Scripta Materialia 185 (2020) 152–158

Contents lists available at ScienceDirect

Scripta Materialia

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

Machine learned feature identification for predicting phase and Young's modulus of low-, medium- and high-entropy alloys

Ankit Roy, Tomas Babuska, Brandon Krick, Ganesh Balasubramanian*

Department of Mechanical Engineering & Mechanics, Lehigh University, Packard Laboratory 561, 19 Memorial Drive West, Bethlehem, PA 18015, USA

ARTICLE INFO

Article history:
Received 21 February 2020
Revised 13 April 2020
Accepted 15 April 2020

Keywords:
High-entropy alloys
Machine learning
Gradient boost algorithm
Crystallographic phase
Young's modulus

ABSTRACT

The growth in the interest and research on high-entropy alloys (HEAs) over the last decade is due to their unique material phases responsible for their remarkable structural properties. A conventional approach to discovering new HEAs requires scavenging an enormous search space consisting of over half a trillion new material compositions comprising of three to six principal elements. Machine learning has emerged as a potential tool to rapidly accelerate the search for and design of new materials, due to its rapidity, scalability, and now, reasonably accurate material property predictions. Here, we implement machine learning tools, to predict the crystallographic phase and Young's modulus of low-, medium- and high-entropy alloys composed of a family of 5 refractory elements. Our results, in conjunction with experimental validation, reveal that the mean melting point and electronegativity difference exert the strongest contributions to the phase formation in these alloys, while the melting temperature and the enthalpy of mixing are the key features impacting the Young's modulus of these materials. Additionally, and more importantly, we find that the entropy of mixing only negligibly influences the phase or the Young's modulus, reigniting the issue of its actual impact on the material phase and properties of HEAs.

High-entropy alloys (HEAs) are compositionally complex alloys containing multiple principal elements in near equiatomic proportions, and have received widespread attention due to their promising structural properties [1–5]. The underlying principle is that for 5 or more elements, the high configurational entropy potentially overcomes the enthalpy of mixing and prevents phase separation or the formation of compounds [6]. While Cantor et al. [7] reported that CrCoFeMnNi assumes a single phase FCC solid solution, presence of multiple phases has been noted for other HEAs [2]. Typically, a single phase solid solution (SS) contributes to excellent mechanical properties [8], while ductility deteriorates in the presence of intermetallics that are inherently brittle [9]. Thus, knowledge of the HEA phase serves as an indicator of its material properties.

Fundamentally, lowering the Gibbs free energy of mixing (ΔG_{mix}) in $\Delta G_{mix} = \Delta H_{mix} - T\Delta S_{mix}$ stabilizes the alloy. Here, T is the absolute temperature, ΔH_{mix} is the enthalpy of mixing, $\Delta S_{mix} = -R\sum_{i=1}^{n}(C_i \ln C_i)$ is the entropy of mixing [10], R the gas constant and C_i is the atomic fraction of the i^{th} component in the alloy. It has been proposed that for a solid solution to form, $\Omega > 1$, while intermetallic compounds would likely form if $\Omega < 1$, where $\Omega = \frac{T_m \Delta S_{mix}}{|\Delta H_{mix}|}$, and T_m is the melting temperature of an alloy composed of n principal elements [11]. A second parameter $\delta = \sqrt{\sum_{i=1}^{n} C_i(1-\frac{r_i}{\bar{r}})^2}$ has been identified that accounts for the effect of atomic size differences between the principal elements ($\bar{r}$ is the average atomic radius and r_i is the atomic radius of the i^{th} component) [12]. From Ω and δ predictions for over 130 alloys available in the literature, it has been reported that a solid solution forms when $\Omega > 1$ and $\delta \leq 6.6\%$. Along the same lines, an earlier report [13] suggested that the conditions of $-22\frac{kJ}{mol} \leq \Delta H_{mix} \leq 7\frac{kJ}{mol}$, $0 \leq \delta \leq 8.5$ and $11 \leq \Delta S_{mix} \leq 19.5\frac{J}{mol*K}$ must be satisfied simultaneously for the formation of solid solutions in equiatomic multicomponent alloys. High values of a purely geometric parameter $\lambda = \frac{\Delta S_{mix}}{\delta^2}$, where δ is representative of strain with respect to a perfect lattice and δ^2 is analogous to strain energy, was found to favor the formation of disordered solid solution [14]. In particular, $\lambda > 0.96$ indicated single phase disordered solid solutions, two phase mixtures for $0.24 < \lambda < 0.96$, while compound formation was predicted for $\lambda < 0.24$.

Recent efforts utilizing machine learning [15] approaches considered two additional descriptors viz., Pauling electronegativity difference $\Delta\chi = \sqrt{\sum_{i=1}^{n} C_i(\chi_i - \bar{\chi})^2}$ and difference in valence electron concentration $VEC = \sum_{i=1}^{n} C_i(VEC)_i$, and used neural network (NN) to predict the phases that form in these complex concentrated alloys. Nevertheless, a gap exists in establishing the relative

* Corresponding author.
E-mail address: bganesh@lehigh.edu (G. Balasubramanian).

https://doi.org/10.1016/j.scriptamat.2020.04.016
1359-6462/

Table 1
The list of features and their mathematical representation as employed in the machine learning framework. C_i corresponds to the mole fraction of the i^{th} element, $\Delta\chi$ to the difference in Pauling electronegativities, ΔH_{mix} to the mixing enthalpy, ΔS_{mix} to the mixing entropy, δ to the difference in atomic radii, while λ is a geometrical parameter, T_m is the melting temperature calculated by the rule of mixtures, a_m is the lattice constant calculated by the rule of mixtures, Ω is a parameter for predicting the solid solution formation, Δa is the difference in lattice constants and ΔT_m is the difference in melting temperatures.

Feature	Description	Reference
$\Delta\chi = \sqrt{\sum_{i=1}^{n} C_i(\chi_i - \bar{\chi})^2}$	Difference in Pauling electronegativities	[30]
$\Delta H_{mix} = \sum_{i=1, i\neq j}^{n} 4H_{ij}C_iC_j$	Mixing enthalpy	[23]
$\Delta S_{mix} = -R\sum_{i=1}^{n}(C_i lnC_i)$	Mixing entropy	[10]
$\delta = \sqrt{\sum_{i=1}^{n} C_i(1 - \frac{r_i}{\bar{r}})^2}$	Difference in atomic radii	[11]
$\lambda = \frac{\Delta S_{mix}}{\delta^2}$	A geometrical parameter	[14]
$T_m = \sum_{i=1}^{n} C_iT_i$	Melting temperature calculated by the rule of mixtures	[31]
$a_m = \sum_{i=1}^{n} C_ia_i$	Resulting lattice constant calculated by the rule of mixtures	[31]
$\Omega = \frac{T_m\Delta S_{mix}}{\lvert\Delta H_{mix}\rvert}$	Parameter for predicting the solid solution formation	[11]
$\Delta a = \sqrt{\sum_{i=1}^{n} C_i(a_i - \bar{a})^2}$	Difference in lattice constants	Proposed by analogy to δ
$\Delta T_m = \sqrt{\sum_{i=1}^{n} C_i(T_i - \bar{T})^2}$	Difference in melting temperatures	Proposed by analogy to $\Delta\chi$

importance of the material features for the prediction of the specific crystallographic phase that any given HEA composition may assume. Here, for a range of equiatomic low- (binary), medium- (ternary, quaternary) and high- (quinary) entropy alloys composed of refractory elements (Mo-Ta-Ti-W-Zr), we employ gradient boost algorithms to predict the Young's modulus (E) and identify the specific crystallographic phases i.e., BCC, FCC or multiphase. The machine learned results, as discussed below, exhibit remarkable agreement with experimental characterization and measurements. We choose the alloys in the Mo-Ta-Ti-W-Zr family as a testbed, since a recent report has predicted a certain MoTaTiWZr HEA to possess an extraordinary Young's modulus of ~335 GPa [3], superior to commercial alloys. Nevertheless, the proposed approach is generic and applicable across a wide gamut of HEA compositions.

Data assemblage and feature selection: Available experimental data is collected from existing literature [2,4,15–20]. The data for crystallographic phase prediction ('phases dataset') consists of 329 entries where 159 are BCC HEAs, 111 are FCC HEAs and 59 are multiphase. Since there are relatively smaller number of HCP HEAs discussed in the literature, we do not include them in 'phases dataset' to avoid creating a biased dataset that reduces the accuracy of the predictive model [16]. The 'multiphase' label in the phase prediction model indicates that the alloy is not a single phase solid solution and could exhibit a secondary phase or an intermetallic phase. It does not indicate a mixture of FCC and BCC as the 'phases dataset' used for training the crystallographic phase prediction model had $L1_2$ and B_2 precipitates in the multiphase alloy data. The dataset for Young's modulus ('E-dataset') consists of only 87 entries due to limited experimental reports. The datasets have been executed using the Jupyter Notebook [17] as a DataFrame. The features that have been adopted for the crystallographic phase prediction are listed in Table 1. T_m is an indirect measure of strength of interatomic interactions [18], while a_m is reflective of the metallic bonding strength. We additionally include ΔT_m as a measure of difference in magnitudes of atomic interactions, and Δa as a measure of the degree of incongruence of the combining lattices. A low ΔT_m and Δa would facilitate atoms of different metals to combine into a single lattice. For the Young's modulus dataset, the features ΔT_m and Δa have not been included, and has been discussed below where in Fig. 3(b) it is seen that one of the features, a_m (lattice constant calculated by the rule of mixtures), has a negligible impact in determining the Young's modulus. Based on this observation, it is redundant to include Δa (difference in lattice constants) as an additional feature. Moreover, the 'phases dataset' does contain both the ΔT_m and Δa as features. This concept stems from the fundamental principle of single phase solid solution formation [19,20] that emphasizes on calculating the 'concentration averaged difference' in parameters like electronegativity or atomic size when describing HEAs because the latter do not have a single dominant solvent or a solute [21]. Hence, ΔT_m and Δa are included for predicting the formation of single phase solid solutions and not for predicting the Young's modulus. Note that $\Delta\chi$ (difference in Pauling electronegativities) is not analogous to the valence electron concentration (VEC) for metals per se but it is closely related to the electron affinity and VEC; it can be used for determining bond strengths and atom positions relative to other species in a matrix [22]. Hence, in this work, $\Delta\chi$ has been used as a representation of all other electronic features namely VEC and electron affinity. Prior literature [23] have used VEC independently to classify the phases formed in HEAs by developing empirical models that suggest formation of FCC, when VEC $\geq$ 8.0; formation of BCC when VEC $\leq$ 6.87 and formation of mixed FCC and BCC phases otherwise. This empirical model has been tried in the current work and discussed below, but yields limited success.

Data analysis is initially performed by calculating the Pearson correlation coefficient P [24] using a heat map illustrated in Fig. 1. Here, $P = 1$ denotes strong positive correlation and $P = -1$ denotes a strong negative correlation. The absence of any significant correlation amongst any pair of features indicates that all metrics should be considered in the model.

Model construction: After assembling the data and evaluating the feature values, we employ the datasets to construct two machine learning models, one for predicting the Young's modulus and the other for the crystallographic phase. In each case, the dataset is classified into a training set (90% of the data) and a test set (10% of the data). For example, in our 'phases dataset', we use around 296 (90%) data points for training the model and around 33 (10%) data points for testing the model. This process is essential to build a robust model with minimum training error (discussed in following sections) before the model can be adopted for actual usage. The errors generated in the predictions of the test dataset are quantified and minimized by altering the parameters of the models, which are retrained iteratively on the training data followed by re-predictions on the test subset until the error is minimized below a threshold. The model is finally used for predicting crystallographic phases and Young's moduli of alloys in Table 2. The final model predictions are validated with experimental measurements.

Table 2
List of the multicomponent alloys synthesized and used as experimental validation dataset to estimate the performance of the constructed machine learning model.

Alloy	VEC	Expt. E(GPa)	ML Predicted E (GPa)	% Error	Expt.Phases	ML Predicted Phases	Correct (✔)
MoZr	5.00	170.18	114.96	32.44	Multiphase	BCC	
MoW	6.00	278.54	126.87	54.45	BCC	BCC	✔
TaTi	4.50	111.35	100.61	9.65	BCC	BCC	✔
TaW	5.50	228.93	117.34	42.83	BCC	FCC	
WZr	5.00	184.15	162.17	11.93	Multiphase	BCC	
MoTi	5.00	149.76	98.95	33.92	BCC	BCC	✔
MoTa	5.50	225.55	210.52	6.66	BCC	BCC	✔
TaZr	4.50	148.78	149.38	0.40	HCP	BCC	
TiZr	4.00	87.67	118.96	35.69	HCP	Multiphase	
TiW	5.00	187.03	154.14	17.59	BCC	BCC	✔
MoWZr	5.33	169.72	178.70	5.29	Multiphase	BCC	
MoTaTi	5.00	173.39	152.74	11.91	BCC	BCC	✔
TaWZr	5.00	200.28	200.07	0.11	Multiphase	BCC	
TaTiZr	4.33	122.64	98.33	19.82	Multiphase	FCC	
MoTiZr	4.67	138.74	131.07	5.53	FCC	BCC	
MoTaW	5.67	300.38	141.68	52.83	BCC	FCC	
MoTiW	5.33	201.33	166.37	17.36	BCC	BCC	✔
TaTiW	5.00	195.83	172.95	11.68	BCC	BCC	✔
TiWZr	4.67	134.90	174.52	29.36	Multiphase	Multiphase	✔
MoTaZr	5.00	200.97	168.76	16.03	Multiphase	BCC	
MoTiWZr	5.00	172.58	185.91	7.72	HCP	BCC	
MoTaTiW	5.25	199.23	177.73	10.79	BCC	BCC	✔
TaTiWZr	4.75	168.65	171.27	1.55	FCC	FCC	✔
MoTaWZr	5.25	221.52	185.62	16.21	Multiphase	Multiphase	✔
MoTaTiZr	4.75	161.75	155.66	3.77	Multiphase	Multiphase	✔
MoTaTiWZr	5.00	157.56	175.46	11.36	BCC	BCC	✔

Model for predicting crystallographic phase: The phase prediction is performed by employing the Gradient Boost Classification [25] algorithm. This algorithm starts with one leaf that is equal to the initial prediction for every data in the phases-dataset. This initial prediction is effectively the logarithm (log) of the odds of a particular output. This log is calculated by taking the log of the number of favorable outputs over the number of unfavorable outputs. To use this log value conveniently, we convert it to a probability representation using the logistic function $P = \frac{e^{\log(odds)}}{1+\ e^{log(odds)}}$. $P = 0.5$ is set as the threshold value, typical of classification problems, and the residuals are calculated and stored. For the data with the favorable output, the $residual = 1 - P$ and for the data with unfavorable output the $residual = 0 - P$. These residuals form the leaves of the current tree. The number of leaves range between 8 and 32; hence, if a leaf contains more than one residual, then it is replaced by a single value called "output from tree" that is obtained from

$$output\ from\ tree = \frac{\sum residual}{\sum [(previous\ probability)_i \times\ (1 - previous\ probability)_i]}$$

Subsequently, the log of the odds is updated as

$$\log(odds) = previous\ prediction + \sum_{i=1}^{number\ of\ trees} (learning\ rate) \times (output\ from\ tree)$$

Once the new log(odds) value is obtained, we again convert it to a probability (P) as discussed previously. The process is repeated by calculating the new residuals and at every successive step the residual value is expected to decrease until it attains a minimum at the last tree. The final probabilities are used to classify the data into favorable or unfavorable outputs.

Model for predicting Young's modulus: Gradient boost for regression [25] is used in the model for predicting E. The algorithm initiates by building a leaf from the training data. This leaf resembles an initial guess for the output of all data in the dataset. For instance, the first guess is essentially the average E of the entire E-dataset. Next, the algorithm builds a tree such that a node of the tree is analogous to a conditional loop. The conditions are based on the errors from a previous tree. Errors are defined as the pseudo residuals or differences between the observed and the predicted E. Note that when the first tree is built, it originates from a single leaf that is the mean of all E values. The maximum number of leaves is set between 8 and 32. Whenever a new tree is built, it calculates the pseudo residuals for all the entries in the dataset and is scaled by a parameter called "learning rate" to avoid overfitting of the model to the data. The product of the "learning rate" (= 0.1) and the pseudo residual obtained at the end of each tree is the actual contribution of that tree. The algorithm continues to build more trees until a specific number of tress or an improved fit is obtained. Given the number of leaves is limited between 8 and 32, we get fewer leaves than the number of pseudo residuals (equal to the size of the dataset). Hence, if multiple pseudo residuals fall under the same leaf, they are replaced by a single average data point. Each successive tree is combined with the existing sequences of trees, and after the required number of trees have been built we predict

$$\begin{aligned} E \quad = \quad & E_{mean} + [(learning\ rate) \times residual]_1 \\ & + [(learning\ rate) \times residual]_2 + \ldots \\ & + [(learning\ rate) \times residual]_n \end{aligned}$$

where the subscripts *1, 2, …, n* represent the number of the trees. At each successive tree, the residual value decreases and reduces to a minimum at the final tree that provides the closest prediction to the measurements.

Experimental validation: We synthesize 26 alloys composed Mo-Ta-Ti-W-Zr elements, as listed in Table 2, by arc-melting compressed pellets of elemental powder mixes (Sigma-Aldrich, purity ≥ 99.9%) in Argon atmosphere (pressure 30 psi) on a water-cooled copper hearth. Powdered metals are used to minimize the occurrences of elemental macro-segregation and achieve homogeneity in the alloys. Powders are mixed thoroughly in a laboratory jar mill (Thomas Scientific Series 8000), at an optimum rpm to ensure a balance between the centrifugal and gravitational forces. Next, the

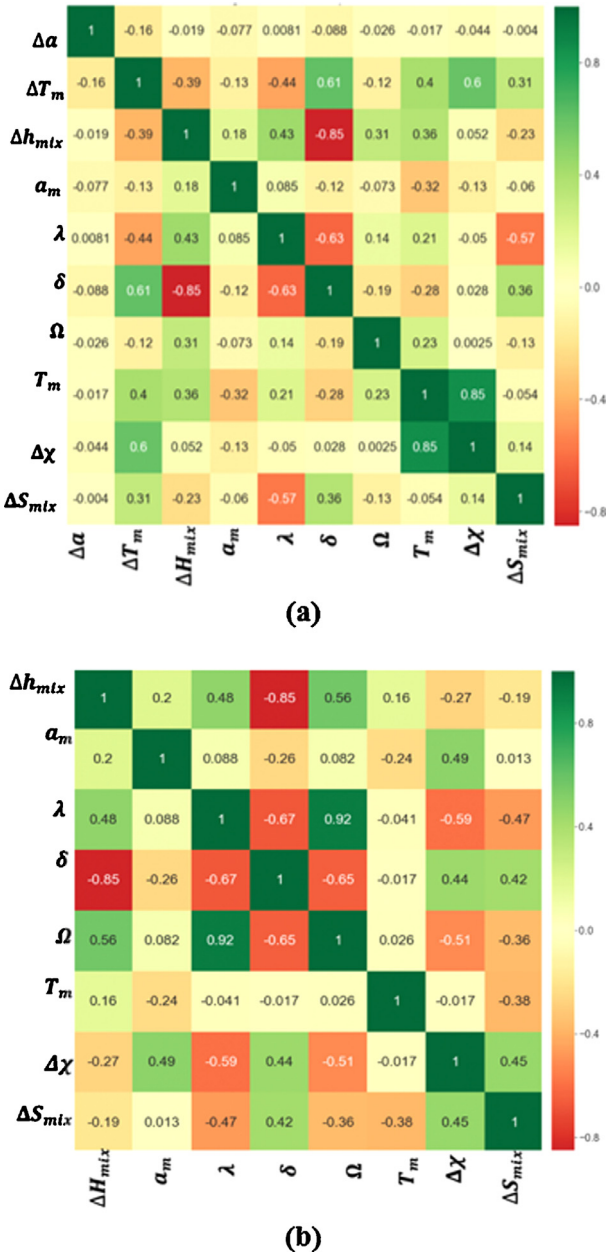

Fig 1. Heat map displaying the correlation values between the features of **(a)** phases dataset and **(b)** Young's modulus (E) dataset, as employed for the different multicomponent alloys. A value close to −1 or 1 implies negative or positive correlation, respectively.

powder mixes are compressed to a pressure of 5000 psi on cylindrical pellets of diameter 20 mm using a Carver hydraulic press. The pellets are then arc-melted (in Edmund Bühler GmbH Mini Arc Melting System), cooled and re-melted a total of four times to ensure improved homogeneity. An X-ray diffractometer (Panalytical Empyrean vs. 7.9f 20170530 X-ray Diffraction Unit) is employed for characterization of the crystal structure, with the 2θ scan ranging from 10 to 90°, with the radiation from a 45 kV, 40 mA copper target. Subsequently, the data is analyzed with Malvern Panalytical HighScore software package [26]. The alloy data in Table 2 serves as the validation set for both classification and regression models. Nanoindentation is performed on mounted and polished samples using a Hysitron TI900 nanoindenter with a Berkovich tip. Indents are performed using load control with a maximum load of 5 mN and a 5–2–5 s load-hold-unload profile. An array of 25 indents (5 by 5 pattern) with a spacing of 10 μm in the x and y was performed on each sample. Modulus and hardness are determined using the Oliver and Pharr method [27].

Crystallographic phase predictions: The tunable parameters in a machine learning algorithm (gradient boost classifier in this case) are known as the hyperparameters. The hyperparameters for gradient boost algorithm are predominantly "learning rate" and "n-estimators" [25,28]. These hyperparameters can be set to an optimum value while training the algorithm on the dataset, to eventually produce minimum training error. Here, 90% of the data from training set is employed to train the model and the remaining 10% is used as a test set to check for error in prediction. We define error as

$$error = \frac{number\ of\ incorrect\ predictions}{total\ number\ of\ data\ points}$$

The error is calculated for all possible combinations of 15 "learning rate" values ranging from 0.001 to 0.5 and 15 "n-estimators" ranging from 50 to 1000. Fig. 2 illustrates the hyperparameters' optimization (the process of the guiding the hyperparameters to an optimum value) to minimize error. A learning rate of 0.05–0.2 with 450–600 estimators produces a minimum error of 0.3077 as identified in region R in Fig. 2(a). Thus, any set of hyperparameter values within that range can be considered for the final model. Here, the learning rate is set to 0.1 and n-estimators is set to 500. The output produced by the model after the experimental dataset for the 26 alloys, is listed in Table 2. On comparison between the measured and predicted values for crystallographic phases, our model is able to accurately identify the lattice structure of 14 out of the possible 23 alloys (the crystal structure test data contains 3 alloys that form HCP lattices, and our existing model can predict only FCC, BCC or a mix of FCC and BCC). The empirical model [23] for determining phases using VEC seems to be of limited accuracy because all alloys in this study show a VEC $\leq$ 6 (Table 2) which categorizes them all as BCC phase. Hence, the empirical model cannot be guaranteed to yield correct results when applied to a refractory system such as the one under investigation.

Young's moduli predictions: We define the mean absolute error as $mean\ absolute\ error = \frac{\sum_{i=1}^{n} |y_i - x_i|}{n}$, where y_i is the prediction and x_i is the measured Young's modulus from experiments, while n is the total number of data points available. From the hyperparameter optimization illustrated in Fig. 2(b), we find that a learning rate of 0.0031 and n-estimators of 1000 produces the minimum mean absolute error of 23.59 GPa, and consequently adopt that for the final model. The model output reveals that 19 of the 26 predictions are within an error of 20%, while 14 are within a 12% margin. We strongly assert that the accuracy is limited by the sparse data available to us, and will improve as the additional simulation and experimental results are incorporated into training the model. In the current state of the model, the root-mean square error (RMSE) for this regression model was calculated to be 87.76%.

The relative importance of the different features in the structure and modulus predictions are illustratively compared in Fig. 3. For crystallographic phase prediction, the order of feature importance is noted to be $T_m > \Delta\chi > \Delta T_m > a_m > \Omega > \Delta H_{mix} > \Delta a > \lambda > \Delta S_{mix} > \delta$, while for Young's modulus, we find the order to be $T_m > \Delta H_{mix} > \lambda > \Omega > \Delta\chi > \delta > \Delta S_{mix} > a_m$. These results evince that certain descriptors like T_m and ΔH_{mix} are more crucial than the Hume-Rothery parameters ($\Delta\chi$ and δ) in determining the Young's modulus, while T_m, $\Delta\chi$, ΔT_m are key in estimating the phases that form in these multicomponent alloys.

T_m and $\Delta\chi$ are predicted to be the key parameters in determining the alloy phases. Since a high T_m of the alloying elements is analogous to their high bonding energies, the constituent metal with the highest melting temperature would preferably not allow

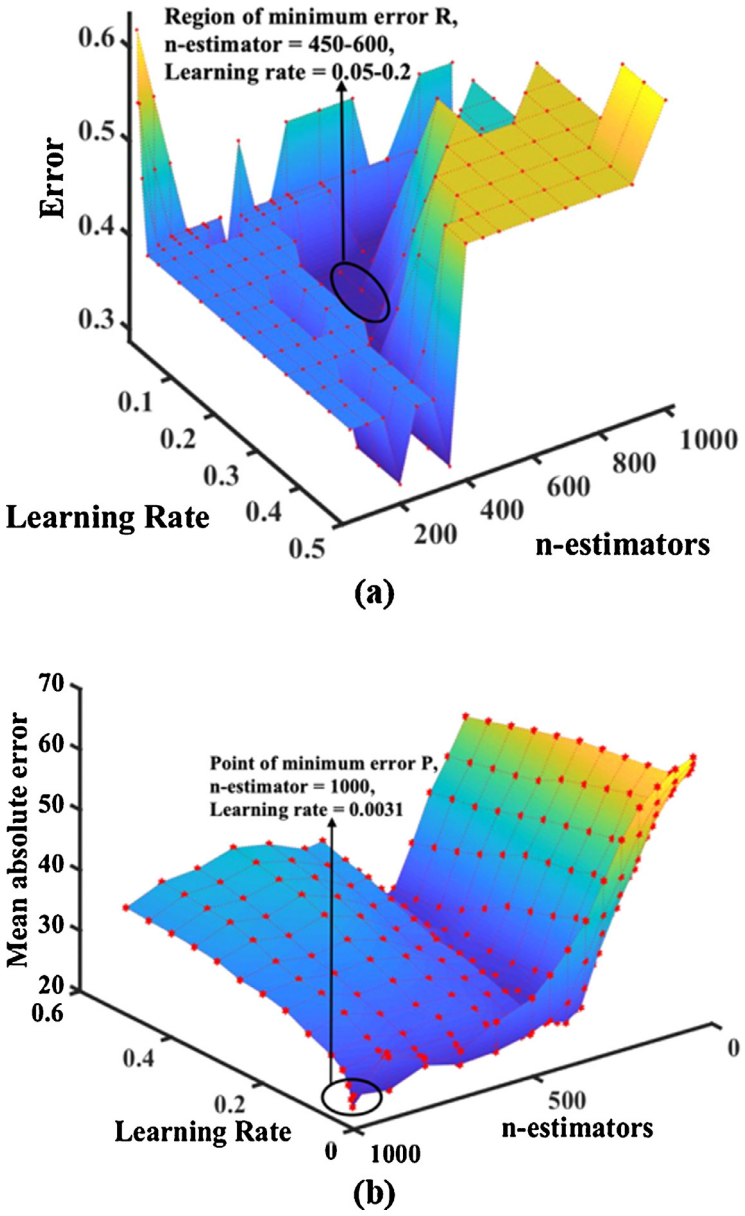

Fig 2. Hyperparameter optimization for the **(a)** gradient boost classifier algorithm to predict the crystallographic phase, with learning rate set to 0.1 and n-estimators set to 500 to minimize the error in phase prediction; **(b)** gradient boost regressor algorithm to predict the Young's modulus, with a learning rate of 0.0031 and 1000 n-estimators to produce the minimum mean absolute error of 23.59 GPa.

the inclusion of another metal atom with a lower bonding energy and hence the system would form multiple phases. On the other hand, the difference in electronegativity is directly responsible for the probability of intermetallic compound (secondary phases) formation. ΔT_m plays a vital role in determining phases as it conveys information on relative bond strengths of the metals in the alloys; the lower the ΔT_m, the higher is the probability of the formation of single phase as the metallic bond strengths in each metal will be identical. Δa and a_m exhibit strong contribution in determining the phases, relative to ΔS_{mix}, as similar lattice parameters enable the formation of single phase to be more probable. Interestingly, contrary to the Hume-Rothery guidelines and an earlier report [11], our machine learned model identifies δ to be the least decisive feature in determining single phase formation in the equiatomic alloys. While Ω plays a considerable role in determining the phase, λ and ΔS_{mix} can be regarded as insignificant features for predicting the crystallographic phases.

Our results conclusively suggest that melting point of alloys, which is an indirect metric of bond strength [18], significantly influences the Young's modulus. Likewise, since a high negative mixing enthalpy and a high electronegativity difference enhances the probability of the formation of intermetallics that are brittle and have low E values, ΔH_{mix} and $\Delta\chi$ play a significant role in estimating the Young's modulus of the alloys [29]. While the thermodynamic parameters, λ and Ω, exert a considerable influence on the Young's modulus predictions, we note that the contribution of the entropy of mixing is rather insignificant as it provides no fundamental measure for the bond strengths. The mean lattice constant obtained from the rule of mixture, has negligible effect on Young's modulus.

In summary, we employ available data and experimental measurements to construct machine-learned correlations to predict the Young's modulus and crystallographic phase of low-, medium- and high-entropy alloys. Thermodynamic, geometric and phenomeno-

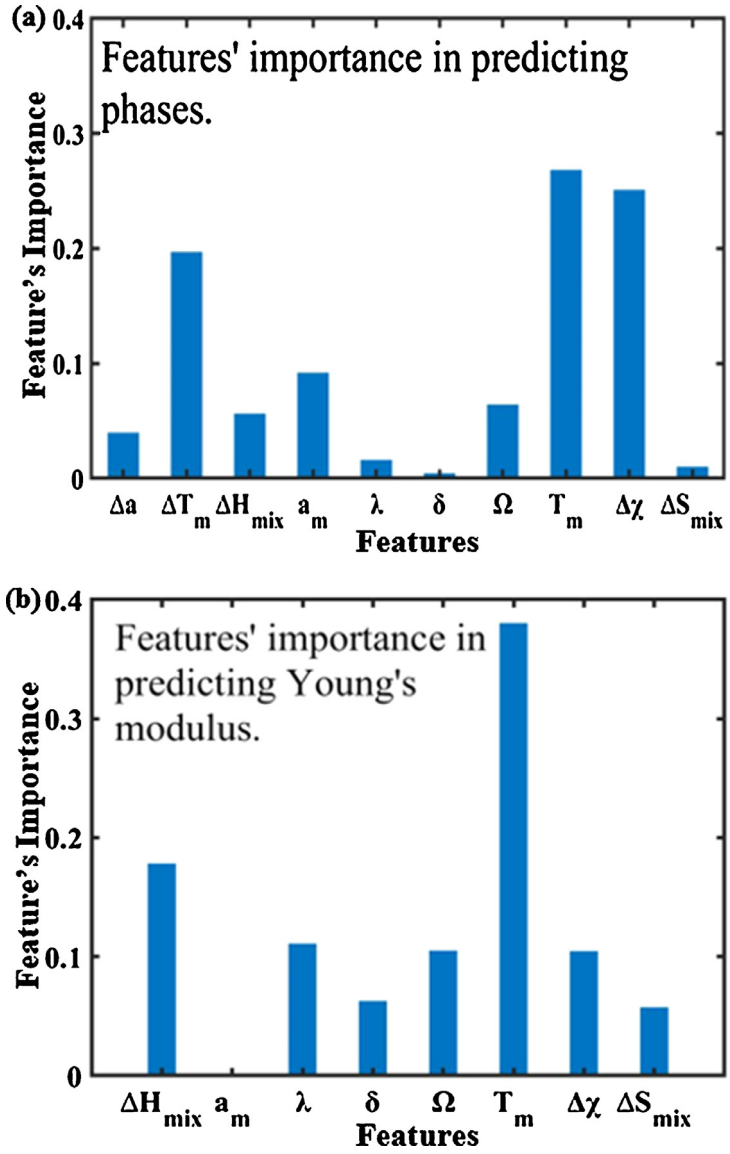

Fig 3. The relative importance of the different features for **(a)** predicting crystallographic phases reveals T_m, $\Delta\chi$ and ΔT_m as the pivotal features, while in **(b)** predicting Young's modulus, T_m and ΔH_{mix} emerge as the crucial ones.

logical parameters are chosen as features and the final models are built to enable predictions for 26 equiatomic alloys in Mo-Ta-Ti-W-Zr elemental family. There is good agreement between the model predictions and the experimental results, which corroborate that additional material features, apart from the Hume-Rothery parameters, contribute significantly in predicting the crystallographic phase and Young's modulus of low-, medium- and high-entropy alloys. A key insight lies in the insignificant effect of the entropy of mixing as a feature for structure and property predictions, even for the HEAs.

Data availability

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

Acknowledgements

The research was supported, in part, by the Office of Naval Research (ONR) through the award N00014-18–1–2484, the Ames Laboratory through the U.S. Department of Energy (DOE), Office of Energy Efficiency and Renewable Energy, Advanced Manufacturing Office (AMO) under design project WBS 2.1.0.19, and the NSF Graduate Fellowship Program Award #1842163. Ames Laboratory is operated by Iowa State University for the U.S. DOE under contract DE-AC02-07CH11358.

References

[1] S. Gorsse, M.H. Nguyen, O.N. Senkov, D.B. Miracle, Data Brief 21 (2018) 2664–2678.
[2] O.N. Senkov, D.B. Miracle, K.J. Chaput, J.-P. Couzinie, J. Mater. Res. 33 (19) (2018) 3092–3128.
[3] P. Singh, A. Sharma, A.V. Smirnov, M.S. Diallo, P.K. Ray, G. Balasubramanian, D.D. Johnson, NPJ Comput. Mater. 4 (1) (2018) 16.
[4] Y. Zhang, T.T. Zuo, Z. Tang, M.C. Gao, K.A. Dahmen, P.K. Liaw, Z.P. Lu, Prog. Mater Sci. 61 (2014) 1–93.
[5] J.M. Rickman, H.M. Chan, M.P. Harmer, J.A. Smeltzer, C.J. Marvel, A. Roy, G. Balasubramanian, Nat. Commun. 10 (1) (2019) 2618.
[6] J.-W. Yeh, S.-K. Chen, S.-J. Lin, J.-Y. Gan, T.-S. Chin, T.-T. Shun, C.-H. Tsau, S.-Y. Chang, Adv. Eng. Mater. 6 (5) (2004) 299–303.
[7] B. Cantor, I.T.H. Chang, P. Knight, A.J.B. Vincent, Mater. Sci. Eng. A 375-377 (2004) 213–218.
[8] A. Gali, E.P. George, Intermetallics 39 (2013) 74–78.
[9] D.J.M. King, S.C. Middleburgh, A.G. McGregor, M.B. Cortie, Acta Mater. 104 (2016) 172–179.

[10] A. Takeuchi, A. Inoue, Mater. Trans. 46 (12) (2005) 2817–2829.
[11] Yang, X. and Y. Zhang, Prediction of High-Entropy Stabilized Solid-Solution in Multi-Component Alloys. Vol. 132. 2012. 233–238.
[12] C. Kittel, P. McEuen, P. McEuen, Introduction to Solid State Physics, 8, Wiley, New York, 1996 Vol..
[13] S. Guo, C.T. Liu, Prog. Nat. Sci. Mater. Int. 21 (6) (2011) 433–446.
[14] A.K. Singh, N. Kumar, A. Dwivedi, A. Subramaniam, Intermetallics 53 (2014) 112–119.
[15] D. Michie, D.J. Spiegelhalter, C. Taylor, Neural Stat. Classif. 13 (1994) 1–298.
[16] A. Chandra, H. Chen, X. Yao, Multi-Objective Machine Learning, Springer, 2006, pp. 429–464.
[17] T. Kluyver, B. Ragan-Kelley, F. Pérez, B.E. Granger, M. Bussonnier, J. Frederic, K. Kelley, J.B. Hamrick, J. Grout, S. Corlay, in: Proceedings of the ELPUB, 2016.
[18] J.M. Rickman, NPJ Comput. Mater. 4 (1) (2018) 5.
[19] W. Hume-Rothery, H.M. Powell, Z. für Krist. Cryst. Mater. 91 (1-6) (1935) 23–47.
[20] M.G. Poletti, L. Battezzati, Acta Mater. 75 (2014) 297–306.
[21] O.N. Senkov, J.M. Scott, S.V. Senkova, D.B. Miracle, C.F. Woodward, J. Alloys Compd. 509 (20) (2011) 6043–6048.
[22] L. Pauling, J. Am. Chem. Soc. 54 (9) (1932) 3570–3582.
[23] S. Guo, C. Ng, J. Lu, C.T. Liu, J. Appl. Phys. 109 (10) (2011) 103505.
[24] J.D. Kelleher, B.M.N., A D'arcy, Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Worked Examples, and Case Studies. 2015.
[25] J.H. Friedman, Ann. Stat. 29 (5) (2001) 1189–1232.
[26] T.D., M. Sadki, E. Bron, U. König, G. Nénert, Powder Diffract. 29 (2014).
[27] W.C. Oliver, G.M. Pharr, J. Mater. Res. 7 (6) (1992) 1564–1583.
[28] T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Springer Science & Business Media, 2009.
[29] B.S. Murty, J.-.W. Yeh, S. Ranganathan, P. Bhattacharjee, High-Entropy Alloys, Elsevier, 2019.
[30] S. Fang, X. Xiao, L. Xia, W. Li, Y. Dong, J. Non Cryst. Solids 321 (1) (2003) 120–125.
[31] O.N. Senkov, G.B. Wilks, D.B. Miracle, C.P. Chuang, P.K. Liaw, Intermetallics 18 (9) (2010) 1758–1765.