Summary of Efforts in Phase Prediction of High Entropy Alloys Using Machine Learning

Swati Singh, Shrikrishna Nandkishor Joshi, and Saurav Goel

1 Introduction

Conventional alloys made from iron, aluminium, titanium, magnesium, and many others are prepared by mixing one or sometimes two principal elements with some secondary alloying elements in small proportions. Cantor alloys also recognized now as high-entropy alloys shows a unique characteristic that by mixing at least five metallic elements such that when the mixing percentage vary from 5 to 35% in an equimolar composition produces a single solid-solution phase which is crystalline in nature [1]. By virtue of this, HEAs offer high specific strength [2, 3], excellent stability at high temperatures, exceptional ductility [4], and high corrosion resistance and others [5]. Unlike conventional alloys, HEAs possess a trend of 'stronger being more ductile' [6, 7].

Cantor suggests that as many as ~ 108 varieties of HEAs can be developed using 44–64 elements from the chemical periodic table [8]. Many of these HEAs are yet to be synthesized in the lab and Yeh [9] suggests that "still a lot more treasure exist in non-equimolar HEAs". In developing an alloy, a considerable

S. Singh · S. N. Joshi (✉) · S. Goel
Department of Mechanical Engineering, Indian Institute of Technology Guwahati, Guwahati, Assam, India
e-mail: snj@iitg.ac.in

S. Singh
e-mail: sswati@iitg.ac.in

S. Goel
e-mail: goeLs@Lsbu.ac.uk

S. Goel
School of Engineering, London South Bank University, 103 Borough Road, London SE1 0AA, UK

Department of Mechanical Engineering, University of Petroleum and Energy Studies, Dehradun 248007, India

S. N. Joshi et al. (eds.), *Low Cost Manufacturing Technologies*,
https://doi.org/10.1007/978-981-19-8452-5_4

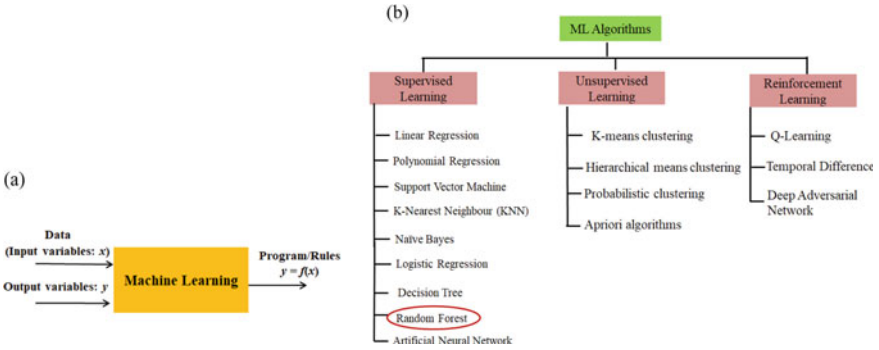

Fig. 1 **a** Illustrative architecture of ML, **b** classification of ML algorithms on the basis of supervised, unsupervised, and reinforcement learning [12–14]. Circle on Random forest is done to highlight that this method was used in this paper to present a case study

number of input parameters such as its composition, synthesis route, processing window, temperature, heating/cooling rate are required to obtain its crystallographic information and mechanical, electrical and functional properties of interest. Thus, relying on traditional laboratory experiments for novel material discovery can be very time-intensive.

Machine learning (ML) has emerged as a sophisticated and reliable technique in replacing repetitive laboratory experiments and computational simulations such as Density Functional Theory (DFT) and Molecular Dynamics (MD) [10]. DFT method predicts material properties using quantum mechanics and can at times be erroneous [11]. MD on the other hand analyse atoms by numerically solving Newton's equations of motion [12]. MD continues to suffer from the drawback on having reliable interatomic potential functions. Thus, ML provides a robust alternative tool based on the reliance of historical data and a mathematic way by pattern recognition technique. This in turn enhances our ability to extract salient features from within the data which are otherwise not readily visible even to an experienced researcher. A summary of various ML algorithms currently being used is shown in Fig. 1a which formed the core of this paper.

1.1 ML Algorithms

ML algorithms are broadly classified into three categories based on their type of learning, namely supervised, unsupervised, and reinforcement learning shown in Fig. 1b [13, 14].

1.1.1 Supervised Learning

As the name suggests, supervised ML algorithms are taught by providing a labelled dataset that includes input and output variables. The objective of supervised learning is to map input (x) and output (y), using a linear or nonlinear function, for instance, $y = f(x)$ [14].

1.1.2 Unsupervised Learning

In unsupervised ML algorithms, unlabelled dataset is used to train an algorithm, where inputs are not labelled with the correct outputs. The goal is to model the underlying structure or to discover the patterns in the data. The unlabelled data in the unsupervised learning is used to train algorithms for clustering and association problems.

1.1.3 Reinforcement Learning

Reinforcement learning is based on reward or punishment methods, where an agent learns to perceive and interpret its complex environment; it takes actions and learns through the trial-and-error method. It is devised to reward the desired behaviour by assigning a positive value to encourage the agent and punishing the undesirable behaviour by assigning negative values to penalize the agent. An agent either gets an award or a penalty based on the actions it performs, and the ultimate goal is to maximize the total reward. Over time, the agent learns to avoid the negative and seek the positive, thus learns the ideal behaviour to optimize its performance.

Current study focuses on a supervised ML algorithm, as the labelled data is employed for phase prediction of HEAs. A detailed description of all supervised ML algorithms has been discussed in Table 1.

1.2 Literature Review

Recently, there has been a surge in the number of publications on phase prediction of HEAs using various ML techniques. Islam et al. [15], Huang et al. [16] and Nassar et al. [17] employed neural networks in their study for phase prediction of HEAs and observed an average accuracy of 83%, 74.3%, and 90%, respectively, by using a relatively smaller number of datasets. These models considered five physical parameters namely mixing entropy (ΔSmix), valence electron concentration (VEC), atomic size difference (δ), mixing enthalpy (ΔHmix), and electronegativity difference ($\Delta\chi$). Choudhury et al. [18], used a random forest regression algorithm for the classification of different phases and crystal structures and obtained an accuracy of

91.66% for the classification of phases and 93.10% for the classification of different crystal structures, respectively.

To significantly improve the phase prediction accuracy, Risal et al. [19], compared Support Vector Machine (SVM), K-Nearest Neighbour (KNN), Random Forest (RF) classifier, and Multi-layer perceptron (MLP). KNN and RF Classifiers performed most effectively and obtained the test accuracy of 92.31% and 91.21%, respectively,

Table 1 A list of important supervised ML algorithms

Algorithms	Category	Description of the method	Advantages	Limitations
Linear regression	Regression	It correlates one independent variable and one dependent variable using a straight line	Easy to implement and understand, and it can extrapolate beyond a specific dataset	Assumes that dependent variables and independent variables are linearly related. Highly prone to noise and sensitive to outliers
Polynomial regression	Regression	Linear regression shows under fitting for nonlinear data, thus a new nth degree polynomial function is used to relate the independent and dependent variables	Capable of accommodating a wide range of functions	A correct polynomial function needs to be selected for better fitting of all the data points. It is more sensitive to outliers; even one or two outliers can significantly affect the outcome
Support vector machine	Regression and classification	It selects a decision boundary that best separates two different groups and predicts whether a new data falls into one category or the other	SVM performs well for high dimensional data as it provides various Kernels such as linear kernel, nonlinear kernel, polynomial kernel, Gaussian, radial basis function (RBF), and sigmoid kernel. It best suits binary classification problems	Performs poorly for overlapped classes. The selection of an appropriate Kernel and hyper-parameter is complex and problematic. A large dataset requires a long time for training

(continued)

Table 1 (continued)

Algorithms	Category	Description of the method	Advantages	Limitations
K-Nearest neighbour	Classification	Depending on the nearest neighbour, it classifies a new sample into one class out of several classes, separated depending on the type of data	Simple and easily accommodates itself when exposed to new data. Multiclass classification can be easily solved	It essentially requires feature scaling, performs poorly on imbalanced data, and cannot handle outliers and missing values
Naïve Bayes	Classification	It is based on 'Bayes theorem', assuming that no features are dependent and each feature is given the same weightage. It can classify data into several categories following the highest possibility	Simple and useful for vast datasets. It is insensitive to irrelevant features. Very fast, scalable, and effective for multiclass classification problems	Its application in the real world is limited because it assumes that all the features are independent and each feature makes an equal contribution
Logistic regression	Regression	An efficient method for binary and linear classification problems, it calculates probability from logistic regression equation to calculate the relationship between input variable and one or more output variables	It is simple, effective, and does not require feature scaling or hyper-parameter tuning	It performs poorly for nonlinear, irrelevant, and highly-correlated data

(continued)

Table 1 (continued)

Algorithms	Category	Description of the method	Advantages	Limitations
Decision tree	Regression and classification	Creates a tree-like structure	It is easy to visualize and interpret the results	It is data-sensitive, which means the outcome might change significantly on changing data slightly, and it is prone to overfitting
Random forest	Regression and classification	It takes votes from various decision trees for classification, and the average of all votes for a regression task	It can smoothly tackle the highly correlated features. It easily handles missing values, a large amount of data, or even imbalanced data and reduces the chances of overfitting, as the final decision depends on the decision of multiple trees	It is difficult to interpret its inner working; appears as a black box
Artificial neural network	Regression and classification	It is similar to human brain functioning and connects neurons to pass information	Performs well even with incomplete data. It works efficiently in recognizing patterns	Its functioning depends on the processing power. No specific rules are defined for selection of neurons or hidden layers. It is truly a black box algorithm

whilst SVM and MLP provided satisfactory performance with accuracies greater than 90%. Several similar studies have recently reported about the phase prediction and various mechanical properties which are summarized briefly in Table 2.

From the aforementioned literature, the random forest algorithm was observed to be the most prominently used algorithm in phase prediction studies, due to its high predictive performance. Therefore, the present study used random forest algorithm with the motive of correctly classifying each phase of HEAs for an imbalanced dataset. Scarce studies on the interpretation of the inner-working of an algorithm have been found in the literature. Thus, an additional attempt to decipher the black-box nature of the employed random forest algorithm has also been made using SHapely Additive exPlanation (SHAP) technique.

1.3 Data Source

A dataset of 1360 HEA samples was used to classify different phases of HEAs [30]. This dataset was observed to be highly imbalanced, as it contained 463 Intermetallic (IM), 441 BCC solid-solution (BCC_SS), 354 FCC solid-solution (FCC_SS), and

Table 2 Application of various ML algorithms in material science

Reference	Supervised ML method	Brief description	Success rate/Accuracy
Tancret et al. [20]	Gaussian processes (GP) + CALPHAD	Study on formation of single-phase solid-solution by considering the previously proposed empirical rules such as Hume-Rothery rules, basic thermodynamic concepts, and CALPHAD	63–80%
Islam et al. [15]	Neural network	Classification of solid-solution, intermetallic, and amorphous phase	83%
Huang et al. [16]	Comparison of k-nearest neighbour (KNN), support vector machine (SVM), and artificial neural network (ANN)	Solid-solution, intermetallic or mixed solid-solution and intermetallic phase formation study	SVM—64.3%, KNN—68.6%, ANN—74.3%
Choudhury et al. [18]	Random forest	Phase selection and crystal structure prediction study	91.66% for phase and 93.10% for crystal structure prediction
Li et al. [21]	Support vector machine (SVM)	Classification of BCC, FCC, and other phases not forming single phase solid-solution in CoCrFeMnNi HEA	96.55% training accuracy and 90.69% validation accuracy
Qi et al. [22]	Random forest	A phenomenological method to predict phases of high-entropy alloys using phase diagram data	> 80%
Zhou et al. [23]	Comparison of artificial neural network (ANN), convolutional neural network (CNN), and support vector machine (SVM)	Design rules for different phases (solid-solution (SS), amorphous (AM), and intermetallic (IM)) of high-entropy alloys	Accuracy of ANN model—98.9% for AM phase, 97.8% for SS phase, and 95.6% for IM phase

(continued)

Table 2 (continued)

Reference	Supervised ML method	Brief description	Success rate/Accuracy
Agrawal et al. [24]	ANFIS—Adaptive neuro fuzzy interface system	Classification of different phases of HEAs, using two approaches, one considering composition (composition-based model) as the inputs, whilst the other considering a set of six crucial parameters (parametric based model) in the formation of HEAs	84.21% for composition-based model, 80% for parametric-based model
Abdoon et al. [25]	Deep neural networks (DNNs)	Phase prediction of HEAs to design alloys	90%
Dai et al. [26]	Logistic regression	Phase prediction study on 407 data of high-entropy alloys using feature engineering	86% accurate with 9 descriptors
Kaufmann and Vecchio [27]	Random forest	High-throughput "ML-HEA" prediction of solid-solution forming-ability for HEAs by coupling thermodynamic and chemical features	(94% for binary) and (82.1% for ternary), when compared to the predictions from CALPHAD
Zhang et al. [28]	Support vector machine (SVM) model using four feature variables and Kernel principal component analysis (4 V-KPCA)	Phase prediction using relationship between phases and nine thermodynamics properties of high-entropy alloys	Testing accuracy—0.9743
Buranich et al. [29]	Linear regression (LR), Random forest (RF), and Gradient boosting regression (GBR)	Designing and screening of new HEAs for application in mechatronics industry	Highest accuracy (above 91%) for GBR
Risal et al. [19]	SVM, KNN, RF Classifiers, and ANN	Phase prediction of HEAs with an aim to significantly improve the phase prediction accuracy	KNN—92.31%, RF—91.21%, SVM and ANN—90%

(continued)

Table 2 (continued)

Reference	Supervised ML method	Brief description	Success rate/Accuracy
Nassar et al. [17]	Two neural networks: NN1 (considering only the compositional data), and NN2 (considering Hume-Rothery (HR) features along with the compositional data)	Phase prediction of HEAs	NN1—92%, NN2—90%
Machaka et al. [30]	Decision tree (DT) classifier, and Random forest (RF) classifier	Phase prediction of high-entropy alloys	DT—73%, RF—85%
Bhandari et al. [31]	Random forest	Yield strength prediction at the desired temperature for HEAs	93–97%, when compared with the experimental report for validation
Lee et. al. [32]	Regularized deep neural network (DNN) model	Identification of the key design parameters for enhancing the performance of phase prediction of HEAs. Furthermore, to overcome the problem of data shortage, a conditional generative adversarial network (GAN) was employed to generate more data	84.75%, 93.17% (when augmented by GAN)
Zeng et al. [33]	eXtreme Gradient Boosting (XGBoost) method	Phase selection rules for identification of single and mixed-phase	> 90%
Krishna et al. [34]	Performance comparison of logistic regression (LR), decision tree (DT), support vector machine (SVM), random forest (RF), gradient boosting classifier (GB), and artificial neural network (ANN)	Phase prediction in multiphase alloy system	LR—62.89%, SVM—83.02%, DT—77.99%, RF—82.39%, GB Classifier—81.13%, ANN—80.50%

102 mixed (FCC + BCC) phase. Five crucial features such as valence electron concentration (VEC), electronegativity difference ($\Delta\chi$), atomic size difference (δ), mixing enthalpy (dH_{mix}), and mixing entropy (ΔS_{mix}) calculated by Miedema's model were used as input parameters in the dataset. The ML modelling framework for the classification of four different phases of HEAs are demonstrated in Fig. 2, where data preprocessing is performed to detect outliers and missing values, and then feature scaling was performed to scale down the data into a finite range.

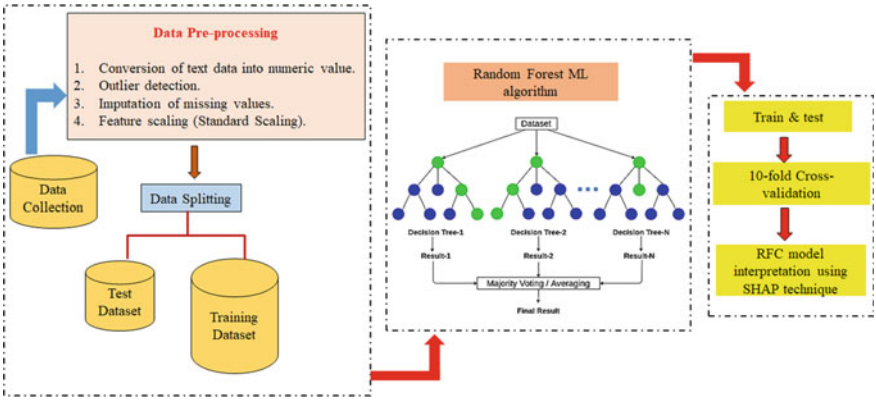

Fig. 2 ML modelling framework for phase classification of high-entropy alloys (HEAs)

The data was then divided into training and test sets (80:20) for the training (1088 HEA samples) and testing (272 HEA samples) purpose of the model. Random forest classifier (RFC) was employed using the scikit-learn library in Python, and further training and testing was performed to evaluate the performance of the RFC model.

1.4 RFC Model Performance

RF is an ensemble of various decision trees (tree-like structures) based on various subsets of the given dataset. For a classification task, it takes votes from various decision trees and makes a final prediction on the basis of majority votes. It is more accurate compared to a single decision tree algorithm, as a large number of trees improve its performance and makes the prediction more stable. The performance of RFC model has been evaluated using a confusion matrix and classification report. Confusion matrix provides the number of correctly predicted and incorrectly predicted classes for each phase [35]. Classification report on the other hand provides precision, recall, f1-score, average classification accuracy, and weighted average accuracy. Out of 272 HEA samples from the test data, 93 samples belong to class 0, i.e. IM phase, 88 samples belong to class 1, i.e. BCC phase, 71 samples belong to class 2, i.e. FCC phase, 20 samples belong to class 3, i.e. FCC + BCC mixed phase, as shown in classification report in Fig. 3a. The confusion matrix in Fig. 3b demonstrates the correctly and incorrectly classified phases, for example out of a total of 93 samples of IM phase, only 72 samples were correctly classified as IM phase, remaining 13 samples were misclassified as BCC phase, 5 samples were misclassified as FCC phase and 2 samples were misclassified as FCC + BCC phase. Similarly, the classification of each phase can be observed. Apart from an average classification accuracy of 86%, the precision, recall, and f1-score for all four phases were investigated in the classification report.

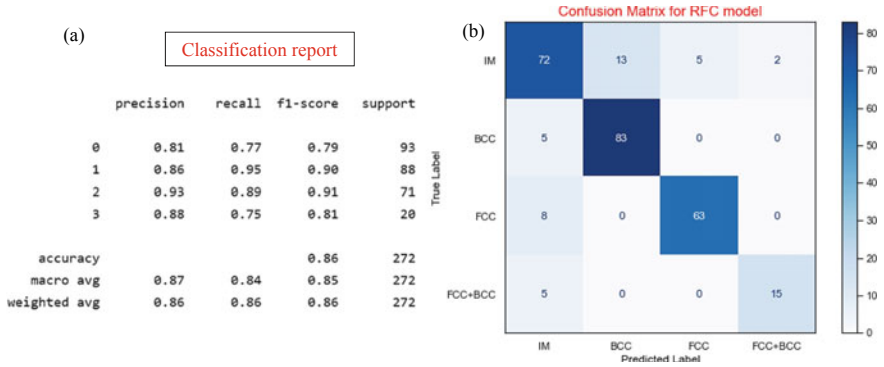

Fig. 3 **a** Confusion matrix, **b** classification report for RFC model in classification of IM, BCC, FCC, and FCC + BCC phase

Training accuracy, test accuracy, ROC_AUC_Score, and tenfold cross-validation ROC_AUC_Score were evaluated and shown in Table 3. The Receiver Operating Characteristic_Area under Curve (ROC_AUC) score ensures better performance of model in predicting different classes of HEAs. tenfold cross-validation was performed to avoid overfitting, where the complete dataset is divided into ten equal folds. Each time, the model was trained with 9 of those folds and the remaining onefold was used for the testing purpose, and the procedure was repeated ten times, by reserving a different tenth fold each time for testing the model. This measure ensured the effectiveness of the model and that the RFC model was not overfitting. The reported training accuracy of 90.9% and testing accuracy of 85.23%, suggested that the model was successfully classifying different phases of HEAs. The difference in the training and test accuracy indicated that the data used to train and test the RFC model was slightly different, which means that the RFC model is capable of predicting phases successfully, even for the unseen data points that have not been used in the present study.

Furthermore, the interpretation of RFC model was performed using SHAP technique, to understand which physical features were influential in governing phases of HEAs. SHapely Additive exPlanation (SHAP) technique has emerged as breakthrough in the field of ML for easy interpretation and explanation of a complex model's prediction; published by Lundberg and Lee in 2017 [36]. SHAP value aims to explain the prediction of an instance by considering the contribution of each feature

Table 3 Comparison of various ML algorithm performance

S. No.	Algorithm	Accuracy
1	Support vector machine	0.79
2	Decision tree classifier	0.81
3	Random forest classifier	0.86
4	XGBoost classifier	0.83

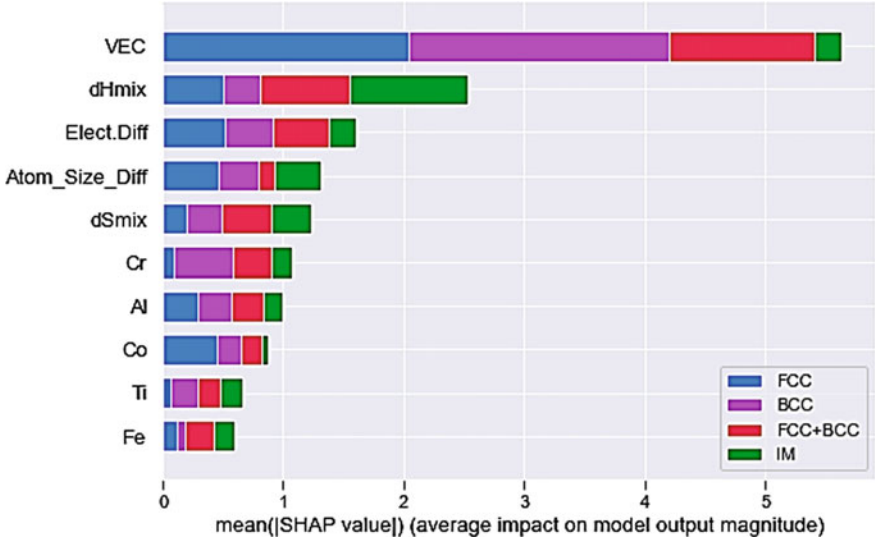

Fig. 4 Feature importance plot for RFC model using SHAP

Table 4 RFC model performance

S. No.	Evaluation	Model performance
1	Training accuracy	0.909672
2	Test accuracy	0.856088
3	ROC_AUC_Score	0.964626
4	Tenfold cross-validation ROC_AUC_Score	0.90311

in making a certain prediction at global as well as local levels. From this analysis, VEC was found to play most crucial role in determining FCC, BCC, and mixed FCC + BCC solid-solution phases whilst mixing enthalpy (dH_{mix}) was found important in determining the formation of solid-solution or an intermetallic phase, as shown in Fig. 4. Tables 4 and 5 show the RFC model performance and cross-validation score in each fold of tenfold cross-validation respectively.

2 Conclusions

The RFC model developed in this study had successfully and reliably predicted BCC, FCC, intermetallic, and FCC + BCC phases in high-entropy alloys. RFC model performance was evaluated using various evaluation metrics such as average classification accuracy, precision, recall, f1-score, ROC_AUC score, and tenfold cross validation ROC_AUC score.

Table 5 Cross-validation score in each fold of tenfold cross-validation

No. of folds	Cross-validation score
Fold 1	0.88633
Fold 2	0.93082
Fold 3	0.78310
Fold 4	0.92372
Fold 5	0.94252
Fold 6	0.93863
Fold 7	0.83451
Fold 8	0.89151
Fold 9	0.96391
Fold 10	0.93750

Mean value of tenfold cross-validation score: 0.9031
Standard deviation: 0.0533

The RFC model showed that the two input parameters, namely valence electron concentration (VEC) and mixing enthalpies were most influential in determining the resulting phase of a given HEA composition. VEC contributed the most in predicting the crystal structure of solid solution phases (BCC, FCC, and FCC + BCC) whilst mixing enthalpy (ΔH_{mix}) played important role in determining formation of solid-solution or intermetallic. Thus, this present study leveraged the reliability of applying ML techniques in material science without any requirement of performing expensive experiments.

Acknowledgements Swati Singh greatly acknowledge the scholarship provided by the Ministry of Education, Government of India. Saurav Goel greatly acknowledge the support provided by the Royal Academy of Engineering via Grants No. IAPP18-19\295 and TSP1332.

References

1. Katiyar N, Goel G, Goel S (2021) Emergence of machine learning in the development of high entropy alloy and their prospects in advanced engineering applications. Emergent Mater 4(6):1635–1648. https://doi.org/10.1007/s42247-021-00249-8
2. Senkova ON, Senkova SV, Woodward C, Miracle DB (2013) Low-density, refractory multi-principal element alloys of the Cr–Nb–Ti–V–Zr system: Microstructure and phase analysis. Acta Mater 61(5):1545–1557. https://doi.org/10.1016/j.actamat.2012.11.032
3. Stepanov N, Shaysultanov D, Salishchev G, Tikhonovsky M (2015) Structure and mechanical properties of a light-weight AlNbTiV high entropy alloy. Mater Lett 142:153–155. https://doi.org/10.1016/j.matlet.2014.11.162
4. Deng Y, Tasan C, Pradeep K, Springer H, Kostka A, Raabe D (2015) Design of a twinning-induced plasticity high entropy alloy. Acta Mater 94:124–133. https://doi.org/10.1016/j.actamat.2015.04.014
5. Zhang Y, Li R (2020) New advances in high-entropy alloys. Entropy 22(10):1158. https://doi.org/10.3390/e22101158

6. Tsai M, Yeh J (2014) High-entropy alloys: a critical review. Mater Res Lett 2(3):107–123. https://doi.org/10.1080/21663831.2014.912690
7. Youssef K, Zaddach A, Niu C, Irving D, Koch C (2014) A novel low-density, high-hardness, high-entropy alloy with close-packed single-phase nanocrystalline structures. Mater Res Lett 3(2):95–99. https://doi.org/10.1080/21663831.2014.985855
8. Cantor B (2021) Multicomponent high-entropy Cantor alloys. Prog Mater Sci 120:100754. https://doi.org/10.1016/j.pmatsci.2020.100754
9. Yeh J (2006) Recent progress in high-entropy alloys. Annales De Chimie Science Des Matériaux 31(6):633–648. https://doi.org/10.3166/acsm.31.633-648
10. Kremer K, Grest G (1990) Molecular dynamics (MD) simulations for polymers. J Phys Condens Matter 2(S):SA295-SA298. https://doi.org/10.1088/0953-8984/2/s/045
11. Neugebauer J, Hickel T (2013) Density functional theory in materials science. Wiley Interdiscip Rev: Comput Mol Sci 3(5):438–448. https://doi.org/10.1002/wcms.1125
12. Goel S, Knaggs M, Goel G, Zhou X, Upadhyaya H, Thakur V et al (2020) Horizons of modern molecular dynamics simulation in digitalized solid freeform fabrication with advanced materials. Mater Today Chem 18:100356. https://doi.org/10.1016/j.mtchem.2020.100356
13. Osisanwo FY, Akinsola JE, Awodele O, Hinmikaiye JO, Olakanmi O, Akinjobi J (2017) Supervised machine learning algorithms: classification and comparison. Int J Comput Trends Technol 48(3):128–138. https://doi.org/10.14445/22312803/ijctt-v48p126
14. Nasteski V (2017) An overview of the supervised machine learning methods. HORIZONS B 4:51–62. https://doi.org/10.20544/horizons.b.04.1.17.p05
15. Islam N, Huang W, Zhuang H (2018) Machine learning for phase selection in multi-principal element alloys. Comput Mater Sci 150:230–235. https://doi.org/10.1016/j.commatsci.2018.04.003
16. Huang W, Martin P, Zhuang H (2019) Machine-learning phase prediction of high-entropy alloys. Acta Mater 169:225–236. https://doi.org/10.1016/j.actamat.2019.03.012
17. Nassar A, Mullis A (2021) Rapid screening of high-entropy alloys using neural networks and constituent elements. Comput Mater Sci 199:110755. https://doi.org/10.1016/j.commatsci.2021.110755
18. Choudhury A, Konnur T, Chattopadhyay P, Pal S (2019) Structure prediction of multi-principal element alloys using ensemble learning. Eng Comput 37(3):1003–1022. https://doi.org/10.1108/ec-04-2019-0151
19. Risal S, Zhu W, Guillen P, Sun L (2021) Improving phase prediction accuracy for high entropy alloys with machine learning. Comput Mater Sci 192:110389. https://doi.org/10.1016/j.commatsci.2021.110389
20. Tancret F, Toda-Caraballo I, Menou E, Rivera Díaz-Del-Castillo P (2017) Designing high entropy alloys employing thermodynamics and Gaussian process statistical analysis. Mater Des 115:486–497. https://doi.org/10.1016/j.matdes.2016.11.049
21. Li Y, Guo W (2019) Machine-learning model for predicting phase formations of high-entropy alloys. Phys Rev Mater 3(9). https://doi.org/10.1103/physrevmaterials.3.095005
22. Qi J, Cheung A, Poon S (2019) High entropy alloys mined from binary phase diagrams. Sci Rep 9(1). https://doi.org/10.1038/s41598-019-50015-4
23. Zhou X, Zhu J, Wu Y, Yang X, Lookman T, Wu H (2022) Machine learning assisted design of FeCoNiCrMn high-entropy alloys with ultra-low hydrogen diffusion coefficients. Acta Mater 224:117535. https://doi.org/10.1016/j.actamat.2021.117535
24. Agrawal A, Choudhary A (2016) Perspective: materials informatics and big data: realization of the "fourth paradigm" of science in materials science. APL Mater 4(5):053208. https://doi.org/10.1063/1.4946894
25. Abdoon Al-Shibaany Z, Alkhafaji N, Al-Obaidi Y, Atiyah A (2020) Deep learning-based phase prediction of high-entropy alloys. IOP Conf Ser: Mater Sci Eng 987(1):012025. https://doi.org/10.1088/1757-899x/987/1/012025
26. Dai D, Xu T, Wei X, Ding G, Xu Y, Zhang J, Zhang H (2020) Using machine learning and feature engineering to characterize limited material datasets of high-entropy alloys. Comput Mater Sci 175:109618. https://doi.org/10.1016/j.commatsci.2020.109618

27. Kaufmann K, Vecchio K (2020) Searching for high entropy alloys: a machine learning approach. Acta Mater 198:178–222. https://doi.org/10.1016/j.actamat.2020.07.065
28. Zhang L, Chen H, Tao X, Cai H, Liu J, Ouyang Y et al (2020) Machine learning reveals the importance of the formation enthalpy and atom-size difference in forming phases of high entropy alloys. Mater Des 193:108835. https://doi.org/10.1016/j.matdes.2020.108835
29. Buranich V, Rogoz V, Postolnyi B, Pogrebnjak A (2020) Predicting the properties of the refractory high-entropy alloys for additive manufacturing based fabrication and mechatronic applications. In: IEEE international conference on "nanomaterials: applications & properties" (NAP-2020) symposium on additive manufacturing and applications (SAMA-2020) Sumy, Ukraine, 9–13 Nov 2020
30. Machaka R (2021) Machine learning-based prediction of phases in high-entropy alloys. Comput Mater Sci 188:110244. https://doi.org/10.1016/j.commatsci.2020.110244
31. Bhandari U, Rafi M, Zhang C, Yang S (2021) Yield strength prediction of high-entropy alloys using machine learning. Mater Today Commun 26:101871. https://doi.org/10.1016/j.mtcomm.2020.101871
32. Lee S, Byeon S, Kim H, Jin H, Lee S (2021) Deep learning-based phase prediction of high-entropy alloys: optimization, generation, and explanation. Mater Des 197, 109260. https://doi.org/10.1016/j.matdes.2020.109260
33. Zeng Y, Man M, Bai K, Zhang Y (2021) Revealing high-fidelity phase selection rules for high entropy alloys: a combined CALPHAD and machine learning study. Mater Des 202, 109532. https://doi.org/10.1016/j.matdes.2021.109532
34. Krishna Y, Jaiswal U, Rahul R (2021) Machine learning approach to predict new multi-phase high entropy alloys. Scripta Mater 197:113804. https://doi.org/10.1016/j.scriptamat.2021.113804
35. Markoulidakis I, Rallis I, Georgoulas I, Kopsiaftis G, Doulamis A, Doulamis N (2021) Multi-class confusion matrix reduction method and its application on net promoter score classification problem. Technologies 9(4):81. https://doi.org/10.3390/technologies9040081
36. Lewis F, Butler A, Gilbert L (2010) A unified approach to model selection using the likelihood ratio test. Methods Ecol Evol 2(2):155–162. https://doi.org/10.1111/j.2041-210x.2010.00063