of multi-objective feature selection and training, the prediction effect of the multi-objective RF is better. Thus we rank the 7 features of the RF-Subset by multi-objective RF. The idea of a RF feature importance evaluation is to see how much each feature contributes to each tree in the RF, then take an average value, and finally compare the contribution between the features. The contribution can usually be measured by the Gini index or the out-of-bag error rate. The result of RF feature importance ranking is shown in Figure 5. The ranking of features is as follows: VEC, ΔS, ΔH, e/a, Λ, D.χ and Ω.

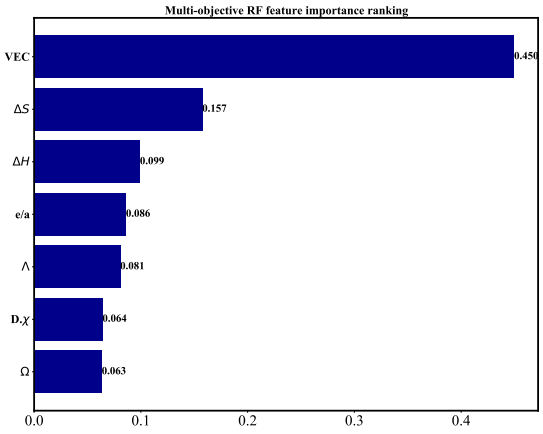

Figure 5. Feature importance ranking for the RF-Subset based on a multi-objective RF.

All 144-sample data are linearly regressed using the RF-Subset to fit the linear analytical formulas for the two target properties of yield strength (YS, MPa) and elongation (E, %) as follows:

$$\begin{aligned} YS \ = & -696.28\mathrm{V}EC + 462.83\Delta S - 1181.44\Delta H - 907.19\mathrm{e}/a \\ & - 797.16\Lambda + 227.84D.\chi - 473.95\Omega + 2823.84, \end{aligned} \tag{3.2}$$

$$\begin{aligned} E \ = & +3.33\mathrm{V}EC - 20.43\Delta S + 56.52\Delta H + 44.35\mathrm{e}/a \\ & - 1.97\Lambda + 2.15D.\chi + 17.14\Omega - 33.46. \end{aligned} \tag{3.3}$$

In these two formulas, the units of ΔS, ΔH, Λ, and Ω are JK^{-1} mole^{-1}, kJmole^{-1}, JK^{-1} mole^{-1}, and K^{-1}, respectively. VEC, e/a, and D.χ are dimensionless values. The coefficients of the first four important features are opposite. In general, the face-centered cubic (FCC) HEAs exhibits a high ductility but poor strength. The body-centered cubic (BCC) HEAs exhibits a high yield strength but is inferior to the FCC structure in terms of ductility. VEC has a negative effect on the yield strength from Eq (3.2) and a positive effect on the strength from Equation (3.3). A low VEC is conducive to

forming a BCC phase and improving an alloy's strength, while a high VEC is beneficial in forming an FCC phase that improves an alloy's ductility [33].

With the default parameters, the multi-objective RF algorithm uses all samples to train a multi-objective prediction model. Based on the multi-objective prediction model, we utilize the SHAP algorithm package developed by Lundberg et al. [34] to calculate the SHAP values, which can be used to analyze the influence of features on the mechanical properties of the HEAs. The SHAP values of four most important features, namely VEC, ΔS, ΔH, and e/a, from the results of the feature ranking from multi-objective RF are calculated. The effects of these four features on the target properties of the HEAs are analyzed through the SHAP values.

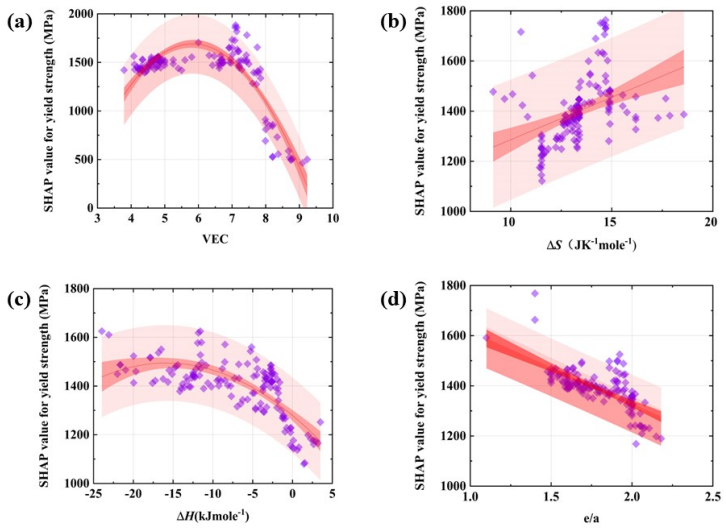

Figure 6. Using the SHAP value which calculated from the RF model, to analyze the effects of the following on the yield strength: VEC (a), ΔS (b), ΔH (c), and e/a (d).

In Figure 6(a)–(d), the x-axis is VEC, ΔS, ΔH and e/a of 144 samples, respectively. The y-axis is the values of the SHAP value which corresponds to VEC, ΔS, ΔH and e/a of 144 samples, respectively, plus the mean value of the yield strength. In Figure 6, the change of the y-axis with the x-axis can be considered as the change of the yield strength of the HEAs with the four characteristic values, which have certain physical significance. Figure 6(a),(c),(d) generally show a downward trend, that is, the SHAP value decreases with an increase of the feature value. Figure 6(b) exhibits an increasing trend. In the linear Eq (3.2), the characteristic coefficient of ΔS is positive, which indicates that the yield strength is positively correlated with ΔS, that is, increasing the value of ΔS can improve the yield strength. The characteristic coefficients of VEC, ΔH, and e/a are negative, indicating that the yield strength is negatively correlated with VEC, ΔH and e/a, that is, increasing the values of these three characteristics will reduce the yield strength of the HEAs. The influence of VEC, ΔS, ΔH, and e/a on the yield strength is analyzed by SHAP value, and the rule obtained is consistent with that obtained by a linear regression model. Since the ductility and yield strength are a pair of trade-off features, the influence of the features in Figure 7 on the ductility is opposite to that in Figure 6. Moreover, through the influence of the four important features analyzed by the SHAP value in Figure 7 on the ductility, the obtained rule is consistent with the rule of the linear regression Eq (3.3).

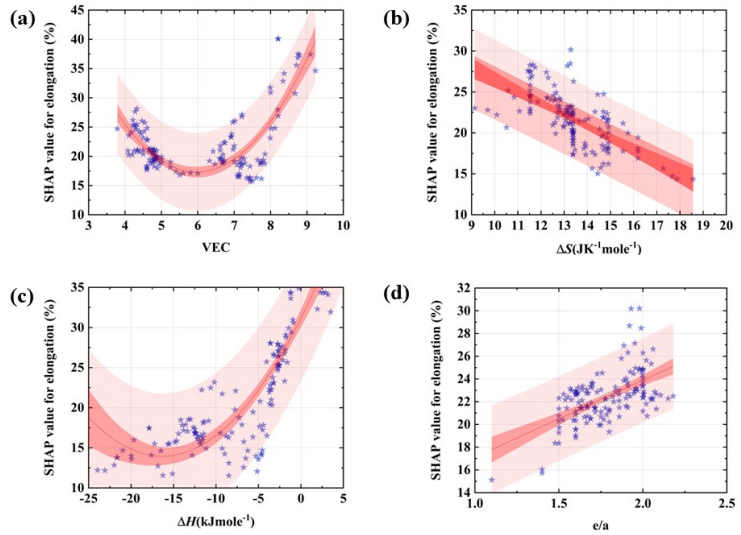

Figure 7. Using the SHAP value, which is calculated from the RF model, to analyze the effects of the following on the ductility: VEC (a), ΔS (b), ΔH (c), and e/a (d).

3.4. Design of high yield strength and high ductility HEAs

Table 3. The feature values and ML-predicted mechanical properties of Hf_ and Nb_ base HEAs, where numbers in parentheses are experimental values.

Alloys	VEC	ΔS	ΔH	e/a	Yield strength (MPa)	ductility (%)
HfNbTaTiWZr	4.676	14.9098	−2.1196	1.837	1552.10 [1550.00]	23.249 [26.300]
HfMoNbTaTi	4.800	13.3809	−1.4400	1.600	1388.89 [1369.00]	24.367 [27.000]
HfCrNbTaTiZr	4.676	14.9098	−2.7890	1.670	1520.00	24.547
$NbTiV_{0.3}Mo_{0.3}Zr$	4.527	12.3113	−0.2680	1.639	1201.38 [1312.00]	37.606 [50.000]
$NbTiV_{0.3}Mo_{0.4}Zr$	4.558	12.4710	−0.7200	1.618	1247.02	38.972
$NbTiV_{0.3}W_{0.3}Zr$	4.553	12.3200	−0.9300	1.732	1252.40	38.390
$NbTiV_{0.3}Mo_{0.1}Zr$	4.436	11.6086	0.8044	1.675	1116.45 [932.00]	41.748 [45.000]
$NbTiV_{0.3}Mo_{0.2}Zr$	4.490	12.0390	0.2320	1.654	1166.34	41.278
$NbTiV_{0.3}W_{0.1}Zr$	4.436	11.6150	0.5566	1.704	1011.08	45.540

The SHAP value of these top features provides guidance to the design of novel HEAs with a high yield strength and high ductility. In general, HEAs with a yield strength larger than 1200 MPa and a ductility larger than 20% are considered to have a high yield strength and high ductility. Combining Figures 6 and 7, when $VEC < 5.0$, $11.5 < \Delta S < 15.0$, $-3.6 < \Delta H < 0.0$, and $1.5 < e/a < 2.0$, the yield strength and ductility of the HEAs are larger than 1200 MPa and 20%, respectively. Because these two mechanical properties are contradictory, we call the value range of these features that make the HEAs have a high yield strength and a high ductility at the same time as the balanced interval. A simple calculation gives the feature values of VEC, ΔS, ΔH, and e/a in Hf_ and Nb_ base HEAs, which are shown in Table 3. Clearly, if the feature values are in the balanced interval, then the alloys have a high predicted value of the yield strength and ductility, and the experimental values are in parentheses.

At the last three lines, the values of ΔH are larger than 0 (not in the balanced interval), and the values of the yield strength are smaller than 1200. Therefore, these four feature values of the HEAs in the balanced interval can obtain the HEAs with a high yield strength and ductility simultaneously.

4. Conclusions

The present work illustrated a ML framework to predict the yield strength and ductility of HEAs. A dataset of 144 samples was constructed, and 20 initial features were proposed. The hybrid method combined a correlation analysis and a wrapper method to produce the seven features of VEC, ΔH, ΔS, Λ, Ω, D.χ, and e/a in the multi-objective RF regression of the yield strength and ductility, alongside the nine features in the MultiTaskLasso regression. The multi-objective RF model regressed the two mechanical properties better. At the same time, we utilized the Lasso and MultiTaskLasso to regress the yield strength and ductility, and compared the differences between the single-objective and multi-objective ML models. Analytic expressions for the two target properties with features after screening were given by a linear regression. Through the interpretable characteristic analysis method, namely the SHAP value, the influence of the feature value change on the yield strength and ductility of HEAs was obtained. Therefore, we designed new alloys in the balanced intervals, that is, VEC < 5.0, $11.5 < \Delta S < 15.0$, $-3.6 < \Delta H < 0.0$, and $1.5 <$ e/a < 2.0, to obtain HEAs with a high yield strength and high ductility simultaneously.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

This work was supported by the National Natural Science Foundation of China [No. U23A2065].

Conflict of interest

The authors declare that there is no conflict of interest.

References

1. Ritchie RO, (2011) The conflicts between strength and toughness. *Nat Mater* 10: 817–822. https://doi.org/10.1038/nmat3115
2. Wei Y, Li Y, Zhu L, Liu Y, Lei X, et al. (2014) Evading the strength-ductility trade-off dilemma in steel through gradient hierarchical nanotwins. *Nat Commun* 5: 3580. https://doi.org/10.1038/ncomms4580
3. Poletti MG, Fiore G, Gili F, Mangherini D, Battezzati L, Development of a new high entropy alloy for wear resistance: $FeCoCrNiW_{0.3}$ and $FeCoCrNiW_{0.3}$ + 5at.% of C. *Mater Des* 115: 247–254. https://doi.org/10.1016/j.matdes.2016.11.027

4. Ding J, Inoue A, Han Y, Kong FL, Zhu SL, Wang Z, et al. High entropy effect on structure and properties of (Fe,Co,Ni,Cr)-B amorphous alloys. *J Alloys Compd* 696: 345–352. https://doi.org/10.1016/j.jallcom.2016.11.223

5. Wu YD, Cai YH, Wang T, Si JJ, Zhu J, Wang YD, et al. (2014) A refractory $Hf_{25}Nb_{25}Ti_{25}Zr_{25}$ high-entropy alloy with excellent structural stability and tensile properties. *Mater Lett* 9: 130–277. https://doi.org/10.1016/j.matlet.2014.05.134

6. Guo X, Xie X, Ren J, Laktionova M, Tabachnikova E, Yu L, et al. (2017) Plastic dynamics of the $Al_{0.5}$CoCrCuFeNi high entropy alloy at cryogenic temperatures: Jerky flow, stair-like fluctuation, scaling behavior, and non-chaotic state. *Appl Phys Lett* 111: 251905. https://doi.org/10.1063/1.5004241

7. Sharma V, Wang C, Lorenzini R, Ma R, Zhu Q, Sinkovits D, et al. (2014) Rational design of all organic polymer dielectrics. *Nat Commun* 5: 4845. https://doi.org/10.1038/ncomms5845

8. Yuan R, Xue D, Zhou Y, Ding X, Sun J, Lookman T, (2019) The search for $BaTiO_3$-based piezoelectrics with large piezoelectric coefficient using machine learning. *IEEE Trans Ultrason Ferroelectr Freq Control* 66: 394–401. https://doi.org/10.1109/TUFFC.2018.2888800

9. Ding QQ, Zhang Y, Chen X, Fu XQ, Chen DK, Chen SJ, et al. (2019) Tuning element distribution, structure and properties by composition in high-entropy alloys. *Nature* 574: 223–227. https://doi.org/10.1038/s41586-019-1617-1

10. Xie TB, Xiong ZP, Liu Z, Deng GY, Cheng XW, (2021) Excellent combination of compressive strength and ductility of $(CoCrFeNi)_x(Co_{0.26}Cr_{0.07}Fe_{0.16}Ni_{0.31}Hf_{0.4)}$ high-entropy alloys. *Mater Des* 202: 109569. https://doi.org/10.1016/j.matdes.2021.109569

11. Pan QS, Zhang LX, Feng R, Lu QH, An K, Chuang AC, et al. (2021) Gradient cell-structured high-entropy alloy with exceptional strength and ductility. *Science* 374: 984–989. https://doi.org/10.1126/science.abj8114

12. Zhou ZQ, Zhou YJ, He QF, Ding ZY, Li FC, Yang Y, (2019) Machine learning guided appraisal and exploration of phase design for high entropy alloys. *npj Comput Mater* 5: 128. https://doi.org/10.1038/s41524-019-0265-1

13. Wei GY, Cui JZ, Wang W, Guo XX, Ren JL, Wang WH, (2022) Short-to-medium range structure and glass-forming ability in metallic glasses. *Phys Rev Mater* 6: 055601. https://doi.org/10.1103/PhysRevMaterials.6.055601

14. Lookman T, Balachandran P, Xue DZ, Yuan RH, (2019) Active learning in materials science with emphasis on adaptive sampling using uncertainties for targeted design. *npj Comput Mater* 5: 21. https://doi.org/10.1038/s41524-019-0153-8

15. Chang HN, Tao YW, Liaw PK, Ren JL, (2022) Phase prediction and effect of intrinsic residual strain on phase stability in high-entropy alloys with machine learning. *J Alloys Compd* 921: 166149. https://doi.org/10.1016/j.jallcom.2022.166149

16. Xiong J, Shi SQ, Zhang TY, (2021) Machine learning of phases and mechanical properties in complex concentrated alloys. *J Mater Sci Technol* 87: 133–142. https://doi.org/10.1016/j.jmst.2021.01.054

17. Wen C, Zhang Y, Wang CX, Xue DZ, Bai Y, Antonov S, et al. (2019) Machine learning assisted design of high entropy alloys with desired property. *Acta Mater* 170: 109–117. https://doi.org/10.1016/j.actamat.2019.03.010
18. Wu LL, Wei GY, Wang G, Wang HY, Ren JL, (2022) Creating win-wins from strength-ductility trade-off in multi-principal element alloys by machine learning. *Mater Today Commun* 32: 104010. https://doi.org/10.1016/j.mtcomm.2022.104010
19. Gorsse S, Nguyen M, Senkov O, Miracle D, (2018) Database on the mechanical properties of high entropy alloys and complex concentrated alloys. *Data Brief* 21: 63–69. https://doi.org/10.1016/j.dib.2018.11.111
20. Lu S, Zhou Q, Ouyang Y, Guo Y, Li Q, Wang J, (2018) Accelerated discovery of stable lead-free hybrid organic-inorganic perovskites via machine learning. *Nat Commun* 9: 3405. https://doi.org/10.1038/s41467-018-05761-w
21. Guo S, (2015) Phase selection rules for cast high entropy alloys: An overview. *Mater Sci Technol* 31: 1223–1230. https://doi.org/10.1179/1743284715Y.0000000018
22. Toda-Caraballo I, Rivera-Díaz-del-Castillo PEJ, (2015) Modelling solid solution hardening in high entropy alloys. *Acta Mater* 85: 14–23. https://doi.org/10.1016/j.actamat.2014.11.014
23. Varvenne C, Luque A, Curtin WA, (2016) Theory of strengthening in fcc high entropy alloys. *Acta Mater* 118: 164–176. https://doi.org/10.1016/j.actamat.2016.07.040
24. Hu GX, Cai X, Rong YH, (2010) *Material Science Foundation*, 3 Eds., Shanghai Jiao Tong University Press.
25. Wang WY, Shang SL, Wang Y, Han FB, Darling KA, Wu YD, et al. (2017) Atomic and electronic basis for the serrations of refractory high-entropy alloys. *npj Comput Mater* 3: 2–10. https://doi.org/10.1038/s41524-017-0024-0
26. Wang Z, Fang Q, Li J, Liu Y, (2018) Effect of lattice distortion on solid solution strengthening of BCC high-entropy alloys. *J Mater Sci Technol* 34: 349–354. https://doi.org/10.1016/j.jmst.2017.07.013
27. Kluyver T, Ragan-Kelley B, Perez F, Granger B, Bussonnier M, Frederic J, et al. (2016) Jupyter notebooks—a publishing format for reproducible computational workflows, In: *Positioning and Power in Academic Publishing: Players, Agents and Agendas*, IOS Press, 87–90. https://doi.org/10.3233/978-1-61499-649-1-87
28. Zhou Y, Jin R, Hoi SC, (2010) Exclusive lasso for multi-task feature selection, In: *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, 9: 988–995.
29. Breiman L, (2001) Random forest. *Mach Learn* 45: 5–32. https://doi.org/10.1023/A:1010933404324
30. Swami A, Jain R, (2013) Scikit-learn: Machine learning in Python. *J Mach Learn Res* 12: 2825–2830.
31. Geman S, Bienenstock E, Doursat R, (1992) Neural networks and the bias/variance dilemma. *Neural Comput* 4: 1–58. https://doi.org/10.1162/neco.1992.4.1.1
32. Wei QH, Xiong J, Sun S, Zhang TY, (2021) Multi-objective machine learning of four mechanical properties of steels. *Sci Sin. Tech* 51: 722–736. https://doi.org/10.1360/SST-2020-0475

33. Chen R, Qin G, Zheng H, Wang L, Su Y, Chiu Y, et al. (2018) Composition design of high entropy alloys using the valence electron concentration to balance strength and ductility. *Acta Mater* 144: 129–137. https://doi.org/10.1016/j.actamat.2017.10.058

34. Lundberg SM, Lee SI, (2017) A unified approach to interpreting model predictions. *Adv Neural Inf Process Syst* 30: 4765–4774.

AIMS Press