Machine-learning model for predicting phase formations of high-entropy alloys

Yao Li and Wanlin Guo[*]

Key Laboratory for Intelligent Nano Materials and Devices of MOE and State Key Laboratory of Mechanics and Control for Mechanical Structures, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China

The phase formations of high-entropy alloys (HEAs) are essential to their properties, but efficient prediction of them remains a challenge. In this work, we propose a support vector machine model that is capable of distinguishing stable body-centered cubic (BCC), face-centered cubic (FCC) HEA phases, and the remaining phases out of the 322 as-cast samples with a cross validation accuracy over 90% after training and test. With the model, we predicted 369 FCC and 267 BCC equiatomic HEAs from the composition space of 16 metallic elements, one order larger than the number of available experimental data. Furthermore, dozens of refractory HEAs with high ratios of melting temperature to density have been screened out. Eleven of them agree with recent experiments and the 20 quinary ones with highest melting temperatures are validated through first-principles calculations. The proposed model is complementary to the calculation of phase diagrams and *ab initio* methods and could serve as an effective guide for designing new HEAs.

I. INTRODUCTION

High-entropy alloys (HEAs) are a unique class of multi-principal element alloys (MPEAs), which can crystallize as single-phase solid solutions [1,2]. Since being proposed in 2004, HEAs have aroused wide interest, not only for their superior properties [3–15] but also for opening up a vast new realm of alloy designs [16–18]. Up to now, an extremely small fraction of the composition space of MPEAs has been studied mainly through trial-and-error approaches and only about 100 HEA compositions have been discovered. With more than 80 metallic elements in the periodic table, it poses a great challenge to recognize phases of MPEAs and find all possible HEA compositions efficiently.

Originally, the high configurational entropy S_C was thought as a criterion of forming single-phase solid solutions. Later, it was found that HEAs cannot be solely identified by S_C and many other theoretical parameters and criteria have been proposed [18]. Zhang *et al.* showed that small mixing enthalpy $\triangle H$ and atomic size difference r_d would be conducive to formations of solid solutions [19]. Guo *et al.* found that when *d*-electrons in the valence band are included, valence electron concentration (VEC) can be used to differentiate body-centered cubic (BCC) and face-centered cubic (FCC) phases [20]. Analogy to the Hume-Rothery rule for binary alloys [21], the electronegativity difference χ_d has been employed by Guo *et al.* [22]. More comprehensive parameters $\Omega = T_m S_C / |\triangle H|$ and $\phi = (S_C - |\triangle H|/T_m)/|S_E| T_m$ were proposed based on thermodynamic analyses [23,24], where T_m is the melting temperature and S_E is the excessive entropy of mixing. Recently, Troparevsky *et al.* employed the formation enthalpies of binary intermetallic compounds H_{IM} to predict the combinations of elements forming single phase [25]. As

new experimental results accumulate, it is found that simply using these parameters to identify HEAs is not sufficiently robust [26]. There is still no rule for assessing, selecting or integrating appropriate parameters.

Alternatively, the calculation of phase diagrams (CAL-PHAD) method and first-principles calculations have been employed to calculate phase stability and prove more robust [27–31]. Using high-throughput CALPHAD calculations, Senkov *et al.* screened out more than 1000 MPEA compositions promising for high-temperature structural applications [17,32]. Tancret *et al.* predicted 62 quinary HEAs with BCC phase through combining CALPHAD method and Gaussian process [33]. Employing a high-throughput *ab initio* method called "LTVC," Lederer *et al.* assessed the solid solution forming ability of 1110 quaternary MPEAs and 130 quinary MPEAs [34]. However, limited by the extent of thermodynamic databases, CALPHAD method is inadequate for alloys constituted by more than five elements and limited to small numbers of elements [26]. First-principles calculations require no experimental input, but they are computationally expensive for MPEAs [35]. Therefore, a new strategy has to be considered to predict phases of HEAs.

Machine learning (ML) has shown advantage in extracting implicit relations from existing data. Since AlphaGo defeated the champion human players in the Go game [36], ML has been recognized as an effective approach to solve complex physical problems including classifying phases. For example, artificial neural networks (ANNs) have been employed to identify phase transitions in a variety of condensed-matter Hamiltonians [37,38]. Recently, Islam *et al.* trained an ANN to identify solid solutions [39]. Due to a large number of parameters from network structures, an ANN model usually requires thousands of labeled samples to achieve satisfied performance. Support vector machine (SVM), another famous ML algorithm, requires less training samples and has many inherent advantages, including sound theoretical basis,

---

*wlguo@nuaa.edu.cn

convexity, excellent generalization ability, and high computation speed [40]. However, few works employed SVM to learn phases of matter.

In this work, we build a SVM model that can distinguish single BCC, single FCC, and the remaining phases of MPEAs out of an experimental data set containing 322 as-cast samples. After training and test, the model manages to achieve a cross-validation (CV) accuracy over 90%. Applying the model to a composition space of 16 elements, we predicted 369 FCC and 267 BCC HEA compositions and compared the predicted phases with the results of CALPHAD and *ab initio* methods. Based on the assessments of melting temperature and density, dozens of refractory HEAs have been further screened out, 11 of which agree with recent experiments. The 20 quinary refractory HEAs with highest melting temperatures have been further validated through first-principles calculations.

II. METHODS

The SVM model is built using the scikit-learn library [41]. The radical basis function $k(x_i, x_j) = e^{-G|x_i - x_j|^2}$ is employed as kernel function, where x_i (x_j) is the feature of ith (jth) samples and G is the kernel coefficient. A penalty parameter C is adopted for model training. Using the kernel, input data can be mapped into a high- or even infinite-dimensional space and then, hyperplanes and supporting vectors are constructed and optimized to distinguish different phases. The one-versus-one method is used for multiclass classification [42]. The employed data set containing 322 as-cast MPEA samples with 90% used for training and the remaining 10% for testing. The 10-fold CV is employed to assess the performance of SVM model: randomly splitting the training set into 10 disjoint subsets and in turns taking nine of them for training with the remaining one for validation. So, the training (cross validation) accuracy is defined as the average value of the 10 training (cross validation) accuracies. G and C are optimized through grid search method: taking the values of G and C as 2^i and 2^j ($-4 \leqslant i \leqslant 7$ and $0 \leqslant j \leqslant 11$), respectively, and selecting the pair of i and j achieving highest validation accuracy.

The thermodynamic parameters of MPEAs are employed as candidate features, which can be calculated by the following equations:

$$S_C = -R \sum_{i=1}^{N} c_i \ln c_i, \tag{1}$$

$$\triangle H = 4 \sum_{i=1, i \neq j}^{N} \triangle H_{ij}^{\text{liq}} c_i c_j, \tag{2}$$

$$T_m = \sum_{i=1}^{N} c_i T_{mi}, \tag{3}$$

$$\text{VEC} = \sum_{i=1}^{N} c_i \text{VEC}_i, \tag{4}$$

$$r_d = \sqrt{\sum_{i=1}^{N} c_i \left(1 - r_i / \sum_{j=1}^{N} c_j r_j\right)^2}, \tag{5}$$

$$\chi_d = \sqrt{\sum_{i=1}^{N} c_i \left(\chi_i - \sum_{j=1}^{N} c_j \chi_j\right)^2}, \tag{6}$$

where c_i, r_i, T_{mi}, VEC_i, and χ_i are molar fraction, atomic radius, melting point, valence electron concentration, and Pauling electronegativity of the ith element, respectively [18]. N is the number of constituent elements and R is the gas constant. H_{ij}^{liq} is the mixing enthalpy of binary liquid alloys derived from Miedema's theory [43,44]. The formation enthalpy of most stable binary compound formed by constituent elements of MPEA samples, H_{IM}, can be derived from Ref. [25] and is also took as a candidate feature. The features of all samples are normalized:

$$P_i^{\text{norm}} = (P_i - P_{\text{min}})/(P_{\text{max}} - P_{\text{min}}), \tag{7}$$

where P_{max} and P_{min} are the maximum and minimum values of feature P, respectively. Both forward and backward wrapper models are used for feature selection [45]. For each feature subset, G and C are optimized.

The SSOS-3×5 method was employed for modeling BCC and FCC quinary HEAs [35]. The formation enthalpies of BCC and FCC structures were calculated through density functional theory as implemented in the VASP code [46,47]. The projector augmented wave method and the LDA exchange-correlation functional were employed for structural relaxations [48–50]. The energy and force converge to 10^{-5} eV and 10^{-2} eV/Å, respectively, with an energy cutoff of 450 eV and a high k-point density ($> 70/\text{Å}^{-3}$).

III. RESULTS AND DISCUSSION

Here, we choose the data set from Ref. [18], which collects 408 experimentally synthesized MPEAs with 648 microstructure characterizations. We only consider the most common as-cast samples to exclude the effects of processing conditions and thermo-mechanical treatments. Samples containing rare earth elements or light elements (except Al) and samples with uncertain phases are excluded to reduce the complexity. Close-packed hexagonal HEAs are not considered for they usually contain rare earth elements.

Then, we obtained a data set of 322 distinct as-cast MPEA samples with 3–9 constituent elements, as list in Table SI in the Supplemental Material [51]. Therein, 18 samples have single BCC phase and 43 samples have single FCC phase. The remaining 261 samples with multiple phases, intermetallic compounds, or amorphous phase are labeled by "NSP" representing "not forming single-phase solid solution." The percentages of ternary to nonary MPEAs and the proportions of BCC, FCC, and NSP phases are shown in Fig. 1(a). Most of the samples consist of five or six elements and the single-phase HEAs only account for a small fraction. There are 18 elements (Al, Ti, V, Cr, Mn, Fe, Co, Ni, Cu, Y, Zr, Nb, Mo, Ag, Hf, Ta, W, and Au) in the data set and their occurrence numbers are shown in Fig. 1(b). We find that Fe, Ni, Cr, Co, and Al occur in more than 200 samples, while Hf, Ag, and Au just occur once.

We calculated the features of 322 samples and plotted them in r_d-versus-$\triangle H$ and Ω-versus-VEC feature spaces [see Figs. 1(c) and 1(d)]. In r_d-versus-$\triangle H$ space, it is found that the samples are highly overlapped and the phases are undistinguishable. In the Ω-versus-VEC space, it seems that VEC could be a good indicator distinguishing BCC and FCC
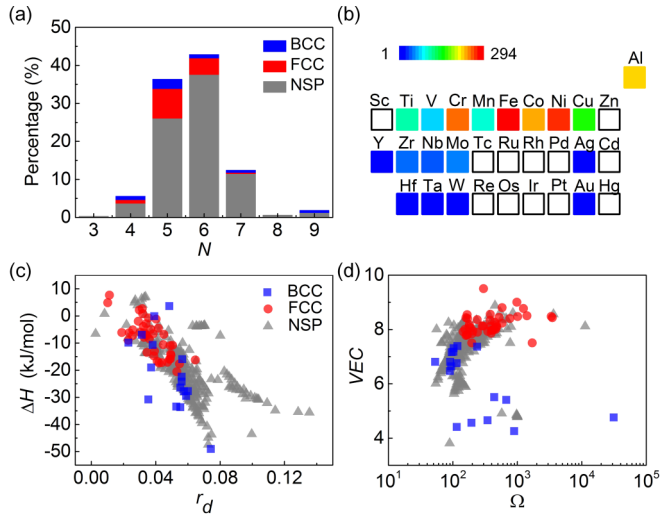
FIG. 1. Statistics of the 322 as-cast MPEA samples in the data set. (a) The percentages of ternary to nonary MPEAs and the proportions of different phases. (b) The occurrence numbers of different elements. The blank squares mark the elements not in the data set. Samples plotted in (c) r_d-vs.-$\triangle H$ and (d) Ω-vs.-VEC feature spaces.

samples, but it still cannot distinguish single-phase samples and NSP samples.

Next, we built a SVM model using the above samples and candidate features. Figure 2(a) shows the procedure of forward feature search based on the wrapper model. The feature set starts with an empty set. Then, candidate features are put into the feature set one by one. When containing one feature, the SVM model with VEC achieves the highest CV

FIG. 2. (a) The process of forward feature search. (b) The optimization of the penalty parameter C and the kernel coefficient G for feature set {VEC, r_d, T_m, $\triangle H$, S_C} through grid search method. The color bars in panels (a) and (b) display the validation accuracies. The variations of training, cross validation and test accuracies in the processes of feature selections using (c) the original data set or (d) the SMOTE data set. The upper (lower) triangles represent the training (cross validation) accuracies and the circles represent the test accuracies.

accuracy of 84.83% and the configurational entropy S_C has the lowest CV accuracy of 81.38%. Compared with the strategy of plotting in VEC feature space, the SVM model is more accurate [see Fig. 1(d)]. When searching for second feature, the rest features are in turns combined with VEC and the feature subset {VEC, r_d} achieves the highest accuracies of 85.86%. Similarly, T_m, $\triangle H$, and S_C have been successively selected out. When containing more features, the validation accuracy decreases due to overfitting. Finally, the feature set {VEC, r_d, T_m, $\triangle H$, S_C} achieves the global highest CV accuracy up to 90.69% and the corresponding training accuracy is up to 96.55% [see Fig. 2(c)]. VEC reflects electronic properties of constituent elements; r_d represents the scale of lattice strain due to different atomic radius; the combination of T_m, $\triangle H$, and S_C describes the thermodynamic properties. The feature selection process here implies that the contributions from electron and lattice strain could be more dominant than thermodynamic effects. The parameters H_{IM} and χ_i reflect the possibilities of forming intermetallics. They have not been selected out for the samples with multiple solid solution phases account for a large proportion of NSP samples in the data set. Characterizing the electronic, strain and thermodynamic features of MPEAs, the feature subset {VEC, r_d, T_m, $\triangle H$, S_C} should give comprehensive assessments of phase formations.

The optimization of parameters C and G for {VEC, r_d, T_m, $\triangle H$, S_C} is shown in Fig. 2(b). The highest CV accuracy achieves at $C = 2^8$ and $G = 2^3$. Figure 2(c) shows the relations between the accuracies and the number of features. All of the training, CV and test accuracies increase at the beginning and then saturate or decrease slightly. At five features, the test accuracy achieves the highest value, further proving that {VEC, r_d, T_m, $\triangle H$, S_C} is the best feature set.

To avoid the influences of unbalanced numbers of BCC, FCC, and NSP samples, a new data set containing 261 BCC, 261 FCC, and 261 NSP samples has been created through the SMOTE method [52]. Using the "SMOTE" data set, 10 times 10-fold CV processes were performed through both forward and backward wrapper models. {VEC, r_d, T_m, $\triangle H$, S_C} and {VEC, r_d, T_m, $\triangle H$, S_C, H_{IM}} have been searched out as best feature set for 9 times. In forward searches, the average training accuracies of {VEC, r_d, T_m, $\triangle H$, S_C} and {VEC, r_d, T_m, $\triangle H$, S_C, H_{IM}} are over 99%. The average CV accuracy of {VEC, r_d, T_m, $\triangle H$, S_C, H_{IM}} (95.83%) is just slightly higher than that of {VEC, r_d, T_m, $\triangle H$, S_C} (95.80%). The average test accuracy reaches up to 94.10% at four features and then saturates or even slightly decreases. To reduce the risk of overfitting, {VEC, r_d, T_m, $\triangle H$, S_C} was chosen as best feature set, showing consistence with the result using the original data set. The results of backward searches are similar as shown in Fig. S1 in the Supplemental Material [51]. All of the training, CV, and test accuracies using the SMOTE data set are much higher than those using the original data set, implying that the influence of unbalanced data here can be clarified. This can also be inferred from the error analyses hereinafter [see Fig. 3(b)]. The SVM model trained by the SMOTE data set has a much higher value of G so it is more likely to suffer from overfitting problem. Therefore, the SVM model with the selected features and optimized parameters were finally trained on the original 322 samples and achieves an accuracy of 96.58%.
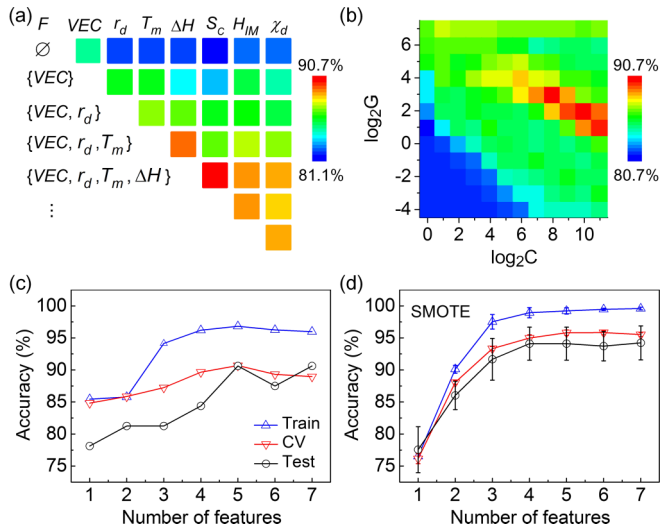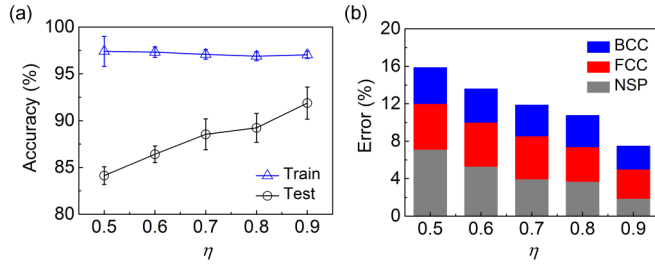
FIG. 3. (a) The variations of training and test accuracies with different η. (b) The constituents of test errors at different η.

To assess the dependency of accuracies on the number of training samples, we define η as the proportion of training samples in total samples and randomly split the original data set with η from 0.5 to 0.9. For each η, we repeated the random splitting process for five times. Figure 3(a) shows the variations of training and test accuracies with different η. When η equals 0.5, the average test accuracy is about 84.13%. With η increasing, the average test accuracy remarkably increases up to 91.88%, while the average training accuracies stay around 97% all the time. The standard deviations of both training and test accuracies are always less than 2%, showing the stability of the model.

From the testing results, we find that the SVM model can always correctly classify BCC and FCC phases and the errors only come from distinguishing FCC or BCC phase with NSP phase. Figure 3(b) shows the average proportions of wrongly

predicted BCC, FCC, and NSP samples in test errors under different η. When η increases, all the three errors decrease. The NSP error decreases much faster than the other two, due to the great number of NSP samples in the data set. To assess the influence of data imbalance, we checked the test accuracies of each phase. For example, when η equals 0.9, the averaged test accuracies of BCC, FCC, and NSP samples are 60%, 75%, and 97.69%, respectively. The BCC and FCC errors would lead to missing some possible HEA compositions. The extremely low NSP error represents that the predicted HEAs have quite high possibilities to be realized experimentally. From the above analyses, we conclude that the SVM model has a high confidence of predicted HEAs, although it may miss some possible compositions in prediction.

The high-performance SVM model can be used to search for new HEAs. We first considered elements in the data set, so we chose the composition space consisting of 16 elements (Al, Ti, V, Cr, Mn, Fe, Co, Ni, Cu, Y, Zr, Nb, Mo Hf, Ta, and W). When N ranges from 3 to 8, there are $\sum_{i=3}^{8} C_{16}^i = 39\,066$ element compositions. With the model, phases of all these candidates are predicted and among them, there are 369 FCC and 267 BCC HEAs containing 4 to 8 constituent elements, as shown in Fig. 4(d) and list in Table SII in the Supplemental Material [51]. When $N = 4$, the ratio of HEAs in the quaternary alloy candidates is about 5.44%. When N increases, the ratio of HEAs decreases due to the increasing possibilities of forming intermetallic components. The number of HEAs continues to increase and reaches the highest value at $N = 6$. Besides, most of the predicted BCC HEAs consist of four
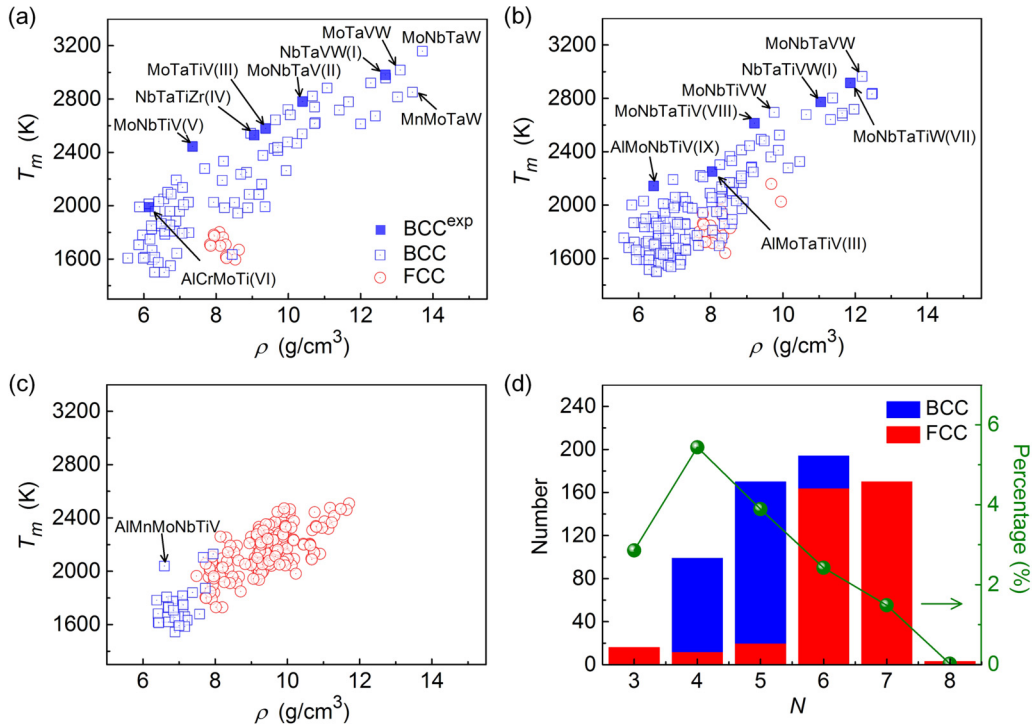
FIG. 4. The melting temperatures and densities of (a) quaternary, (b) quinary, and (c) senary equiatomic HEAs, predicted from the composition spaces of 16 different elements. The squares and circles represent BCC and FCC phases, respectively. The solid squares mark the BCC HEAs not in the data set and validated by recent experiments. (I) Ref. [53], (II) Ref. [54], (III) Ref. [55], (IV) Ref. [56], (V) Ref. [57], (VI) Ref. [58], (VII) Ref. [59], (VIII) Ref. [60], and (IX) Ref. [61]. (d) Statistics of the predicted single-phase MPEAs. The green line shows the ratios of element compositions forming single-phase in the corresponding candidates with N elements.

TABLE I. The 20 quinary refractory HEAs with highest melting temperatures. $\Delta E = E_{BCC} - E_{FCC}$, where E_{BCC} and E_{FCC} are the enthalpies of forming BCC and FCC phases, respectively. The unit of ΔE is eV/atom.

Alloy	ΔE	Alloy	ΔE
MoNbTaVW	−0.11703	MnMoTaTiW	−0.11474
MoNbTaTiW	−0.10939	MnNbTaTiW	−0.07252
HfNbTaTiW	−0.1065	MoNbTaTiV	−0.08532
MnMoNbTaW	−0.126	HfNbTaTiZr	−0.04927
MoTaTiVW	−0.08773	AlMoNbVW	−0.11295
NbTaTiVW	−0.10592	MnMoNbTaTi	−0.12461
MnMoTaVW	−0.10477	AlMoNbTiW	−0.09257
MoNbTiVW	−0.09808	AlMoNbTaV	−0.11182
MnNbTaVW	−0.0725	AlTaTiVW	−0.08474
HfMoNbTaTi	−0.07351	AlMoNbTaTi	−0.09446

or five elements, while most of the FCC equiatomic HEAs consist of six or seven elements.

Some of the predicted HEAs have practically important properties, e.g., light-weight and high-temperature resistant. The density ρ can be estimated by

$$\rho = \sum_{i=1}^{N} c_i \rho_i, \tag{8}$$

where ρ_i is the density of the ith element.

The melting temperatures and densities of predicted BCC and FCC HEAs are plotted in Figs. 4(a)–4(c) (also list in Table SII in the Supplemental Material [51]). We can find that there are some scattered quaternary and quinary BCC HEA points having both significantly high melting temperatures and high ratios of melting temperature to density. These BCC HEAs contain refractory elements e.g. Mo, Nb, Ta, V, and W, so have been called "refractory high-entropy alloys" [7]. Dozens of refractory HEAs in outside of the data set have been predicted. Among them, MoNbTaV, MoNbTiV, MoTaTiV, AlCrMoTi, NbTaTiZr, NbTaVW, MoNbTaTiV, MoNbTaTiW, NbTaTiVW, AlMoNbTiV, and AlMoTaTiV agree with recent experiments [see solid squares in Figs. 4(a) and 4(b)] [53–61]. Some Mn-based refractory HEAs have also been predicted but have rarely been experimentally studied. When containing more than five elements, the BCC and FCC HEA points are all clustered and no refractory HEAs have been screened out (see Fig. S4 in the Supplemental Material for details [51]). We noted that several refractory HEAs containing Hf have not been screened out, e.g., HfNbTaZr, HfMoTaTiZr, and HfMoNbTaTiZr [62,63]. This should be attributed to the shortage of Hf samples in the data set and the error properties of SVM model.

It is daunting to calculate properties of all the predicted HEAs through first-principles calculations and the SSOS method [35]. Therefore, calculations were carried out for the 20 quinary refractory HEAs with highest melting temperatures, all of which are predicted to form BCC phase. As listed in Table I, for all of the 20 element compositions, the enthalpies of forming BCC phase E_{BCC} are less than the enthalpies of forming FCC phase E_{FCC}. The energy
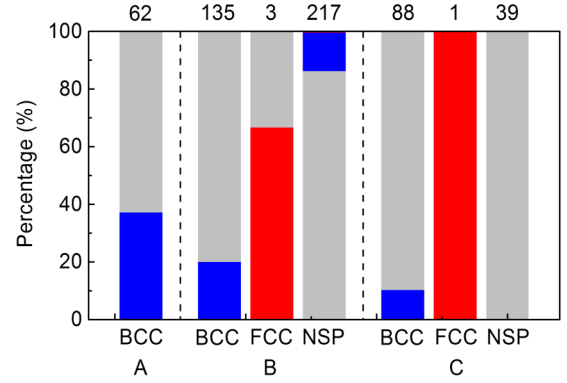

FIG. 5. Comparisons between the phases of MPEAs predicted by SVM model and studied by other methods. The columns present different phases in different data sets, while the colors blue, red and gray represent BCC, FCC, and NSP phases predicted by the SVM model, respectively. The numbers of different phases are indicated at the top of columns.

differences range from −0.04927 to −0.126 eV/atom, showing that BCC phase is more stable than FCC phase even in high temperatures.

The predictions were compared with previous results based on CALPHAD and *ab initio* methods, as shown in Fig. 5. The data set A contains the 62 quinary BCC HEAs predicted through the combination of CALPHAD method and Gaussian process [33]. The data set B and C contain 355 quaternary and 128 quinary MPEAs, respectively, screened out from the composition space of 16 elements through the LTVC method [34]. The data set A and C contain 27 same element compositions, but 7 of them are identified as different phases. So, results from different methods could have divergences. Most of the element compositions forming NSP phase in data set B and C are correctly predicted by our SVM model. The model seems more cautious when predicting HEAs, for many BCC HEAs in data set A and C are predicted to form NSP phase. The phases of samples in the data set are labeled through characterizations of x-ray diffraction and scanning electron microscope in room temperature, while other methods have different criteria of single-phase formation, e.g., using a temperature range between 700 and 1300°C to assess phase stabilities [33]. Another reason could be the shortage of BCC HEA samples for certain elements in the data set. There are also many HEA compositions predicted by our SVM model have never been studied, should be helpful for future experimental investigation. With extremely high computation speed and improving reliability with accumulating samples, the SVM model is complementary to current CALPHAD and *ab initio* methods. Using the SVM model in cooperation with other methods could achieve higher confidence in guiding experimental designs.

We have also considered the composition space containing more elements (Sc, Zn, Ru, Rh, Pd, Ag, Cd, Re, Os, Ir, Pt, and Au in addition; see Fig. S2 in the Supplemental Material [51]). The number of predicted HEA compositions increases remarkably. As 10 elements are in outside of the data set, these predictions are reserved for future assessments and verifications.

Although it has achieved 96.55% and 90.69% training and validation accuracies, respectively, the performance of the SVM model can still be improved. As the accumulation of experimental data, the generalization properties can be further improved and the support vectors between different phases can be further optimized. With the increasing capabilities of computers, more accurate method will be applied to calculate the features, e.g., calculating the entropy and enthalpy through the density functional theory. Besides, more effective features could be found and assessed through the feature searching process, which would strengthen the understanding of phase formations of HEAs. For future developments, we will include the effects of processing conditions, thermo-mechanical treatments and temperatures into the SVM model and consider more complex microstructures, e.g., dual-phase structures.

IV. CONCLUSION

In conclusion, we build a SVM model to predict phase formations of HEAs. A data set containing 322 as-cast MPEA samples is built for training and testing based on available experimental data. Five physical parameters—valence electron concentration, atomic size difference, melting temperature, mixing enthalpy, and configurational entropy—are selected out as features through both forward and backward wrapper models. The optimal SVM model can differentiate single

BCC, single FCC, and the remaining phases of MPEAs with an accuracy over 90%. With the model, 369 FCC and 267 BCC equiatomic HEAs are predicted from the composition space of 16 metallic elements, one order larger than the number of available experimental data. Among them, dozens of refractory high-entropy alloys have been further screened out based on assessments of melting temperatures and densities. Eleven of these refractory HEAs are validated by recent experiments and the 20 quinary ones with highest melting temperatures are validated through first-principles calculations. By comparison, it is found that the SVM model is complementary to current CALPHAD and *ab initio* methods. The proposed SVM model could serve as an effective guide for designing new HEAs and would inspire new applications of machine learning in material science.

[1] J.-W. Yeh, S.-K. Chen, S.-J. Lin, J.-Y. Gan, T.-S. Chin, T.-T. Shun, C.-H. Tsau, and S.-Y. Chang, Adv. Eng. Mater. **6**, 299 (2004).

[2] B. Cantor, I. Chang, P. Knight, and A. Vincent, Mater. Sci. Eng. A **375**, 213 (2004).

[3] B. Gludovatz, A. Hohenwarter, D. Catoor, E. H. Chang, E. P. George, and R. O. Ritchie, Science **345**, 1153 (2014).

[4] M. A. Hemphill, T. Yuan, G. Wang, J. Yeh, C. Tsai, A. Chuang, and P. Liaw, Acta Mater. **60**, 5723 (2012).

[5] Z. Li, K. G. Pradeep, Y. Deng, D. Raabe, and C. C. Tasan, Nature (London) **534**, 227 (2016).

[6] D. Li, C. Li, T. Feng, Y. Zhang, G. Sha, J. J. Lewandowski, P. K. Liaw, and Y. Zhang, Acta Mater. **123**, 285 (2017).

[7] O. N. Senkov, G. Wilks, D. Miracle, C. Chuang, and P. Liaw, Intermetallics **18**, 1758 (2010).

[8] O. N. Senkov, G. Wilks, J. Scott, and D. B. Miracle, Intermetallics **19**, 698 (2011).

[9] C. Liu, H. Wang, S. Zhang, H. Tang, and A. Zhang, J. Alloys Compd. **583**, 162 (2014).

[10] C. Lu, L. Niu, N. Chen, K. Jin, T. Yang, P. Xiu, Y. Zhang, F. Gao, H. Bei, S. Shi *et al.*, Nat. Commun. **7**, 13564 (2016).

[11] P. Koželj, S. Vrtnik, A. Jelen, S. Jazbec, Z. Jagličić, S. Maiti, M. Feuerbacher, W. Steurer, and J. Dolinšek, Phys. Rev. Lett. **113**, 107001 (2014).

[12] M. Lucas, L. Mauger, J. Munoz, Y. Xiao, A. Sheets, S. Semiatin, J. Horwath, and Z. Turgut, J. Appl. Phys. **109**, 07E307 (2011).

[13] O. Schneeweiss, M. Friák, M. Dudová, D. Holec, M. Šob, D. Kriegner, V. Holý, P. Beran, E. P. George, J. Neugebauer *et al.*, Phys. Rev. B **96**, 014437 (2017).

[14] R. Lizárraga, E. Holmström, and L. Vitos, Phys. Rev. Mater. **2**, 094407 (2018).

[15] F. O. von Rohr and R. J. Cava, Phys. Rev. Mater. **2**, 034801 (2018).

[16] Y. Ye, Q. Wang, J. Lu, C. Liu, and Y. Yang, Mater. Today **19**, 349 (2016).

[17] O. N. Senkov, J. Miller, D. Miracle, and C. Woodward, Nat. Commun. **6**, 6529 (2015).

[18] D. B. Miracle and O. N. Senkov, Acta Mater. **122**, 448 (2017).

[19] Y. Zhang, Y. J. Zhou, J. P. Lin, G. L. Chen, and P. K. Liaw, Adv. Eng. Mater. **10**, 534 (2008).

[20] S. Guo, C. Ng, J. Lu, and C. T. Liu, J. Appl. Phys. **109**, 103505 (2011).

[21] U. Mizutani, *Surface Properties And Engineering Of Complex Intermetallics* (World Scientific, Singapore, 2010), pp. 323–399.

[22] S. Guo and C. T. Liu, Prog. Nat. Sci.: Mater. Int. **21**, 433 (2011).

[23] X. Yang and Y. Zhang, Mater. Chem. Phys. **132**, 233 (2012).

[24] Y. F. Ye, Q. Wang, J. Lu, C. T. Liu, and Y. Yang, Scr. Mater. **104**, 53 (2015).

[25] M. C. Troparevsky, J. R. Morris, P. R. C. Kent, A. R. Lupini, and G. M. Stocks, Phys. Rev. X **5**, 011041 (2015).

[26] M. C. Gao, C. Zhang, P. Gao, F. Zhang, L. Ouyang, M. Widom, and J. Hawk, Curr. Opin. Solid State Mater. Sci. **21**, 238 (2017).

[27] A. Abu-Odeh, E. Galvan, T. Kirk, H. Mao, Q. Chen, P. Mason, R. Malak, and R. Arróyave, Acta Mater. **152**, 41 (2018).

[28] C. Chattopadhyay, A. Prasad, and B. Murty, Acta Mater. **153**, 214 (2018).

[29] A. J. S. F. Tapia, D. Yim, H. S. Kim, and B.-J. Lee, Intermetallics **101**, 56 (2018).

[30] J. M. Sanchez, I. Vicario, J. Albizuri, T. Guraya, and J. C. Garcia, J. Mater. Res. Technol. **8**, 795 (2019).

[31] F. Tian, L. Delczeg, N. Chen, L. K. Varga, J. Shen, and L. Vitos, Phys. Rev. B **88**, 085128 (2013).

[32] O. Senkov, J. Miller, D. Miracle, and C. Woodward, Calphad **50**, 32 (2015).

[33] F. Tancret, I. Toda-Caraballo, E. Menou, P. E. J. R. Díaz-Del *et al.*, Mater. Des. **115**, 486 (2017).

[34] Y. Lederer, C. Toher, K. S. Vecchio, and S. Curtarolo, Acta Mater. **159**, 364 (2018).

[35] C. Jiang and B. P. Uberuaga, Phys. Rev. Lett. **116**, 105501 (2016).

[36] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. Van Den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot *et al.*, Nature (London) **529**, 484 (2016).

[37] J. Carrasquilla and R. G. Melko, Nat. Phys. **13**, 431 (2017).

[38] E. P. Van Nieuwenburg, Y.-H. Liu, and S. D. Huber, Nat. Phys. **13**, 435 (2017).

[39] N. Islam, W. Huang, and H. L. Zhuang, Comput. Mater. Sci. **150**, 230 (2018).

[40] C. Cortes and V. Vapnik, Machine Learn. **20**, 273 (1995).

[41] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg *et al.*, J. Mach. Learn. Res. **12**, 2825 (2011).

[42] J. H. Friedman, Technical Report, Statistics Department, Stanford University (1996).

[43] F. D. Boer, R. Boom, W. C. M. Mattens, A. R. Miedema, and A. K. Niessen, in *Cohesion in Metals: Transition Metal Alloys*, edited by F. R. de Boer and D. G. Pettifor, Vol. 1 (North-Holland, Amsterdam, 1988).

[44] R. Zhang, S. Zhang, Z. He, J. Jing, and S. Sheng, Comput. Phys. Commun. **209**, 58 (2016).

[45] R. Kohavi and G. H. John, Artif. Intell. **97**, 273 (1997).

[46] G. Kresse and J. Hafner, Phys. Rev. B **47**, 558 (1993).

[47] G. Kresse and J. Furthmüller, Phys. Rev. B **54**, 11169 (1996).

[48] J. P. Perdew and A. Zunger, Phys. Rev. B **23**, 5048 (1981).

[49] P. E. Blöchl, Phys. Rev. B **50**, 17953 (1994).

[50] G. Kresse and D. Joubert, Phys. Rev. B **59**, 1758 (1999).

[51] See Supplemental Material at http://link.aps.org/supplemental/ 10.1103/PhysRevMaterials.3.095005 for the 322 as-cast samples with microstructures and calculated features, the 10 times 10-fold cross validations using the SMOTE data set through both forward and backward wrapper models, and the predicted HEAs from the composition spaces of 16 elements with estimated melting temperatures.

[52] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, J. Artif. Intell. Res. **16**, 321 (2002).

[53] H. Yao, J. Qiao, M. Gao, J. Hawk, S. Ma, H. Zhou, and Y. Zhang, Mater. Sci. Eng. A **674**, 203 (2016).

[54] H. Yao, J.-W. Qiao, M. C. Gao, J. A. Hawk, S.-G. Ma, and H. Zhou, Entropy **18**, 189 (2016).

[55] D. X. Qiao, H. Jiang, X. X. Chang, Y. P. Lu, and T. J. Li, in *Materials Science Forum*, Vol. 898 (Trans Tech Publications, Zurich, 2017), pp. 638–642.

[56] M. Todai, T. Nagase, T. Hori, A. Matsugaki, A. Sekita, and T. Nakano, Scr. Mater. **129**, 65 (2017).

[57] S. Y. Chen, X. Yang, K. A. Dahmen, P. K. Liaw, and Y. Zhang, Entropy **16**, 870 (2014).

[58] H. Chen, A. Kauffmann, S. Laube, I.-C. Choi, R. Schwaiger, Y. Huang, K. Lichtenberg, F. Müller, B. Gorr, H.-J. Christ *et al.*, Metall. Mater. Trans. A **49**, 772 (2018).

[59] Z. Han, N. Chen, S. Zhao, L. Fan, G. Yang, Y. Shao, and K. Yao, Intermetallics **84**, 153 (2017).

[60] H. Yao, J. Qiao, J. Hawk, H. Zhou, M. Chen, and M. Gao, J. Alloys Compd. **696**, 1139 (2017).

[61] M. Zhang, X. Zhou, and J. Li, J. Mater. Eng. Perform. **26**, 3657 (2017).

[62] S. Maiti and W. Steurer, Acta Mater. **106**, 87 (2016).

[63] C.-C. Juan, M.-H. Tsai, C.-W. Tsai, C.-M. Lin, W.-R. Wang, C.-C. Yang, S.-K. Chen, S.-J. Lin, and J.-W. Yeh, Intermetallics **62**, 76 (2015).