



Accelerated discovery of single-phase refractory high entropy alloys assisted by machine learning

Yonggang Yan, Dan Lu, Kun Wang^{*}

School of Engineering, Alfred University, Alfred, NY 14802, USA

ARTICLE INFO

Keywords:

Refractory high entropy alloys
Machine learning
Experimental validation

ABSTRACT

Herein, we proposed a strategy to design single-phase refractory high entropy alloys (RHEAs) with the assistance of machine learning algorithms. Based on an extensive dataset (1807 entries) built in this work, we applied multiple machine learning algorithms to train the dataset. After the blind test, we found that the Gradient boosting (GB) model can distinguish the single-phase-solid solution and non-single-phase-solid solution alloys with a test accuracy of 96.41%. Given the GB model, we predicted over 100 equiatomic oxidation-resistance RHEAs from the composition space of eight metallic elements. After that, we synthesized ten of these predicted single-phase RHEAs by mechanical alloying. The XRD patterns show that all of them are single-phase BCC solid solution. The experimental results agree well with the prediction results, indicating the excellent performance of the machine learning model in single-phase RHEAs prediction. With the aid of the machine learning method, single-phase oxidation-resistant RHEAs were successfully designed. Our work presents a novel strategy with outstanding performance and evident effectiveness on the accelerated discovery of novel metallic materials used for extreme environments.

1. Introduction

High-temperature oxidation of the alloys — the chemical interaction between alloys and the atmospheric oxygen at elevated temperatures—drastically deteriorates the mechanical performances of the alloys and rapidly leads to failure during their service if oxidation is unhindered [1,2]. Usually, 0.1–10 wt% oxidation-resistant elements, such as Al, Ti, Cr, Si etc., are added to the base-alloys to improve the oxidation-resistance. The oxidation-resistant elements prefer to form an oxidation layer on the surface to prevent primary elements from being oxidized in the deeper region at elevated temperatures [3–12]. Currently, the Ni-base superalloys, such as IN617 [2,13–17], IN738LC [2,18–22], René N4 [2,23,24], and René N5 [2,22,25–27], have good oxidation-resistance abilities at high temperatures up to ~ 1000 °C, and primarily used in gas turbine. To further improve the efficiency of the gas turbine, the application temperatures need to be pushed to higher temperature, i.e. > 1200 °C. However, there is little room available for Ni-based superalloys to further improve the application temperature due to the limitation of their melting points. Therefore, the development of new metallic structural materials used at extreme environments are very critically needed.

The emerging refractory high entropy alloys (RHEAs) composed of multiple primary refractory metal elements such as Ti, Zr, Hf (Group IV), V, Nb, Ta (Group V), Cr, Mo, W (Group VI), were proposed with the inspiration of the idea of high entropy alloys (HEAs) in the past decade [28–31]. On the one hand, as a unique group of high entropy alloys (HEAs), RHEAs inherits the core characteristics from the HEAs [29,32], the outstanding high-temperature properties of the HEAs are primarily attributed to the sluggish diffusion [33–38]. On the other hand, the refractory elements with high melting points improve the high-temperature mechanical properties. Therefore, RHEAs are considered a potential replacement for the existing superalloys used in aerospace, next-generation nuclear reactors, and energy industries. Since the first results reported by Senkov et al. in 2010 [30], RHEAs have attracted the majority of investigations [11,39–48]. Regarding the high-temperature oxidation-resistant RHEAs, intensive efforts have been made to pursue the alloys with better high-temperature oxidation-resistance capability in the combinations of the multiple refractory elements and the oxidation-resistant elements [7,8,11,41,43,46,49–52]. For instance, Gorr and his colleagues investigated the high-temperature oxidation behaviors of a family of RHEAs with the composition of X-AlCrMoTi (X represents W, Nb, or Ta) [53,54], and NbMoCrTiAl RHEAs with the

^{*} Corresponding author.

E-mail address: wang@alfred.edu (K. Wang).

<https://doi.org/10.1016/j.commsci.2021.110723>

Received 26 March 2021; Received in revised form 24 June 2021; Accepted 11 July 2021

Available online 24 July 2021

0927-0256/© 2021 Elsevier B.V. All rights reserved.

addition of 1 at.% Si [49], to discover alloys with high oxidation-resistance at elevated temperatures. The TaMoCrTiAl RHEA shows the most superior oxidation-resistant ability compared with other RHEAs studied. And the addition of a small amount of Si significantly improves the high-temperature oxidation-resistance although without the formation of the silicon oxide layer. In addition, to the best of our knowledge, the single-phase-solid solution (SS) including face centered cubic (FCC), body centered cubic (BCC) and hexagonal closed packed (HCP) structures, may generally exhibit good mechanical properties with favorable ductility and acceptable strength, while the intermetallic phase or amorphous phase tend to present embrittlement [55,56]. Thus, we aim to design the single-phase-SS RHEAs with superior oxidation resistance.

Early attempts of the HEAs design mainly relied on trial-and-error experiments. Unfortunately, it is rather restricted by some limitations. For instance, the experimental process generally imposes high requirements in terms of experimental resources and equipment. Furthermore, the experimental process has to be conducted over a long period for a small number of HEAs. Thus, it is almost impossible to explore the HEAs with high efficiency in the extremely large composition space only relying on experiments. Later on, the computational methods, like Density Functional Theory (DFT) [29,57–63], Molecular-Dynamics-simulation [64–67], and Calculation of Phase Diagrams (CALPHAD) [68–76], have been applied to explore the phase in composition space far more efficiently. However, these methods also exhibit expensive computational costs as well as uncertainties. More recently, some physical criteria, based on thermodynamic or topological parameters, have been widely used to predicting the phase of multiple elemental combination. Typically, the mixing entropy (ΔS_{mix}) and mixing enthalpy (ΔH_{mix}) are the thermodynamic parameters used to determine the HEAs phase. Besides, some parameters derived from Hume-Rothery (H-R) rules, such as atomic radius difference (δ), electronegativity difference ($\Delta\chi$), valence electron concentration (VEC), are widely used to study the phase of HEAs. For example, Zhang et al. [77] found that the formation of solid solution (SS) HEAs requires $-20 < \Delta H_{mix} < 5\text{kJ/mol}$, $12 < \Delta S_{mix} < 17.5\text{kJ/K}\cdot\text{mol}$, and $0 < \delta < 6.4$ (Percentage of the atomic size difference). Guo et al. [78] observed that the SS HEAs have ΔH_{mix} in the range of $-22 \sim 7\text{ kJ/mol}$, ΔS_{mix} in the range of $11 \sim 19.5\text{ kJ/mol}$, and the δ smaller than 8.5%. Yang and Zhang [79] placed a criterion that simple solid solutions form when $\Omega < 1.1$ and $\delta < 6.6\%$. The value of Ω is calculated by the formula $\Omega = \frac{T_m \Delta S_{mix}}{\Delta H_{mix}}$.

Machine learning, a robust computational technique, is considered to start a revolution in material informatics potentially. In the past few years, machine learning (ML) have attracted intensive research attention in materials science community, such as mechanical properties prediction [80–84], phases identification [56,85–87,56,88–93], thermal conductivity [94–97], guiding the synthesis route [98–100] etc., due to its high efficiency and favorable prediction accuracy. As compared to the conventional materials simulation methods, machine learning can quickly review large volumes of data and discover specific trends and patterns and substantially reduce the computational costs and development cycle.

Thus far, ML has been used in the phase prediction of HEAs [56,89,90,92,93]. For example, in Islam and his colleagues' recent work [85], a backpropagation neural network (BP-ANN) model was constructed to distinguish single-phase-SS, amorphous (AM), and Intermetallic (IM) compound of HEAs based on 118 as-cast HEAs. ΔS_{mix} , ΔH_{mix} , VEC, δ and $\Delta\chi$ were selected as the input features to predict the phases of HEAs. An average accuracy of 83.0% was achieved. Pei et al. [101] adopted the Gaussian process classification (GPC) algorithm to screen the most important elemental properties related to the formation of solid solution (SS). Molar volume (V_m), bulk modulus (K), and melting temperature (T_m) were identified as key features, and further used to form a new physical rule ($\gamma \geq 1$) of predicting the solid solution formation of HEAs. Although many works have been conducted in the phase

prediction of HEAs, the amounts of the training samples in most of the dataset are fairly insufficient. It's believed that the size and quality of the dataset would significantly affect the performance of the ML models, even more important than the ML algorithms themselves [102]. In addition, most of the results are only validated by the ML test in the previous work. While the blind test (the testing data are not involved in the training process) can provide impartial validation, it still lacks the new results to verify the model's reliability.

In addition, the microstructure of the HEAs greatly affect the properties. Firstly, the entropy hypothesis indicates that alloys comprising five or more elements would form single-phase solid solutions. Nevertheless, HEAs often contain multiple phases when testing at room temperature. In particular, many of the second phases are intermetallics that can be brittle and hinder mechanical performance. In addition, Single-phase alloys enable the systematical investigation of correlations between varies number, types and concentrations of alloying elements and the physical and mechanical properties. It helps to uncover fundamental mechanisms without the confounding effects of secondary phases. [29] Thus, it is significant to develop and study single-phase SS HEAs for the discovery of novel metallic structural materials.

Herein, our overarching goal is to discover new high-temperature oxidation-resistant RHEAs retaining single-phase-SS with the assistance of the ML. We built up an extensive dataset (total of 1807 samples) by summarizing published results to improve ML models' performance. Gradient boosting (GB) model shows the best performance among the trained models, affording the highest prediction accuracy of over 96% tested on a more extensive subset. Then, the GB model was employed to identify attractive compositions for further studies fast. Here, approximately 100 intrinsic single-phase-SS RHEAs were screened out from the composition space of eight metallic elements, including W, Mo, Ta, Nb, V, Ti, Cr and Al, to design the novel oxidation-resistant RHEAs. Ten out of these newly predicted RHEAs were successfully synthesized via mechanical alloying to verify the model's feasibility.

2. Methods

2.1. Machine learning

We summarized a dataset from the historical literatures. The collected dataset contains 1807 entries, consisting of 800 single-phase-SS alloys and 1000 non-single-phase-SS alloys. The single-phase-SS alloys are labeled as "1", while non-single-phase SS alloys are labeled as "0". Eight features (descriptors), including empirical parameters and weight average elemental properties of the alloys, are utilized (See Table 1). The following equations are applied to obtain the numerical values of the eight features:

$$VEC = \sum_{i=1}^n c_i VEC_i, \quad (1)$$

$$K = \sum_{i=1}^n c_i K_i, \quad (2)$$

Table 1
Description of the features.

Symbol	Description of Feature
δ	Atomic Size Difference
ΔH_{mix}	Mixing Enthalpy
ΔS_{mix}	Mixing Entropy
$\Delta\chi$	Pauli Electronegativity Difference
V_m	Molar Volume
K	Bulk Modulus
T_m	Melting Temperature
VEC	Valence Electron Concentration

$$T_m = \sum_{i=1}^n c_i T_i, \quad (3)$$

$$V_m = \sum_{i=1}^n c_i V_i, \quad (4)$$

$$\Delta H_{mix} = 4 \sum_{i=1, j < i}^n H_{ij} c_i c_{ij}, \quad (5)$$

$$\Delta S_{mix} = R \sum_{i=1}^n c_i \ln(c_i), \quad (6)$$

$$\Delta \chi = \sqrt{\sum_{i=1}^n c_i (\chi_i - \bar{\chi})^2}, \quad (7)$$

$$\delta = 100 \times \sqrt{\sum_{i=1}^n c_i (1 - \frac{r_i}{\bar{r}})^2}, \quad (8)$$

where VEC_i, K_i, T_m , and V_m represent the valence electron, bulk modulus, melting temperature, and molar volume of the i_{th} element, respectively. c_i refers to the atomic concentrations of the i_{th} element, and n is the total number of elemental species in a HEA. χ_i and r_i denote the Pauling electronegativity and radius of i_{th} element, respectively. The $\bar{\chi}$ and \bar{r} are the weight-averaged Pauling electronegativity and weight-averaged atomic radius, which are calculated as $\bar{\chi} = \sum_{i=1}^n c_i \chi_i$ and $\bar{r} = \sum_{i=1}^n c_i r_i$, respectively. The VEC_i, K_i, T_m , and V_m, χ_i and r_i are also taken from ref

[103]. R in Eq. (6) is the ideal gas constant. H_{ij} is the enthalpy of atomic pairs of the i_{th} and j_{th} elements computed with the Miedema method by Takeuchi and Inoue [104].

Each feature is normalized via the Eq. (9) to make the value of all features in the comparable magnitude,

$$X_i^{norm} = \frac{X_i - X_i^{min}}{X_i^{max} - X_i^{min}} \quad (9)$$

where X_i^{max} and X_i^{min} are the maximum and minimum values of X_i , respectively.

Before feeding the data to the ML process, we performed a statistical data analysis on the dataset as shown in Fig. 1. The diagonal panels show that the single-phase formation of the HEAs depends on more than one input features. In other words, this means that a clear boundary of the single-phase and non-single-phase cannot be drawn using a single feature. Therefore, multiple features out of the eight features would be involved in the ML process.

Five representative entries in Table 2 are shown to describe the characteristics of the dataset. The dataset consists of alloy information, eight input features, and the classification labels.

After the dataset was established, we constructed several models using different ML algorithms. In the course of this process, our original data were randomly split into two subsets for training and testing. The training subset contains 1440 alloys (80%), and the remaining data (20%) belong to the testing subset. Additionally, ten-fold cross-validation (CV) was employed to evaluate the trained models. In order to quantify the quality of predictions, we calculate the accuracy score, F1 score, precision score, and recall score, respectively. In this case, the

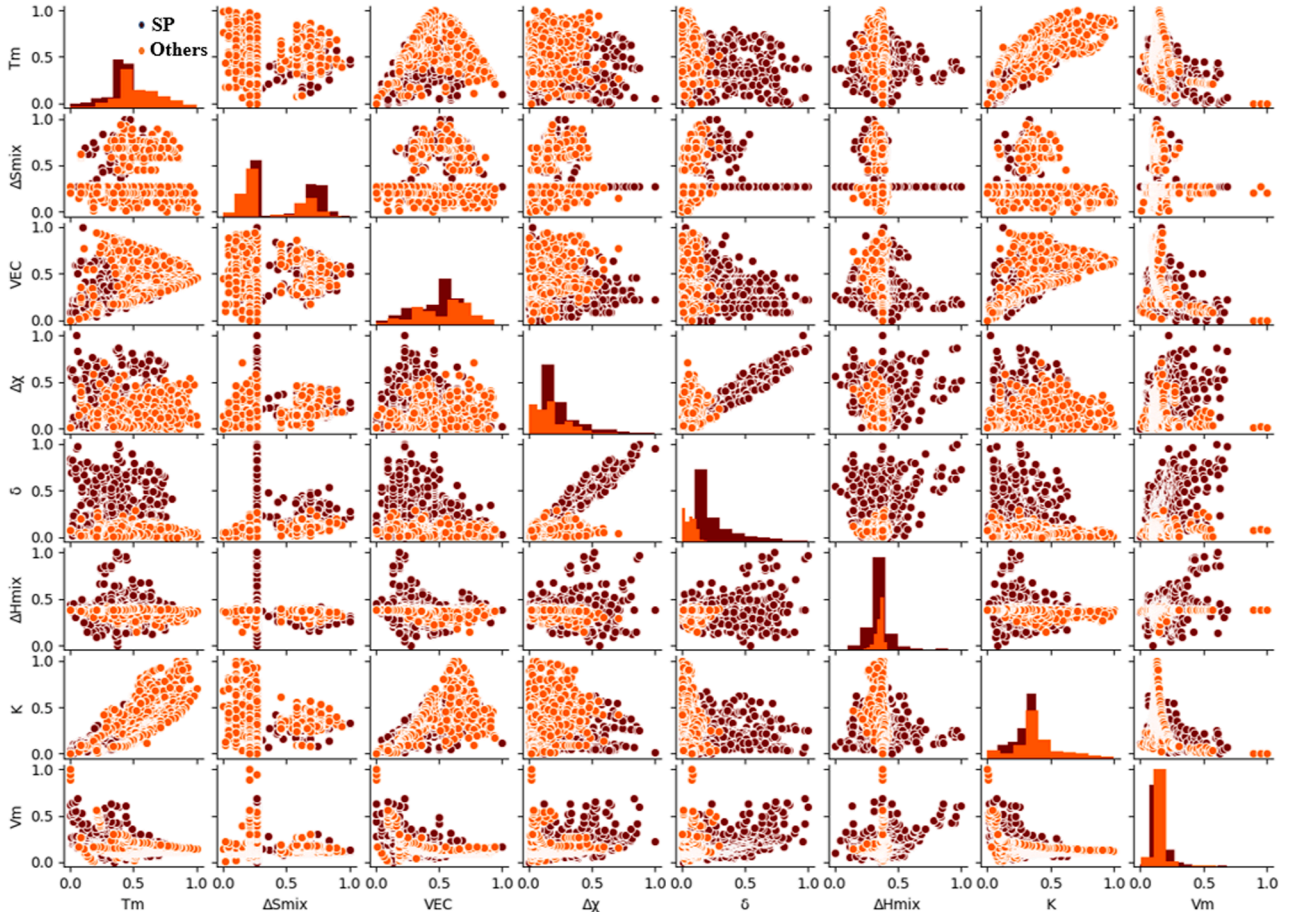


Fig. 1. Diagonal panels: histograms displaying the distributions of the single-phase SS alloy and non-single-phase alloy at different values of one of the eight features. Off-diagonal panels: scatter plots exhibiting the distributions of the two kinds of alloys (single-phase v.s. non-single-phase) as a function of a pair of the eight features.

Table 2

Five representative entries of the dataset after normalization. Each entry has eight input features; The single-phase is denoted as “Class 1”, while non-single-phase is denoted as “Class 0”.

Alloy	Class	T_m	ΔS_{mix}	VEC	$\Delta\chi$	δ	ΔH_{mix}	K	V_m
AlCoCrCu0.5FeNi	1	0.40	0.76	0.60	0.15	0.12	0.34	0.34	0.11
Co1Cu1Fe1Mn1Ni1	1	0.40	0.69	0.73	0.17	0.07	0.39	0.35	0.10
Co1Cr1Mn1Ni1	1	0.44	0.58	0.64	0.19	0.07	0.35	0.35	0.10
Cr1Nb1Ti1V1Zr1	0	0.57	0.69	0.35	0.15	0.16	0.36	0.30	0.15
Mo1Nb1Ti1V2Zr1	0	0.61	0.66	0.35	0.32	0.33	0.37	0.34	0.15

recall represents the fraction of correctly predicted true single-phase SS samples in the total real single-phase SS samples. Simultaneously, precision intuitively indicates the correctly predicted real single-phase-SS samples ratio to the total predicted single-phase SS samples [105,106]. The F1 score indicates the weighted average of the precision and recall [107]. The accuracy denotes the ability of the model to label the samples [102] correctly.

Removal of the irrelevant and redundant features is a well-accepted way to reduce the computation time and improve model robustness. Therefore, we iterate through all feature combinations and then train ML models with each combination, to achieve this goal. Nine different machine learning algorithms, including Gaussian Process (GP), Gradient Boosting (GB), Decision Tree (DT), Random Forest (RF), Support Vector Classifier (SVC), Artificial Neural Network (ANN), Logistic Regression (LR), Gaussian Naïve Bayes (GNB), and K-Nearest Neighbor (KNN), are selected as candidate algorithms for predicting the SP HEAs. Although Logistic Regression is named as regression algorithm, it is also applied for the binary classification problem. Logistic regression is characterized by a logistic function to model the conditional probability of the label Y versus variable X $P(Y|X)$. In this work, label Y takes the single-phase SS or non-single-phase SS. X is the input eight features. All the algorithms are utilized by making use of scikit-learning package in the python program. To prevent the models from overfitting, we implement ten-fold grid-search CV method to tune the hyper-parameters. Herein, we valued their performance to find the one with the highest accuracy. Finally, the top performer is identified based on CV scores and testing accuracy and used to predict the new elemental compositions.

In the perspective of feature importance, permutation importance method is employed to evaluate the relative importance of each feature. The process of permutation importance is as follows: (1) Select the best GB model. (2) Shuffle the values in a single column (input feature) of the dataset and keep other columns unchanged, then feed the shuffled dataset to the selected model to make prediction. Then estimate the degree of the model's performance deterioration to judge the importance of the shuffled feature. (3) Return to the original data, repeat the step (2) with the next column (input feature) to estimate the degree of the model's performance deterioration by shuffling the 2nd input feature. (4) The step (2) will be repeated many times until the importance of each input feature has been evaluated. (5) The degree of model's performance deterioration by shuffling each feature will be compared to ordering the importance of the input feature.

2.2. Experiments

In this study, the alloyed powders of 10 ML-predicted RHEAs were prepared by high-energy ball milling (HEBM). High purity raw W (99.9%, 1–5 μm), Mo (99.98%, 1–2 μm), Ta (99.8%, 1–5 μm), Nb (99.8%, 1–5 μm), V (99.8%, –325 mesh), and Cr (99.5%, 10 μm) powders were purchased from Atlantic Equipment Engineers. Al (99.5%, –325 mesh), and Ti (99.5%, –325 mesh) powders were purchased from Fisher Scientific. The equi-molar amounts of metallic powders were mixed and mechanically alloyed via HEBM for six hours with 500 revolutions per minute (RPM).

The phase of the as-synthesized RHEAs powders was characterized using X-ray Diffraction (XRD) instrument (Bruker D2 Phaser) with Cu K α

radiation.

3. Results and discussion

3.1. Machine learning prediction

Fig. 2 shows the flow chart of the design strategy of single-phase RHEAs with the assistance of ML models. The experimental dataset is established to be fed in the training models of various ML algorithms. Then newly predicted RHEAs are selected for the subsequent experimental validation. The best performer was identified based on average CV accuracy and testing accuracy through the iteration loop of the ML algorithms and feature combinations [108–107]. The average CV precision, recall and F1 scores were regarded as additional indicators in evaluating the prediction performance as well. Here, the average accuracy intuitively represents the ability of the model to correctly label the phases. In contrast, the average precision is the model's capability to label the real single-phase as single-phase. Generally, the higher the average CV accuracy and testing accuracy of the ML model, the better the model is.

Fig. 3 shows the average CV scores and testing accuracy for the best model from each algorithm. The average CV accuracy agrees well with the testing accuracy, indicating that no overfitting exists among all the models. As seen in Fig. 3 and Table.2, the average CV accuracy ranges from 88.93% to 97.37%, while testing accuracy is in the range of 88.4 % – 96.41%. Among them, the GB model shows the best performance with the highest prediction CV accuracy score of 97.37% and testing accuracy value of 96.41%.

GNB model based on feature T_m and δ exhibits the lowest average CV accuracy of 88.93% and testing accuracy 88.40%. GNB model, as a simplified algorithm, tends to be affected by the assumption of conditional independence between every pair of features [110]. More importantly, the GNB model is based on the Gaussian distribution of each feature value and use the summation of each feature's likelihood as score to classify the labels. Fig. 1 shows the distribution of features, i.e. V_m , K, ΔH_{mix} , are significantly overlapped for each other, while T_m and δ are relatively separated. This may be the reason why the GNB model trained by feature T_m and δ exhibits a highest performance. On the other hand, our work in turn confirmed that GNB is a type of simplified algorithm to pursue fast computation rather than accuracy comparing to other more sophisticated methods.

LR model based on feature subset [T_m , $\Delta\chi$, δ , and ΔH_{mix}] achieved an average CV accuracy of 90.24% and testing accuracy of 89.78%. The LR model requires some preconditions such as linear relationship of the coefficient and multicollinearity, which may lead to a relatively lower accuracy for the classification problem (single-phase-SS vs non-single-phase-SS) in present work [111].

As can be seen, the KNN model trained by the feature set [T_m , $\Delta\chi$, δ , and ΔH_{mix}] acquired an average CV accuracy of 95.23% and testing accuracy of 95.30%. The best KNN model has a lower k value of 3. That means the first three closest training samples to the query (testing) sample are chosen to define its label.

Generally, SVC is a well-accepted algorithm for binary classification problems. The prerequisite for training a successful SVC model is the rich experience of appropriately tuning the hyperparameters [106].

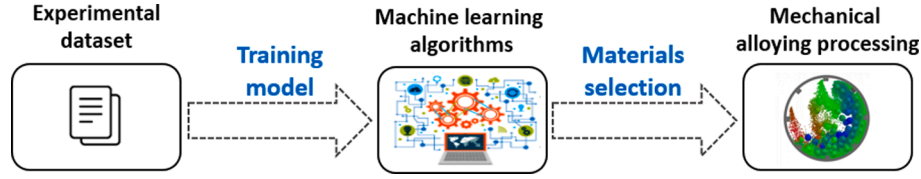


Fig. 2. The flow chart showing the materials design process in this work. The dataset is built up from the published experimental results; Then the dataset is used for machine learning model training; The optimized the machine learning model is applied to predict new single-phase refractory high entropy alloys; Mechanical alloying processing is conducted for experimental validation.

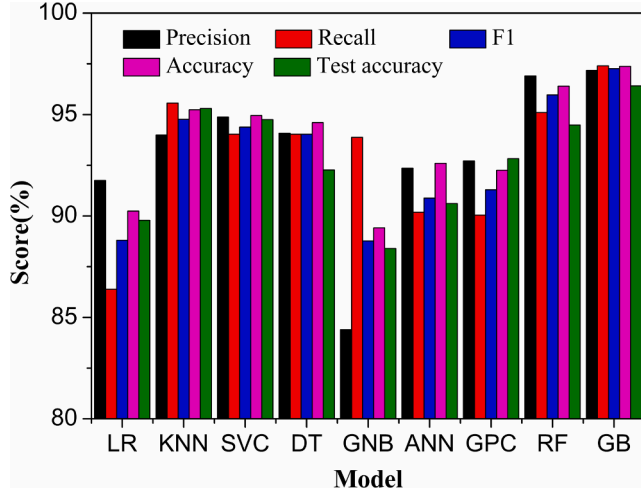


Fig. 3. The ten-fold CV scores (Precision, Recall, Accuracy, and F1) and test accuracy of Logistic Regression (LR), K-Nearest Neighbor (KNN), Support Vector Classifier (SVC), Decision Tree (DT), Gaussian Naïve Bayes (GNB), Artificial Neural Network (ANN), Gaussian Process Classification (GPC), Random Forest (RF), and Gradient Boosting (GB) machine learning models.

Based on the feature space [T_m , ΔS_{mix} , VEC , $\Delta\chi$, δ , ΔH_{mix} , and V_m], the SVC model with the radial-basis function (rbf) kernel reached an average CV accuracy of 95.57% and testing accuracy of 94.75%, which has a significant improvement as compared to previous works [86,109,112]. The performance improvement can be attributed to the extended dataset (1807 samples) and the optimal feature selection by traversing all the possible feature combinations.

Our GPC model with Matérn kernel, trained by feature subset [ΔS_{mix} , $\Delta\chi$ and δ], obtained a decent average CV accuracy of 92.25 % and testing accuracy of 92.82%, which is comparable to the average CV accuracy obtained by Pei et al. [101]. The size of the dataset used by Pei et al. is also fairly large, with 1252 samples. However, our model used thermodynamic parameters and elemental properties as input features instead of the pure elemental properties (V_m , K , T_m , and VEC) used in their work.

Regarding the ANN model, we only set one hidden layer in the network to reach a trade-off between the accuracy and computation time. We constantly tuned hyper-parameters and the neuron number of the hidden layer to seek optimal accuracy. Finally, the ANN model with 80 neurons achieves an average CV accuracy of 92.59% and testing accuracy of 90.61%, which was based on the feature set [T_m , ΔS_{mix} , VEC , $\Delta\chi$, and δ]. The ANN model does not show superior performance among the models we employed. In addition, the ANN model is the most time-consuming model compared to other models in current work.

When it comes to the RF model and GB model, they have been proved to be effective in processing high-dimensional data. The best RF model in the present work achieved an average CV accuracy of 95.57 % and testing accuracy of 94.75%, which is trained by the feature set [T_m , ΔH_{mix} , ΔS_{mix} , $\Delta\chi$, and δ]. The RF model shows a more powerful classification ability than the best DT model, achieving 94.60 % of an average

CV accuracy and 92.27% of testing accuracy. As an ensemble of multiple decision trees (DT), the RF model equally combines the final majority vote of each decision tree trained by the randomly created ‘bootstrapped’ datasets to generate a final decision. The results also agree well with the well-accepted conclusion that the RF model’s accuracy tends to better than the DT model. The RF model in our case achieved higher accuracy in phase prediction of HEAs than the previous studies [107,113,114]. Except for the extended dataset and the optimal model employed in this work, the binary classification problem (single-phase-SS vs. non-single-phase-SS) makes the ML more probably obtaining higher accuracy. Although the GB model is an ensemble of multiple decision trees (DT) as well, in contrast to the RF model, the GB model builds each decision tree sequentially based on the newly created dataset updated by the previous tree’s errors, which iteratively reduce the error of the established trees. Generally, the GB algorithm is one of the robust ML algorithms widely used to achieve satisfactory prediction results [115,116]. As shown in Table 3, the GB model trained by the dataset with the top 7 features [T_m , $\Delta\chi$, ΔS_{mix} , VEC , δ , V_m and ΔH_{mix}] achieved the highest average CV accuracy of 97.37% among the nine ML models. Therefore, GB model is selected for the subsequent analysis due to the best performance among all the models employed.

The performance of the ML model can be further improved by tuning hyper-parameters. For GB models, we calculated the models’ accuracy that varies with `n_estimators` and `max_depth` by fixing other hyper-parameters, as shown in Fig. 4. It is shown that the best GB model has a `n_estimators` value of 100 and `max_depth` value of 5. The `n_estimators`

Table 3
Ten-fold CV performance of each model.

Model	Cross validation				Test accuracy	Features
	Precision	Recall	F1	Accuracy		
LR	91.75 ± 4.1	86.38 ± 6.58	88.80 ± 3.90	90.24 ± 3.24	89.78	T_m , ΔH_{mix} , $\Delta\chi$ and δ
KNN	93.99 ± 2.20	95.56 ± 2.69	94.76 ± 2.17	95.23 ± 1.97	95.30	T_m , $\Delta\chi$, δ , and ΔH_{mix}
SVC	94.88 ± 3.15	94.03 ± 3.52	94.38 ± 2.11	94.95 ± 1.91	94.75	T_m , ΔS_{mix} , VEC , $\Delta\chi$, δ , ΔH_{mix} and V_m
DT	94.08 ± 2.78	94.03 ± 2.10	94.03 ± 2.07	94.60 ± 1.88	92.27	T_m , $\Delta\chi$, δ , and ΔH_{mix}
GNB	84.39 ± 3.32	93.88 ± 3.41	88.77 ± 2.33	89.41 ± 1.58	88.40	T_m and δ
ANN	92.35 ± 3.74	90.18 ± 4.99	90.88 ± 3.18	92.59 ± 2.37	90.61	T_m , ΔS_{mix} , VEC , $\Delta\chi$, and δ
GPC	92.71 ± 2.60	90.04 ± 3.51	91.29 ± 1.70	92.25 ± 1.46	92.82	ΔS_{mix} , $\Delta\chi$ and δ
RF	96.90 ± 2.33	95.10 ± 3.13	95.97 ± 2.40	96.40 ± 2.13	94.48	T_m , ΔS_{mix} , ΔH_{mix} , $\Delta\chi$, and δ
GB	97.17 ± 2.25	97.40 ± 1.55	97.26 ± 1.27	97.37 ± 1.23	96.41	T_m , $\Delta\chi$, ΔS_{mix} , VEC , δ , V_m and ΔH_{mix}

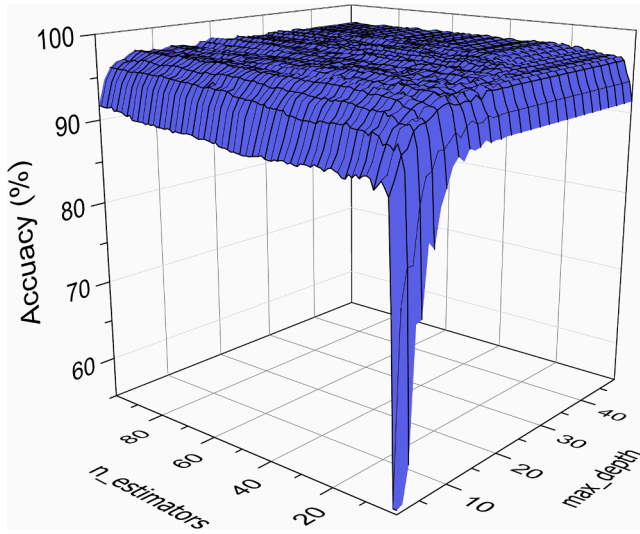


Fig. 4. Gradient Boosting (GB) model's accuracy as a function of $n_estimators$ and max_depth hyperparameters.

denote the number of boosting stages, while the max_depth indicates the maximum number of nodes in each tree. Usually, a larger $n_estimators$ number results in better performance of the GB model.

The prediction accuracies of each model trained with their optimal feature combination are shown in Table 3. The GB model was trained by the largest feature space (7 features). This is consistent with the well-accepted conclusion that the GB model is appropriate to process a dataset with high dimensionality. Furthermore, Fig. 5 depicts the average CV scores (F1, Recall, Precision, and Accuracy) and test accuracy of GB model with the best performance at various given number of input features. According to the results in Fig. 5, the importance order of the input features is also evaluated based on the assumption that the most important feature exists in the feature group that can train the model with the best performance. The screening steps are stated as follows, (1) The subsets only containing one feature is used to train GB models. The most important feature is the subset that can train the GB model with the best performance. For example, When the dataset contains only one feature, the GB model trained by δ achieved the highest CV accuracy of 87.82 % and test accuracy of 88.12 %. δ -based model is

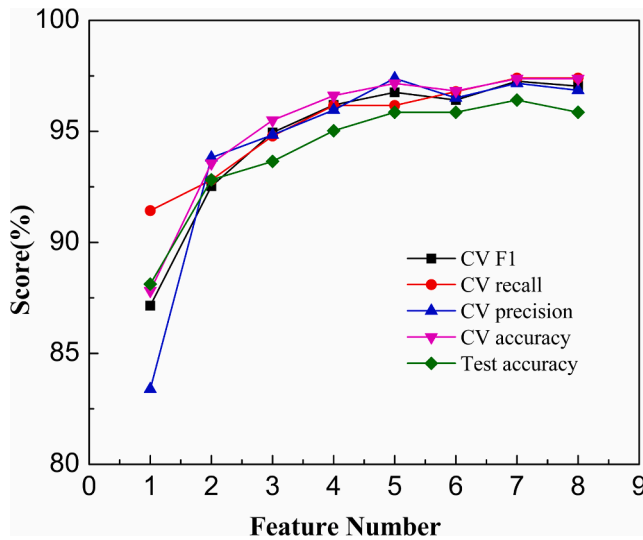


Fig. 5. Average CV scores (F1, Recall, Precision, and Accuracy) and test accuracy of the Gradient Boosting (GB) model with the best performance at various given number of input features.

the best one feature-based GB models, thus indicating δ is the most important feature. Thus, we can gain the 1st important feature from the one-feature subset training. (2) The two-feature subsets are used for GB model training. All the two-feature subsets contain δ (the 1st important feature screened from step (1)) and another feature. Also, the 2nd important feature can be obtained from the two-feature subset that can achieve the best performance of GB model. For instance, subset with δ and T_m has the best two-feature trained GB model, indicating the T_m is the 2nd important feature. (3) The step (1) will be repeated for many times by adding one feature to the subset each time until the importance order of all the features is determined. On this basis, it's confirmed that the feature importance in descending order is δ , T_m , $\Delta\chi$, ΔH_{mix} , ΔS_{mix} , VEC, V_m , and K. Additionally, when the training dataset contains four features δ , $\Delta\chi$, T_m , and ΔH_{mix} , GB model achieved an average CV accuracy of 96.61% and test accuracy of 95.03 %. Then, the performance of the GB model becomes stable with the further increase of feature number.

Fig. 6 shows the importance order of the features through permutation feature importance based on the best GB model. The principle of the permutation method is illustrated in "Methods" section. The returned raw permutation importance scores is an nd_array ($n_features$, $n_repeats$). $n_repeats$ represents number of times to permute a feature. $n_features$ represents the number of the features. In our work, we set the $n_repeats$ as 8, and the dataset has 7 features. Thus, the returned raw permutation importance scores is an 8×7 array. The box plot is employed to show the permutation importance scores. A box plot shows a five-number summary of a set of data. The five-number summary includes the minimum, first quartile, median, third quartile, and maximum. In a box plot, the box was drawn from the first quartile to the third quartile. The vertical orange line going through the box indicates the median score, while the whiskers show the minimum and maximum scores. The circles shown in Fig. 6 represent the outlier in each column (feature) when using Box plot. Thus, ΔS_{mix} , ΔH_{mix} and K columns have one outlier, respectively. Permutation feature importance measures each feature's importance through the prediction error increment of the model trained by the dataset that has the feature's values permuted. Regarding the ranking position of δ , $\Delta\chi$, T_m and ΔH_{mix} , this method got the same result as derived from Fig. 5, indicating that δ , $\Delta\chi$, T_m , and ΔH_{mix} play a crucial role in the phase formation of HEAs. However, the importance order results of VEC, ΔS_{mix} , V_m , and K shows a discrepancy between two methods. When comes to the basis of these two methods, the discrepancy tends to be caused by the different ML models applied. For the feature importance order obtained from Fig. 5 (δ , T_m , $\Delta\chi$, ΔH_{mix} , ΔS_{mix} , VEC, V_m and K), it is evaluated based on the feature combination at different given feature number. In other word, the

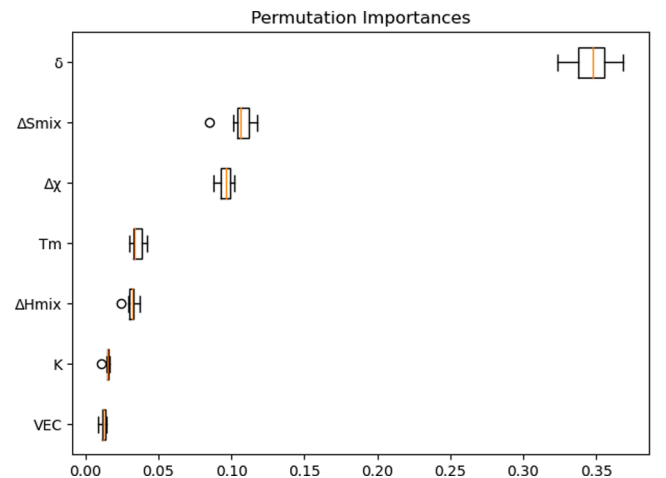


Fig. 6. Importance scores of descriptors (T_m , $\Delta\chi$, ΔS_{mix} , VEC, δ , V_m and ΔH_{mix}) derived from permutation importance method.

feature importance order is obtained on the basis of several different GB models, which are trained by different database (different feature groups). In contrast, the feature importance derived from permutation importance method (δ , ΔS_{mix} , $\Delta\chi$, T_m , ΔH_{mix} , K , and VEC) is evaluated using only one specific model (the best GB model in this work), which is trained by the features δ , T_m , $\Delta\chi$, ΔH_{mix} , ΔS_{mix} , VEC and K . Additionally, each subset contains one shuffled feature during the evaluation. That means the two methods used completely different models and datasets (derived from the same original database). Among the features, δ , representing the scale of lattice strain due to different atomic radius, is the most crucial feature for single-phase formation. This is consistent with the conclusion pointed by Huang et al. that δ is the critical parameter for phase selection [86]. Hume-Rothery rules suggest that the elements with similar electronegativity tend to limit the formation of intermetallic compounds (IMs) [117,118]. And our work also indicates that $\Delta\chi$ is highly related to the formation of single-phase HEAs. Mathematically speaking, this is because their influences on the output property are very close. In contrast to the previous observations [85,119], our study shows that VEC has the least effect on single-phase formation in HEAs. This is because VEC is highly related to the crystal structure [120], while our study only focuses on the single-phase-SS formation rather than detailed crystal structure. In addition, T_m and ΔH_{mix} are the secondary important features for the single-phase-SS formation of HEAs. This agrees well with the simple solid solution formation criterion Ω placed by Yang and Zhang [79], which is the combination of T_m , ΔH_{mix} , and ΔS_{mix} .

Overall, the combination of the GB algorithm and features [T_m , $\Delta\chi$, ΔS_{mix} , VEC , δ , V_m and ΔH_{mix}] is the best strategy to screen out single-phase-SS RHEAs in the present work. Besides, atomic size difference (δ) is the most critical feature for the single-phase-SS formation in RHEAs.

3.2. Experimental validation

Given by the GB models, ~100 new single-phase-SS quaternary and quinary RHEAs were screened out. The compositions of the new RHEAs are created based on the oxidation-resistance property consideration. The quaternary RHEAs include at least two principal elements out of five refractory elements (V, Nb, Mo, Ta, and W) and at least one principal element out of three oxidation-resistant elements (Al, Ti, and Cr). And the quinary RHEAs include at least three principal elements out of five refractory elements (V, Nb, Mo, Ta, and W) and at least one principal element out of three oxidation-resistant elements (Al, Ti, and Cr). The refractory elements can improve the high-temperature mechanical properties, while the Al, Ti, and Cr elements tend to enhance the oxidation-resistance ability of the RHEAs at elevated temperatures due to the formation of the protective oxide layer on the surface. Moreover, the addition of Al or Ti element enables a reduction of the alloys' density. The density (ρ) of these RHEAs can be estimated by

$$\rho = \sum_{i=1}^n c_i \rho_i \quad (10)$$

Table 4

The predicted and experimental phase of the new RHEAs.

RHEAs	T_m	ΔS_{mix}	VEC	$\Delta\chi$	δ	ΔH_{mix}	V_m	ρ	Prediction	Experimental
WMoAlTiCr	2056.06	13.38	5.00	0.33	5.34	-10.08	15.54	8.8	SP	SP(BCC)
WMoNbAlTi	2170.06	13.38	4.80	0.34	2.76	-12.32	16.72	9.08	SP	SP(BCC)
WMoVNBAl	2218.46	13.38	5.00	0.32	3.18	-9.12	15.96	9.39	SP	SP(BCC)
WMoVNBTi	2420.00	13.38	5.20	0.34	3.73	-4.16	16.16	9.76	SP	SP(BCC)
WMoAlTi	2093.33	11.53	4.75	0.35	3.01	-11.75	16.40	9.20	SP	SP(BCC)
WNbAlTi	2056.83	11.53	4.50	0.34	2.43	-15.50	17.03	8.79	SP	SP(BCC)
WNbTiCr	2368.50	11.53	5.25	0.33	5.90	-6.25	15.90	9.91	SP	SP(BCC)
MoNbVAl	1917.58	11.53	4.75	0.24	3.5	-11.5	15.97	6.28	SP	SP(BCC)
MoNbAlTi	1857.08	11.53	4.50	0.25	2.54	-15.25	16.93	6.51	SP	SP(BCC)
MoNbTiCr	2168.75	11.53	5.25	0.25	5.92	-5.50	15.80	7.63	SP	SP(BCC)

where ρ_i is the density of the i_{th} element.

We selected ten newly predicted single-phase-SS RHEAs for fabrication by mechanical alloying. The values of some features for these RHEAs are listed in Table 4. The phases of newly-fabricated RHEAs are determined by X-ray diffraction (XRD). Fig. 7 depicts the XRD patterns of the as-milled RHEAs powders. It is shown that all the samples are single-phase-SS with BCC structure. The experimental results agree well with prediction results, indicating that our model is reliable for the single-phase selection of the RHEAs. The broad peaks result from severe lattice strain and nanocrystalline effect induced by mechanical alloying. Although the alloyed powders after mechanical alloying are in the far-from equilibrium state, the results indicate the capability of the single-phase solid solution formation via this processing method. Also, WC contaminations cannot be entirely avoided during the ball milling process, as shown in the XRD patterns, although it will not affect the phase determinations in the present work.

4. Conclusions

In this work, we combined machine learning prediction and experimental validation to accelerate the search of single-phase RHEAs with oxidation-resistance ability. The performance of various ML models has been improved evidently by expanding the dataset in this work. Among ML models, the GB model with the feature subset [T_m , $\Delta\chi$, ΔS_{mix} , VEC , δ , V_m and ΔH_{mix}] achieved a highest average CV accuracy of 97.37% and testing accuracy of 96.41%, indicating that GB is the best model for the single-phase-SS and non-single-phase-SS classification problem. In addition, the δ is found to play a leading role in screening out single-phase-SS formation. The GB model successfully predicted over 100 new single-phase-SS RHEAs from the composition space of eight metal

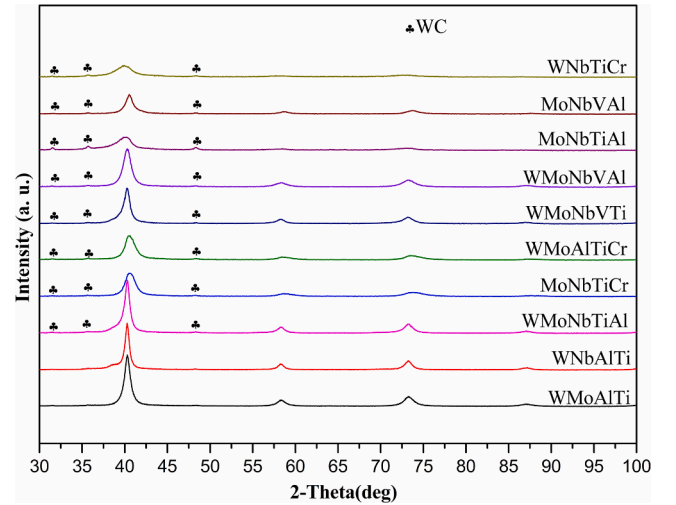


Fig. 7. XRD patterns of ten as-milled refractory high entropy alloys (RHEAs) powders (newly predicted by machine learning model).

elements. Subsequently, 10 of these predicted RHEAs were prepared by mechanical alloying. The XRD patterns showed that all of them are single-phase BCC structures, verifying the great prediction ability of the ML model in the present work. In summary, we successfully designed single-phase oxidation-resistant RHEAs with the aid of ML and experimental validation. Our work would inspire a novel strategy to accelerate the development of new alloys through a combination of the machine learning algorithm and experimental validation.

5. Data availability

The data and methods reported in this paper are available from the corresponding author upon reasonable request.

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.

CRediT authorship contribution statement

Yonggang Yan: Writing - original draft, Writing - review & editing. **Dan Lu:** Writing - review & editing. **Kun Wang:** Conceptualization, Supervision, Writing - review & editing.

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.

Acknowledgement

This work is supported by the Faculty Startup Fund in the New York State College of Ceramics at Alfred University.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.commatsci.2021.110723>.

References

- [1] M. Kutz, Handbook of environmental degradation of materials, William Andrew (2018).
- [2] D.J. Young, High temperature oxidation and corrosion of metals, Elsevier, 2008.
- [3] B.G. Kim, G.M. Kim, C.J. Kim, Oxidation behavior of TiAl-X (X = Cr, V, Si, Mo or Nb) intermetallics at elevated temperature, *Scr. Metall. Mater.* 33 (7) (1995) 1117–1125.
- [4] Z.G. Zhang, X. Teng, Y.L. Mao, C.X. Cao, S.J. Wang, L. Wang, Improvement of Oxidation Resistance of γ -TiAl at 900 and 1000 °C Through Hot-dip Aluminizing, *Oxid. Met.* 73 (3–4) (2010) 455–466.
- [5] M.D.P. Moricca, S.K. Varma, Isothermal oxidation behaviour of Nb–W–Cr Alloys, *Corros. Sci.* 52 (9) (2010) 2964–2972.
- [6] M. Azimovna Azim, S. Burk, B. Gorr, H.-J. Christ, D. Schliephake, M. Heilmaier, R. Bornemann, P.H. Bolívar, Effect of Ti (Macro-) Alloying on the High-Temperature Oxidation Behavior of Ternary Mo–Si–B Alloys at 820–1,300 °C, *Oxid. Met.* 80 (3–4) (2013) 231–242.
- [7] C.-H. Chang, M.S. Titus, J.-W. Yeh, Oxidation Behavior between 700 and 1300 °C of Refractory TiZrNbHfTa High-Entropy Alloys Containing Aluminum, *Adv. Eng. Mater.* 20 (6) (2018) 1700948.
- [8] B. Gorr, F. Mueller, H.-J. Christ, H. Chen, A. Kauffmann, R. Schweiger, D. V. Szabó, M. Heilmaier, Development of Oxidation Resistant Refractory High Entropy Alloys for High Temperature Applications: Recent Results and Development, *Strategy* (2018) 647–659.
- [9] Z. Tong, H. Liu, J. Jiao, W. Zhou, Y. Yang, X. Ren, Laser additive manufacturing of CrMnFeCoNi high entropy alloy: Microstructural evolution, high-temperature oxidation behavior and mechanism, *Opt. Laser Technol.* 130 (2020), 106326.
- [10] S. Wang, Z. Chen, P. Zhang, K. Zhang, C.L. Chen, B.L. Shen, Influence of Al content on high temperature oxidation behavior of AlxCoCrFeNiTi0.5 high entropy alloys, *Vacuum* 163 (2019) 263–268.
- [11] F. Müller, B. Gorr, H.-J. Christ, H. Chen, A. Kauffmann, M. Heilmaier, Effect of microalloying with silicon on high temperature oxidation resistance of novel refractory high-entropy alloy Ta–Mo–Cr–Ti–Al, *Mater. High Temp.* 35 (1–3) (2018) 168–176.
- [12] S. Sheikh, L. Gan, A. Ikeda, H. Murakami, S. Guo, Alloying effect on the oxidation behavior of a ductile Al0.5Cr0.25Nb0.5Ta0.5Ti1.5 refractory high-entropy alloy, *Materials Today, Advances* 7 (2020), 100104.
- [13] A. Duval, F. Miserque, M. Tabarant, J.-P. Nogier, A. Gédéon, Influence of the Oxygen Partial Pressure on the Oxidation of Inconel 617 Alloy at High Temperature, *Oxid. Met.* 74 (5–6) (2010) 215–238.
- [14] P.J. Ennis, W.J. Quadackers, H. Schuster, Effect of selective oxidation of chromium on creep strength of Alloy 617, *Mater. Sci. Technol.* 8 (1) (1992) 78–82.
- [15] K.A. Al-Hatab, M.A. Al-Bukhaiti, U. Krupp, Cyclic oxidation kinetics and oxide scale morphologies developed on alloy 617, *Appl. Surf. Sci.* 318 (2014) 275–279.
- [16] H.-M. Tung, J.F. Stubbins, Incipient oxidation kinetics of alloy 617 and residual stress of the oxide scale formed in air at temperatures between 850 and 1000 °C, *J. Nucl. Mater.* 424 (1–3) (2012) 23–28.
- [17] S. Gao, B. He, L. Zhou, J. Hou, Effects of Ta on the high temperature oxidation behavior of IN617 alloy in air, *Corros. Sci.* 170 (2020), 108682.
- [18] A.J. Ardell, Precipitation hardening, *Metall. Trans. A* 16 (12) (1985) 2131–2165.
- [19] J. Zýka, I. Andršová, B. Podhorná, K. Hrbáček, Mechanical Properties and Microstructure of IN738LC Nickel Superalloy Castings, *Mater. Sci. Forum* 782 (2014) 437–440.
- [20] M. Montazeri, F.M. Ghaini, The liquation cracking behavior of IN738LC superalloy during low power Nd:YAG pulsed laser welding, *Mater. Charact.* 67 (2012) 65–73.
- [21] Y. Danis, C. Arvieu, E. Lacoste, T. Larrouy, J.-M. Quenisset, An investigation on thermal, metallurgical and mechanical states in weld cracking of Inconel 738LC superalloy, *Mater. Des.* 31 (1) (2010) 402–416.
- [22] H.T. Mallikarjuna, N.L. Richards, W.F. Caley, Effect of alloying elements and microstructure on the cyclic oxidation performance of three nickel-based superalloys, *Materialia* 4 (2018) 487–499.
- [23] A. Giorgetti, C. Monti, L. Tognarelli, F. Mastromatteo, Microstructural evolution of René N4 during high temperature creep and aging, *Results Phys.* 7 (2017) 1608–1615.
- [24] C. Sun, M. Kirk, M. Li, K. Hattar, Y. Wang, O. Anderoglu, J. Valdez, B. P. Uberuaga, R. Dickerson, S.A. Maloy, Microstructure, chemistry and mechanical properties of Ni-based superalloy René N4 under irradiation at room temperature, *Acta Mater.* 95 (2015) 357–365.
- [25] M. Bensch, J. Preußner, R. Hüttner, G. Obigodi, S. Virtanen, J. Gabel, U. Glatzel, Modelling and analysis of the oxidation influence on creep behaviour of thin-walled structures of the single-crystal nickel-base superalloy René N5 at 980 °C, *Acta Mater.* 58 (5) (2010) 1607–1617.
- [26] M. Walker, X. Huang, Isothermal Oxidation of René N5 at 1150 °C, (2018).
- [27] H. Yao, L. Yang, Z. Bao, S. Zhu, F. Wang, Low inter-diffusivity γ '-base bondcoats for single crystal superalloy René N5. II: Cyclic oxidation behavior at 1100 °C, *Corrosion Science* 159 (2019) 108127.
- [28] J.W. Yeh, S.K. Chen, S.J. Lin, J.Y. Gan, T.S. Chin, T.T. Shun, C.H. Tsau, S. Y. Chang, Nanostructured High-Entropy Alloys with Multiple Principal Elements: Novel Alloy Design Concepts and Outcomes, *Adv. Eng. Mater.* 6 (5) (2004) 299–303.
- [29] E.P. George, D. Raabe, R.O. Ritchie, High-entropy alloys, *Nat. Rev. Mater.* 4 (8) (2019) 515–534.
- [30] O.N. Senkov, G.B. Wilks, D.B. Miracle, C.P. Chuang, P.K. Liaw, Refractory high-entropy alloys, *Intermetallics* 18 (9) (2010) 1758–1765.
- [31] O.N. Senkov, D.B. Miracle, K.J. Chaput, J.-P. Couzinie, Development and exploration of refractory high entropy alloys—A review, *J. Mater. Res.* 33 (19) (2018) 3092–3128.
- [32] Y. Zhang, T.T. Zuo, Z. Tang, M.C. Gao, K.A. Dahmen, P.K. Liaw, Z.P. Lu, Microstructures and properties of high-entropy alloys, *Prog. Mater. Sci.* 61 (2014) 1–93.
- [33] S. Praveen, H.S. Kim, High-Entropy Alloys: Potential Candidates for High-Temperature Applications - An Overview, *Adv. Eng. Mater.* 20 (1) (2018) 1700645.
- [34] K.-Y. Tsai, M.-H. Tsai, J.-W. Yeh, Sluggish diffusion in Co–Cr–Fe–Mn–Ni high-entropy alloys, *Acta Mater.* 61 (13) (2013) 4887–4897.
- [35] J. Dąbrowa, W. Kucza, G. Cieślak, T. Kulik, M. Danielewski, J.-W. Yeh, Interdiffusion in the FCC-structured Al–Co–Cr–Fe–Ni high entropy alloys: Experimental studies and numerical simulations, *J. Alloy. Compd.* 674 (2016) 455–462.
- [36] M. Vaidya, S. Trubel, B.S. Murty, G. Wilde, S.V. Divinski, Ni tracer diffusion in CoCrFeNi and CoCrFeMnNi high entropy alloys, *J. Alloy. Compd.* 688 (2016) 994–1001.
- [37] Y. Xu, W. Li, L. Qu, X. Yang, B. Song, R. Lupoi, S. Yin, Solid-state cold spraying of FeCoCrNiMn high-entropy alloy: an insight into microstructure evolution and oxidation behavior at 700–900 °C, *J. Mater. Sci. Technol.* 68 (2021) 172–183.
- [38] G. Laplanche, U.F. Volkert, G. Eggeler, E.P. George, Oxidation Behavior of the CrMnFeCoNi High-Entropy Alloy, *Oxid. Met.* 85 (5–6) (2016) 629–645.
- [39] O.N. Senkov, S.V. Senkova, D.M. Dimiduk, C. Woodward, D.B. Miracle, Oxidation behavior of a refractory NbCrMo0.5Ta0.5TiZr alloy, *J. Mater. Sci.* 47 (18) (2012) 6522–6534.
- [40] J. Zheng, X. Hou, X. Wang, Y.e. Meng, X. Zheng, L. Zheng, Isothermal oxidation mechanism of a newly developed Nb–Ti–V–Cr–Al–W–Mo–Hf alloy at 800–1200 °C, *Int. J. Refract. Metal Hard Mater.* 54 (2016) 322–329.
- [41] T.M. Butler, K.J. Chaput, J.R. Dietrich, O.N. Senkov, High temperature oxidation behaviors of equimolar NbTiZrV and NbTiZrCr refractory complex concentrated alloys (RCCAs), *J. Alloy. Compd.* 729 (2017) 1004–1019.

- [42] O.A. Waseem, J. Lee, H.M. Lee, H.J. Ryu, The effect of Ti on the sintering and mechanical properties of refractory high-entropy alloy Ti₅₀W₂₀V₁₀Cr₁₀ fabricated via spark plasma sintering for fusion plasma-facing materials, *Mater. Chem. Phys.* 210 (2018) 87–94.
- [43] L.-C. Li, M.-X. Li, M. Liu, B.-Y. Sun, C. Wang, J.-T. Huo, W.-H. Wang, Y.-H. Liu, Enhanced oxidation resistance of MoTaTiCrAl high entropy alloys by removal of Al, *Science China Materials* 64(1) (2020) 223–231.
- [44] F. Müller, B. Gorr, H.-J. Christ, J. Müller, B. Butz, H. Chen, A. Kauffmann, M. Heilmaier, On the oxidation mechanism of refractory high entropy alloys, *Corros. Sci.* 159 (2019), 108161.
- [45] P. Zhang, Y. Li, Z. Chen, J. Zhang, B. Shen, Oxidation response of a vacuum arc melted NbZrTiCrAl refractory high entropy alloy at 800–1200 °C, *Vacuum* 162 (2019) 20–27.
- [46] L. Rudolf Kanyane, A. Patricia Popoola, N. Malatji, Development of spark plasma sintered TiAlSiMoW multicomponent alloy: Microstructural evolution, corrosion and oxidation resistance, *Results Phys.* 12 (2019) 1754–1761.
- [47] J. Jayaraj, P. Thirathipiwat, J. Han, A. Gebert, Microstructure, mechanical and thermal oxidation behavior of AlNbTiZr high entropy alloy, *Intermetallics* 100 (2018) 9–19.
- [48] Q.-D. Qin, J.-b. Qu, Y.-e. Hu, Y.-J. Wu, X.-D. Su, Microstructural characterization and oxidation resistance of multicomponent equiatomic CoCrCuFeNi–TiO high-entropy alloy, *Int. J. Miner. Metall. Mater.* 25 (11) (2018) 1286–1293.
- [49] B. Gorr, F. Mueller, H.-J. Christ, T. Mueller, H. Chen, A. Kauffmann, M. Heilmaier, High temperature oxidation behavior of an equimolar refractory metal-based alloy 20Nb 20Mo 20Cr 20Ti 20Al with and without Si addition, *J. Alloy. Compd.* 688 (2016) 468–477.
- [50] Y.-K. Kim, Y.-A. Joo, H.S. Kim, K.-A. Lee, High temperature oxidation behavior of Cr-Mn-Fe-Co-Ni high entropy alloy, *Intermetallics* 98 (2018) 45–53.
- [51] Y.Y. Liu, Z. Chen, Y.Z. Chen, J.C. Shi, Z.Y. Wang, S. Wang, F. Liu, Effect of Al content on high temperature oxidation resistance of Al_xCoCrCuFeNi high entropy alloys (x=0, 0.5, 1, 1.5, 2), *Vacuum* 169 (2019) 108837.
- [52] T.M. Butler, J.P. Alfano, R.L. Martens, M.L. Weaver, High-Temperature Oxidation Behavior of Al-Co-Cr-Ni-(Fe or Si) Multicomponent High-Entropy Alloys, *Jom* 67 (1) (2015) 246–259.
- [53] B. Gorr, F. Müller, M. Azim, H.-J. Christ, T. Müller, H. Chen, A. Kauffmann, M. Heilmaier, High-Temperature Oxidation Behavior of Refractory High-Entropy Alloys: Effect of Alloy Composition, *Oxid. Met.* 88 (3–4) (2017) 339–349.
- [54] B. Gorr, M. Azim, H.-J. Christ, T. Mueller, D. Schliephake, M. Heilmaier, Phase equilibria, microstructure, and high temperature oxidation resistance of novel refractory high-entropy alloys, *J. Alloy. Compd.* 624 (2015) 270–278.
- [55] R. Kozak, A. Sologubenko, W. Steurer, Single-phase high-entropy alloys – an overview, *Zeitschrift für Kristallographie - Crystalline Materials* 230(1) (2015).
- [56] J.M. Rickman, H.M. Chan, M.P. Harmer, J.A. Smeltzer, C.J. Marvel, A. Roy, G. Balasubramanian, Materials informatics for the screening of multi-principal elements and high-entropy alloys, *Nat Commun* 10 (1) (2019) 2618.
- [57] F. Tian, L.K. Varga, N. Chen, J. Shen, L. Vitos, Ab initio design of elastically isotropic TiZrNbMoV high-entropy alloys, *J. Alloy. Compd.* 599 (2014) 19–25.
- [58] L.-Y. Tian, G. Wang, J.S. Harris, D.L. Irving, J. Zhao, L. Vitos, Alloying effect on the elastic properties of refractory high-entropy alloys, *Mater. Des.* 114 (2017) 243–252.
- [59] Z. Li, F. Körmann, B. Grabowski, J. Neugebauer, D. Raabe, Ab initio assisted design of quinary dual-phase high-entropy alloys with transformation-induced plasticity, *Acta Mater.* 136 (2017) 262–270.
- [60] J. Ding, Q. Yu, M. Asta, R.O. Ritchie, Tunable stacking fault energies by tailoring local chemical order in CrCoNi medium-entropy alloys, *Proc Natl Acad Sci U S A* 115 (36) (2018) 8919–8924.
- [61] D. Ma, B. Grabowski, F. Körmann, J. Neugebauer, D. Raabe, Ab initio thermodynamics of the CoCrFeMnNi high entropy alloy: Importance of entropy contributions beyond the configurational one, *Acta Mater.* 100 (2015) 90–97.
- [62] C. Niu, A.J. Zaddach, C.C. Koch, D.L. Irving, First principles exploration of near-equiatom NiFeCrCo high entropy alloys, *J. Alloy. Compd.* 672 (2016) 510–520.
- [63] Y. Gao, L. Qiao, D. Wu, Y. Zhang, Y. Zou, First principle calculation of the effect of Cr, Ti content on the properties of VMoNbTaW_xM (M = Cr, Ti) refractory high entropy alloy, *Vacuum* 179 (2020), 109459.
- [64] M. Widom, Modeling the structure and thermodynamics of high-entropy alloys, *J. Mater. Res.* 33 (19) (2018) 2881–2898.
- [65] Y. Qi, X. Chen, M. Feng, Molecular dynamics-based analysis of the effect of voids and HCP-Phase inclusion on deformation of single-crystal CoCrFeMnNi high-entropy alloy, *Mater. Sci. Eng., A* 791 (2020), 139444.
- [66] Q. Fang, Y. Chen, J. Li, C. Jiang, B. Liu, Y. Liu, P.K. Liaw, Probing the phase transformation and dislocation evolution in dual-phase high-entropy alloys, *Int. J. Plast.* 114 (2019) 161–173.
- [67] W.Y. Wang, S.L. Shang, Y. Wang, F. Han, K.A. Darling, Y. Wu, X. Xie, O. N. Senkov, J. Li, X.D. Hui, K.A. Dahmen, P.K. Liaw, L.J. Kecskes, Z.-K. Liu, Atomic and electronic basis for the serrations of refractory high-entropy alloys, *npj Comput. Mater.* 3 (1) (2017).
- [68] O.N. Senkov, C. Zhang, A.L. Pilchak, E.J. Payton, C. Woodward, F. Zhang, CALPHAD-aided development of quaternary multi-principal element refractory alloys based on NbTiZr, *J. Alloy. Compd.* 783 (2019) 729–742.
- [69] O.N. Senkov, J.D. Miller, D.B. Miracle, C. Woodward, Accelerated exploration of multi-principal element alloys for structural applications, *Calphad* 50 (2015) 32–48.
- [70] M.C. Gao, C.S. Carney, Ö.N. Doğan, P.D. Jablonksi, J.A. Hawk, D.E. Alman, Design of Refractory High-Entropy Alloys, *Jom* 67 (11) (2015) 2653–2669.
- [71] F.G. Courty, T. Butler, K. Chaput, A. Saville, J. Copley, J. Foltz, P. Mason, K. Clarke, M. Kaufman, A. Clarke, Phase equilibria, mechanical properties and design of quaternary refractory high entropy alloys, *Mater. Des.* 155 (2018) 244–256.
- [72] F. Zhang, C. Zhang, S.L. Chen, J. Zhu, W.S. Cao, U.R. Kattner, An understanding of high entropy alloys from phase diagram calculations, *Calphad* 45 (2014) 1–10.
- [73] S. Gorsse, O. Senkov, About the Reliability of CALPHAD Predictions in Multicomponent Systems, *Entropy (Basel)* 20 (12) (2018) 899, <https://doi.org/10.3390/e20120899>.
- [74] L. Guo, J. Gu, X. Gong, S. Ni, M. Song, CALPHAD aided design of high entropy alloy to achieve high strength via precipitate strengthening, *Science China Materials* 63(2) (2019) 288–299.
- [75] O.N. Senkov, J.D. Miller, D.B. Miracle, C. Woodward, Accelerated exploration of multi-principal element alloys with solid solution phases, *Nat Commun* 6 (2015) 6529.
- [76] C. Zhang, F. Zhang, S. Chen, W. Cao, Computational Thermodynamics Aided High-Entropy Alloy Design, *Jom* 64 (7) (2012) 839–845.
- [77] Y. Zhang, Y.J. Zhou, J.P. Lin, G.L. Chen, P.K. Liaw, Solid-Solution Phase Formation Rules for Multi-component Alloys, *Adv. Eng. Mater.* 10 (6) (2008) 534–538.
- [78] S. Guo, C.T. Liu, Phase stability in high entropy alloys: Formation of solid-solution phase or amorphous phase, *Progress in Natural Science: Materials International* 21 (6) (2011) 433–446.
- [79] X. Yang, Y. Zhang, Prediction of high-entropy stabilized solid-solution in multi-component alloys, *Mater. Chem. Phys.* 132 (2–3) (2012) 233–238.
- [80] C. Wen, Y. Zhang, C. Wang, D. Xue, Y. Bai, S. Antonov, L. Dai, T. Lookman, Y. Su, Machine learning assisted design of high entropy alloys with desired property, *Acta Mater.* 170 (2019) 109–117.
- [81] C. Zou, J. Li, W.Y. Wang, Y. Zhang, D. Lin, R. Yuan, X. Wang, B. Tang, J. Wang, X. Gao, H. Kou, X. Hui, X. Zeng, M. Qian, H. Song, Z.-K. Liu, D. Xu, Integrating data mining and machine learning to discover high-strength ductile titanium alloys, *Acta Mater.* 202 (2021) 211–221.
- [82] 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.
- [83] Y. Li, Y. Liu, S. Luo, Z. Wang, K. Wang, Z. Huang, H. Zhao, L. Jiang, Neural network model for correlating microstructural features and hardness properties of nickel-based superalloys, *J. Mater. Res. Technol.* 9 (6) (2020) 1467–1477.
- [84] 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.
- [85] N. Islam, W. Huang, H.L. Zhuang, Machine learning for phase selection in multi-principal element alloys, *Comput. Mater. Sci.* 150 (2018) 230–235.
- [86] W. Huang, P. Martin, H.L. Zhuang, Machine-learning phase prediction of high-entropy alloys, *Acta Mater.* 169 (2019) 225–236.
- [87] Z. Qin, Z.i. Wang, Y. Wang, L. Zhang, W. Li, J. Liu, Z. Wang, Z. Li, J. Pan, L. Zhao, F. Liu, L. Tan, J. Wang, H. Han, L. Jiang, Y. Liu, Phase prediction of Ni-base superalloys via high-throughput experiments and machine learning, *Materials Research Letters* 9 (1) (2021) 32–40.
- [88] K. Kaufmann, K.S. Vecchio, Searching for high entropy alloys: A machine learning approach, *Acta Mater.* 198 (2020) 178–222.
- [89] U.K. Jaiswal, Y. Vamsi Krishna, M.R. Rahul, G. Phanikumar, Machine learning-enabled identification of new medium to high entropy alloys with solid solution phases, *Comput. Mater. Sci.* 197 (2021), 110623.
- [90] Y.V. Krishna, U.K. Jaiswal, R.M. R. Machine learning approach to predict new multiphase high entropy alloys, *Scripta Materialia* 197 (2021) 113804.
- [91] J.M. Rickman, T. Lookman, S.V. Kalinin, Materials informatics: From the atomic-level to the continuum, *Acta Mater.* 168 (2019) 473–510.
- [92] S. Risal, W. Zhu, P. Guillen, L. Sun, Improving phase prediction accuracy for high entropy alloys with Machine learning, *Comput. Mater. Sci.* 192 (2021), 110389.
- [93] A. Roy, G. Balasubramanian, Predictive descriptors in machine learning and data-enabled explorations of high-entropy alloys, *Comput. Mater. Sci.* 193 (2021), 110381.
- [94] S. Wu, Y. Kondo, M.-A. Kakimoto, B. Yang, H. Yamada, I. Kuwajima, G. Lambard, K. Hongo, Y. Xu, J. Shiomi, C. Schick, J. Morikawa, R. Yoshida, Machine-learning-assisted discovery of polymers with high thermal conductivity using a molecular design algorithm, *npj Comput. Mater.* 5 (1) (2019).
- [95] Y. Zhang, C. Ling, A strategy to apply machine learning to small datasets in materials science, *npj Comput. Mater.* 4 (1) (2018).
- [96] X. Qian, S. Peng, X. Li, Y. Wei, R. Yang, Thermal conductivity modeling using machine learning potentials: application to crystalline and amorphous silicon, *Materials Today Physics* 10 (2019), 100140.
- [97] J. Wan, J.-W. Jiang, H.S. Park, Machine learning-based design of porous graphene with low thermal conductivity, *Carbon* 157 (2020) 262–269.
- [98] B. Tang, Y. Lu, J. Zhou, T. Chouhan, H. Wang, P. Golani, M. Xu, Q. Xu, C. Guan, Z. Liu, Machine learning-guided synthesis of advanced inorganic materials, *Mater. Today* 41 (2020) 72–80.
- [99] P. Raccuglia, K.C. Elbert, P.D.F. Adler, C. Falk, M.B. Wenny, A. Molloy, M. Zeller, S.A. Friedler, J. Schrier, A.J. Norquist, Machine-learning-assisted materials discovery using failed experiments, *Nature* 533 (7601) (2016) 73–76.
- [100] S. Liu, B.B. Kappes, B. Amin-ahmadi, O. Benafan, X. Zhang, A.P. Stebner, Physics-informed machine learning for composition – process – property design: Shape memory alloy demonstration, *Appl. Mater. Today* 22 (2021) 100898, <https://doi.org/10.1016/j.apmt.2020.100898>.
- [101] Z. Pei, J. Yin, J.A. Hawk, D.E. Alman, M.C. Gao, Machine-learning informed prediction of high-entropy solid solution formation, Beyond the Hume-Rothery rules, *npj Computational Materials* 6 (1) (2020), <https://doi.org/10.1038/s41524-020-0308-7>.

- [102] Y. Yan, D. Lu, K. Wang, Overview: Recent studies of machine learning in phase prediction of High Entropy Alloys, *JOM* (2021).
- [103] Periodic Table (created by Theodore Gray, with assistance from Nick Mann, and in partnership with Max Whitby of RGB Research).
- [104] A. Takeuchi, A. Inoue, Classification of bulk metallic glasses by atomic size difference, heat of mixing and period of constituent elements and its application to characterization of the main alloying element, *Mater. Trans.* 46 (12) (2005) 2817–2829.
- [105] F. Li, J. Han, T. Cao, W. Lam, B. Fan, W. Tang, S. Chen, K.L. Fok, L. Li, Design of self-assembly dipeptide hydrogels and machine learning via their chemical features, *Proc. Natl. Acad. Sci.* 116 (23) (2019) 11259–11264.
- [106] Y. Xie, C. Zhang, X. Hu, C. Zhang, S.P. Kelley, J.L. Atwood, J. Lin, Machine Learning Assisted Synthesis of Metal-Organic Nanocapsules, *J Am Chem Soc* 142 (3) (2020) 1475–1481.
- [107] A. Choudhury, T. Konnur, P.P. Chattopadhyay, S. Pal, Structure prediction of multi-principal element alloys using ensemble learning, *Engineering Computations* 37 (3) (2019) 1003–1022.
- [108] L. Zhang, H. Chen, X. Tao, H. Cai, J. Liu, Y. Ouyang, Q. Peng, Y. Du, Machine learning reveals the importance of the formation enthalpy and atom-size difference in forming phases of high entropy alloys, *Mater. Des.* 193 (2020), 108835.
- [109] Y. Li, W. Guo, Machine-learning model for predicting phase formations of high-entropy alloys, *Physical Review Materials* 3 (9) (2019), <https://doi.org/10.1103/PhysRevMaterials.3.095005>.
- [110] I. Rish, An empirical study of the naive Bayes classifier, *IJCAI 2001 workshop on empirical methods in artificial intelligence*, 2001, pp. 41–46.
- [111] B. Boehmke, B.M. Greenwell, *Hands-on machine learning with R*, CRC Press, 2019.
- [112] N. Qu, Y. Chen, Z. Lai, Y. Liu, J. Zhu, The phase selection via machine learning in high entropy alloys, *Procedia Manuf.* 37 (2019) 299–305.
- [113] R. Machaka, Machine learning-based prediction of phases in high-entropy alloys, *Comput. Mater. Sci.* 188 (2021), 110244.
- [114] J. Qi, A.M. Cheung, S.J. Poon, High Entropy Alloys Mined From Binary Phase Diagrams, *Sci Rep* 9 (1) (2019) 15501.
- [115] A. Natekin, A. Knoll, Gradient boosting machines, a tutorial, *Front. Neurorob.* 7 (2013) 21.
- [116] J.H. Friedman, Greedy function approximation: a gradient boosting machine, *The Annals of Statistics* 29 (5) (2001) 1189–1232.
- [117] U. Mizutani, Hume-Rothery rules for structurally complex alloy phases, *MRS Bulletin* 37(2) (2012) 169–169.
- [118] B.S. Murty, J.-W. Yeh, S. Ranganathan, P. Bhattacharjee, *High-entropy alloys*, Elsevier, 2019.
- [119] 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, *Comput. Mater. Sci.* 175 (2020), 109618.
- [120] 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>.