



# Learning phase selection and assemblages in High-Entropy Alloys through a stochastic ensemble-averaging model

D. Beniwal <sup>a,b</sup>, P.K. Ray <sup>a,\*</sup>

<sup>a</sup> Department of Metallurgical and Materials Engineering, Indian Institute of Technology Ropar, Rupnagar 140001, Punjab, India

<sup>b</sup> Department of Mechanical Engineering, University of Petroleum and Energy Studies, Dehradun 248007, Uttarakhand, India



## ARTICLE INFO

### Keywords:

High-Entropy Alloys  
Phase selection  
Phase fractions  
Concentrated solid solutions  
Materials informatics  
Machine learning

## ABSTRACT

Efficient alloy design requires a knowledge of phase selection, phase-fractions and microstructure of the material. Prediction of the phase equilibria, though well-established for traditional alloys, is still an open challenge for novel materials and concentrated multicomponent systems such as the High-Entropy Alloys (HEAs). In this paper, we present a novel data-driven approach for learning the phase-equilibria in HEAs through the use of a stochastic ensemble averaging method. The proposed model employs a model-averaging technique on an ensemble of 150 artificial neural networks that have been trained on a 323 alloy dataset. A seven-label classification is presented using a three-element vector description of phases. This allows us to fit the classification boundaries to a relatively larger number of phase combination labels while using only three target parameters for training thereby improving accuracy. The phase prediction capabilities (i.e., formation of FCC, BCC, intermetallics or different combinations of these phases) were tested on 320 alloys outside of the training dataset. Additionally, quantitative estimate of the phase equilibria, i.e., relative phase-fractions, were estimated and compared with experimental measurements and CALPHAD predictions in three different high-entropy systems, viz.,  $\text{Fe}_x\text{-Ni}_y\text{-}(\text{AlCoCr}_{0.5})_{1-x-y}$ ,  $\text{Al}_x\text{-Ti}_y\text{-}(\text{CrFeNi})_{1-x-y}$  and  $\text{Cr}_x\text{-Moy}\text{-}(\text{VNbTi})_{1-x-y}$ . The model's capability of going a step beyond current state-of-the-art model allows greater insights into target composition spaces for the alloy designer and establishes the first such approach for HEAs.

## 1. Introduction

The recent discovery of complex multi-principal element alloys, also known as the High-Entropy Alloys (HEAs), by Cantor et al. [1] and Yeh et al. [2] has resulted in a paradigm shift in alloy design. While the initial interest in HEAs was driven by the formation of single-phase concentrated solid solution alloys, subsequent efforts on investigating the thermal stability and decomposition behavior of these alloys revealed the potential for designing tunable microstructures with excellent physical and chemical properties which are dependent on the phases formed and their phase fractions [3,4]. Therefore, prediction of phases and their respective volume fractions becomes especially critical for alloy design. While a number of researchers have developed rules and models for predicting whether the phase being formed is a solid solution or an intermetallic (IM) [5,6], very few have actually predicted the crystal structure successfully and none, to our knowledge, have predicted the phase fractions, which provides an additional depth for tailoring alloy properties. Here, we present a machine learning (ML)

model for accelerated phase exploration in single and multi-phase HEAs. Previous attempts [7–12] toward phase classification in HEAs have adopted a non-probabilistic approach wherein the presence or absence of a phase is predicted. Unlike these, we have adopted a probabilistic approach in our model by predicting the probability of occurrence of each phase which may be interpreted as a measure of relative phase stability. This fundamental change in approach allows us to further transform these probabilities into relative phase fraction estimates using a softmax activation function (discussed in §2.3). Also, while the previous attempts have considered three or four-label classification, we have tackled a seven-label classification problem through transformation of phase information into three-element binary vectors for improving accuracy (discussed in §2.2).

Various empirical and semi-empirical design principles have been developed over the last decade to predict formation of SS and IM phases in HEAs based on: a) thermodynamic parameters such as configurational entropy, mixing enthalpy and formation enthalpy, and b) lattice distortion parameters such as size asymmetry, strain energy and

\* Corresponding author.

E-mail address: [pratik@iitrpr.ac.in](mailto:pratik@iitrpr.ac.in) (P.K. Ray).

geometric packing state [13–18]. While these models point toward some significant trends, they suffer from two major drawbacks: a) they do not distinguish between the crystal structure of the SS phase (FCC or BCC), and b) they use very few parameters for fitting classification boundaries to the experimental data which allows better interpretability but inadvertently leads to lower accuracy.

In recent years, various ML models have been reported to guide the phase selection in HEAs. Menou et al. [12] probed stability regions for single SS phases through a genetic algorithm using criteria built through statistical analysis of previously reported empirical and semi-empirical rules. Huang et al. [11] employed three ML algorithms for exploring phase selection rules for SS and IM phase classification and obtained best performance using artificial neural networks (ANNs). Zhou et al. [10] used convolutional neural networks driven by 13 design parameters for predicting formation of SS, IM and amorphous phases in HEAs. Guo et al. [19] explored (statistically) the stability of FCC and BCC phases in HEAs as a function of valence electron concentration (VEC) and observed that a high VEC ( $\geq 8$ ) stabilizes FCC crystal structure. Li and Guo [8] used a support vector machine model to classify HEAs into single phase FCC/BCC alloys and non-single phase (NSP) alloys. While they achieved a cross-validation accuracy of 90%, 81% of the predicted alloys are labeled as NSP, which does not provide further insight into their structure i.e., whether they are (FCC/BCC + IM) or (FCC + BCC) or only IM. Roy et al. [7] employed a Gradient Boost Classification algorithm using ten features for classification of HEAs into single phase FCC/BCC or multiphase, and validated their model experimentally on ternary, quaternary and quinary alloys from Mo-Ta-Ti-W-Zr alloy system. Here, the ‘multiphase’ label represents a variety of phase combinations including (FCC + BCC), (FCC + IM) and (BCC + IM) HEAs, thereby limiting the applicability of the model for targeted alloy design.

Despite the strides made in ML driven phase exploration in HEAs, two significant facets remain largely unexplored and have been addressed in this work:

- A reliable phase classification indicating both the type of SS phase (FCC and/or BCC) and the presence/absence of IM phase for a wide range of HEAs – This is difficult because the classification accuracy usually decreases when the data is split into more labels since the training data per label is reduced. We have overcome this by expressing the phase information as a three-element binary vector which considerably improves accuracy (§2.2).
- Quantitative estimation of phase fractions in HEAs – This has not been attempted before due to the lack of a database wherein phase fractions have been compiled; and in fact, even if compiled, we believe such a database for HEAs will not be very exhaustive due to the limited availability of experimentally measured phase fractions in literature. To overcome this, we have developed a methodology to estimate phase fraction indirectly by using a probabilistic approach, wherein first we predict the probability of occurrence of each phase and then transform these probabilities into relative phase fraction estimates (§2.3).

**Table 1** compares the performance of ML model presented in this work with some of the models reported in recent literature for phase classification in HEAs and illustrates the unique features of the approach presented here vis-à-vis prior efforts reported in the literature.

## 2. Methods

The machine learning strategy used in this work, as shown in Fig. 1, comprises of: a) database preparation and feature engineering (§2.1), b) converting categorical phase information into binary vectors (§2.2), c) training deep ANNs to predict phase fractions and analyzing cross-validation performance (§2.3), d) exploring new alloy systems using an ensemble of trained ANNs (§2.4 and §3.2–3.4), and e) validating predictions based on experimental and CALPHAD results reported in recent literature (§3.2–3.4). We have also revisited some of the previously proposed empirical rules using our expanded dataset and have identified additional parameters that can assist in establishing stability domains for different phases (§3.1).

### 2.1. Dataset creation and feature-engineering

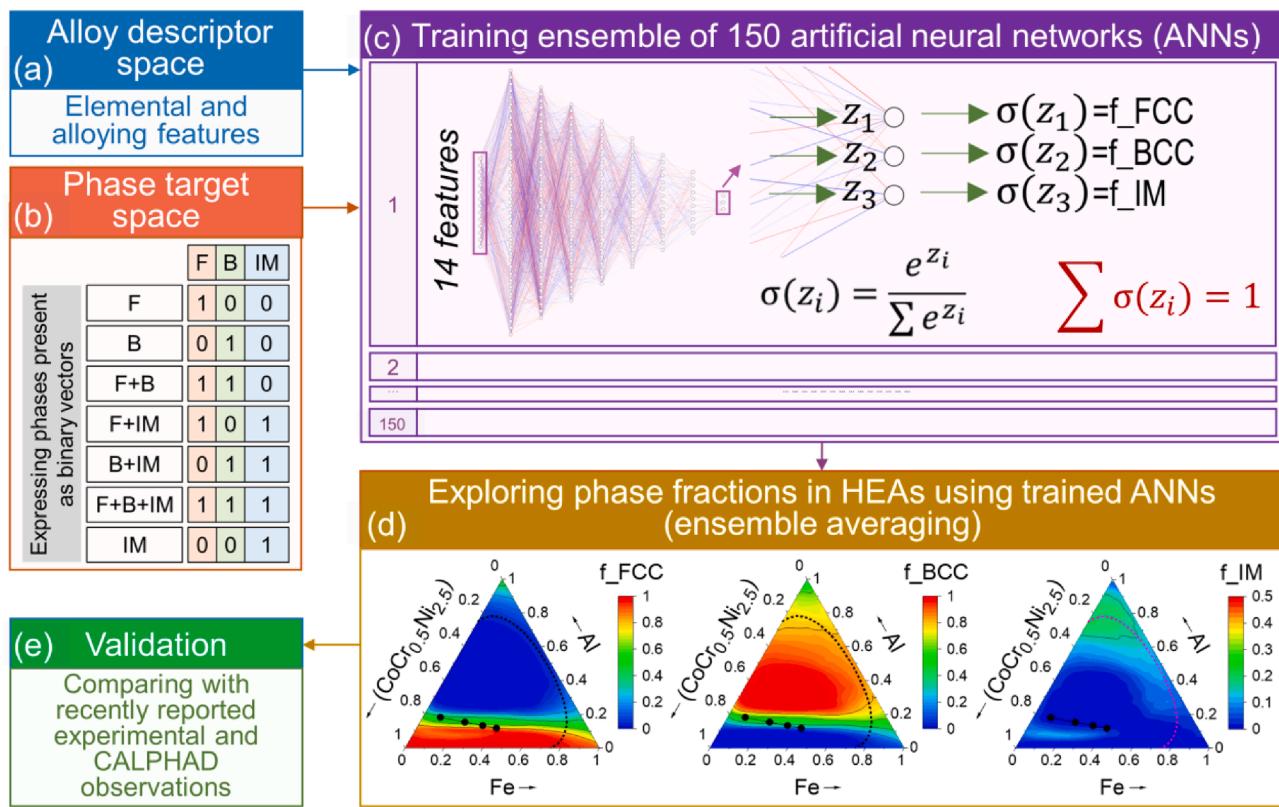
A dataset of 323 as-cast HEAs (listed in Table S4), containing experimentally determined phase information, was compiled through refinement of database reported in literature [21,22]. For comparison, the size of the dataset is similar to that used by Li and Guo (322 alloys) and Roy et al. (329 alloys) in their recent work [7,8]. The dataset spans a compositional space of 24 elements including 3-d transition metals, refractory metals and main-group elements, and the phase distribution spans seven labels: FCC (39 alloys), BCC (117 alloys), FCC + BCC (40 alloys), FCC + IM (50 alloys), BCC + IM (49 alloys), FCC + BCC + IM (21 alloys) and IM (7 alloys).

We explored an expansive set of 18 features (Table 2) comprising: a) composition-averaged and asymmetry metrics built from elemental properties, and b) thermodynamic parameters that are representative of alloying energetics. The binary chemical enthalpies of mixing used here (S. No. 15–18 in Table 2) have been calculated in accordance with Miedema’s model [23,24] and take into account the interfacial surface energy for non-ionic interfaces and dipole-layer energy for ionic interfaces. The binary elastic enthalpy of mixing (S. No. 14 in Table 2) has been calculated using the classical elasticity method by Eshelby and Friedel [24–26]. The binary enthalpies have been extended to high-entropy alloys using the extended regular solid solution model proposed by Takeuchi and Inoue [27]. It should be noted that Miedema’s chemical enthalpies have limitations that arise primarily from the simplistic treatment of atomic interactions in the ‘macroscopic atom’ picture used by Miedema [28]. But while they may not be as accurate as enthalpies calculated from first-principles calculations, they do align reasonably well with the experimental data and are fairly easy to calculate and integrate with machine learning approach. The presence of strongly correlated multicollinear variables can cause some features to appear insignificant, and thus hierarchical clustering was performed

**Table 1**

Comparison of the present work with other phase selection ML models available in the literature for HEAs. ✓ indicates that the model has the predictive capability. ✗ indicates that the predictive capability is lacking in the model. Classification accuracy has been included wherever available.

Reference	Phase fractions	Classification accuracy for different classification labels (%)								
		SS & (SS + IM)	FCC, BCC & DP	FCC	BCC	(FCC + BCC)	(FCC + IM)	(BCC + IM)	(FCC + BCC + IM)	IM
Present work	✓	90.1	87.6	89.7	93.1	69.2	82.0	77.1	71.4	71.4
Huang et al. [11]	✗	78.9	✗	✗	✗	✗	✗	✗	✗	✓
Li and Guo [8]	✗	✗	✗	60.0	75.0	✗	✗	✗	✗	✗
Roy et al. [7]	✗	✗	✗	✓	✓	✗	✗	✗	✗	✗
Zhang et al. [9]	✗	88.7	91.3	✓	✓	✓	✗	✗	✗	✗
Lee et al. [20]	✗	89.5	✗	✗	✗	✗	✗	✗	✗	✓



**Fig. 1.** Flowchart of machine learning strategy: (a) Alloy descriptor space comprising features derived from elemental and alloying properties, (b) Phase target space comprising seven vectorized labels used to train the ML models (F is 'FCC', B is 'BCC', IM is 'Intermetallic'; 0 → phase absent, 1 → phase present), (c) An ensemble of 150 deep ANNs trained to predict the phase fraction of each phase ( $f_{\text{FCC}}$ ,  $f_{\text{BCC}}$  and  $f_{\text{IM}}$ ), (d) Exploring phase fractions over large compositional spaces in HEAs using trained ANNs (ensemble averaging), (e) Validating predictions through experimental and CALPHAD results reported in recent literature.

on the basis of Spearman rank-order correlations to identify multicollinear features (Fig. S1.b). A dendrogram (Fig. S1.a) was created to visualize the hierarchical relationship between features. It is most accurate at the bottom as it represents the summary of distance matrix and some information is always lost while creating it. Thus, only the three bottom-most clusters (with the smallest height of connecting links) were considered for feature elimination and four features were removed, resulting in a set of 14 features (Fig. S2) that were used for machine learning models. The importance of each feature (Fig. S3.(a and b)) was calculated through implementation of Random Forest [29] algorithm that is based on ensembles of randomized decision trees. The valence electron concentration (VEC) and average cohesive energy appear as the most important features for phase selection, followed by the average density and size asymmetry (measure of lattice strain). The importance of these features has been explored in §3.1.2 and §3.1.3.

## 2.2. Preparing phase target space

The dataset contained seven categorical phase labels, each representing a distinct phase combination, and these were converted to binary vectors before implementing machine learning algorithms. Each phase label was converted into a three-element binary vector – [F, B, I], wherein the elements F, B and I correspond to the presence (value = 1) or absence (value = 0) of FCC, BCC and IM phase respectively. All of these conversions are illustrated in Fig. 1.b. Even though we are performing seven-label classification, this transformation allows us to define our ML approach in terms of only three target parameters and thereby increases the training data available for each parameter, for instance, the 'F' index prediction, representing presence of FCC phase, will learn from not only FCC HEAs, but also from (FCC + BCC), (FCC + IM) and (FCC + BCC + IM) HEAs in the dataset.

**Table 2**

List of all 18 features explored in this work, along with their mathematical expression and description.

S. No.	Feature	Calculation	Feature Description
1	$\delta V_m$	$\sqrt{\sum_{i=1}^N c_i \left(1 - \frac{V_{m,i}}{\bar{V}_m}\right)^2}$	Molar volume asymmetry
2	$\delta R_{\text{met}}$	$\sqrt{\sum_{i=1}^N c_i \left(1 - \frac{r_i}{\bar{r}}\right)^2}$	Metallic radius asymmetry
3	$\Delta S_{\text{config}}$	$-R \sum_{i=1}^N c_i (\ln c_i)$	Configurational entropy
4	$\delta R_{\text{cov}}$	$\sqrt{\sum_{i=1}^N c_i \left(1 - \frac{r_i}{\bar{r}}\right)^2}$	Covalent radius asymmetry
5	$T_m$	$\sum c_i T_{m,i}$	Average melting point
6	$E_{\text{coh}}$	$\sum c_i E_{\text{coh},i}$	Average cohesive energy
7	$\delta E_{\text{coh}}$	$\sqrt{\sum_{i=1}^N c_i \left(1 - \frac{E_{\text{coh},i}}{\bar{E}_{\text{coh}}}\right)^2}$	Cohesive energy asymmetry
8	$\delta K$	$\sqrt{\sum_{i=1}^N c_i \left(1 - \frac{K_i}{\bar{K}}\right)^2}$	Bulk modulus asymmetry
9	$\delta E$	$\sqrt{\sum_{i=1}^N c_i \left(1 - \frac{E_i}{\bar{E}}\right)^2}$	Young's modulus asymmetry
10	$\delta G$	$\sqrt{\sum_{i=1}^N c_i \left(1 - \frac{G_i}{\bar{G}}\right)^2}$	Shear modulus asymmetry
11	VEC	$\sum_{i=1}^N c_i VEC_i$	Valence electron concentration
12	$\chi^{\text{Allen}}$	$\sum c_i \chi_i$	Average Allen electronegativity
13	$\rho$	$\sum c_i \rho_i$	Average density
14	$\Delta H^{\text{el}}$	$\sum_{i < j} 4H_{ij}^{\text{el}} c_i c_j$	Elastic enthalpy of mixing
15	$\Delta H_{SS}^{\text{chem}}(S)$	$\sum_{i < j} 4H_{ij}^{\text{chem}, S} c_i c_j$	Miedema's chemical enthalpy of mixing (solid)
16	$\Delta H_{IM}^{\text{chem}}(S)$	$\sum_{i < j} 4H_{ij}^{\text{chem}, S} c_j$	Miedema's IM formation enthalpy (solid)
17	$\Delta H_{SS}^{\text{chem}}(L)$	$\sum_{i < j} 4H_{ij}^{\text{chem}, L} c_j$	Miedema's chemical enthalpy of mixing (liquid)
18	$\Delta H_{IM}^{\text{chem}}(L)$	$\sum_{i < j} 4H_{ij}^{\text{IM}, L} c_i c_j$	Miedema's IM formation enthalpy (liquid)

		ACTUAL							(Sum) Precision
		FCC	BCC	FCC + BCC	FCC + IM	BCC + IM	FCC+ BCC+IM	IM	
PREDICTED	FCC	35	0	5	5	0	0	0	45 78% 22%
	BCC	0	108	6	2	9	1	0	126 86% 14%
	FCC + BCC	1	2	27	0	0	4	0	34 79% 21%
	FCC + IM	2	0	0	41	2	0	2	47 87% 13%
	BCC + IM	0	5	0	0	37	1	0	43 86% 14%
	FCC+ BCC+IM	0	1	1	0	0	15	0	17 88% 12%
	IM	1	0	0	2	0	0	5	8 63% 38%
	(Sum) Recall	39 90% 10%	116 93% 7%	39 69% 31%	50 82% 18%	48 77% 23%	21 71% 29%	7 71% 29%	320 84% 16%

**Fig. 2. Cross-validation performance:** Confusion matrix for seven-label classification of alloy phases. The diagonal (green cells) contains True Positive count for each class. The last column and row (black cells) represent the precision and recall respectively for each class. The red highlighted cells represent occasions where the model predicted a non-existent phase.

### 2.3. Model training and cross-validation performance

The core of our ML strategy comprises a deep ANN with 14 nodes in the input layer followed by seven hidden layers (40, 35, 30, 25, 20, 15 and 10 nodes in subsequent layers), each of which employs a rectified linear unit activation (ReLU). The final output layer has 3 nodes - one node each for FCC, BCC and IM prediction. The model is trained to predict the probability of existence of each phase (FCC/BCC/IM) which are subsequently transformed into phase fraction estimates ( $F_i$ ) using softmax activation for the output layer. The softmax activation function ensures that the outputs follow a probability distribution with summation equal to 1 and calculates phase fractions as:  $F_i = \frac{e^{z_i}}{\sum e^{z_i}}$ , where  $z_i$  is the input received by unit  $i$  in the output layer and  $F_i$  is the phase fraction output returned by each unit (Fig. 1.c). The programming environment used for implementing the machine learning framework comprised of Python 3.8.1 and associated open-source libraries: a) ‘pandas|1.0.3’ and ‘numpy|1.18.2’ for data processing, b) ‘scikit-learn|0.22.2’ for implementation of decision trees, random forests and hierarchical clustering, and c) ‘tensorflow|2.2.0rc2’ and ‘keras|2.3.1’ for implementation of ANNs. Each ANN was trained using ‘RMSprop’[30] as the optimizer (learning rate of 0.0005) and ‘Binary Crossentropy’[31] as the loss function. For more details on the implementation of ANNs, readers may refer to [supplementary data](#) (§S1).

We used five-fold cross-validation for the training process i.e., the training-set to testing-set size ratio was always 80:20. Each cross-validation was independent of the others and the trained weights or initialized parameters were not carried forward. The performance, as depicted in Fig. 2, was calculated using only the cross-validation results

i.e., for each alloy, only that prediction was considered when it was part of the validation set and thus did not participate in the training process. An overall accuracy of 84% was obtained over the seven-class classification and in only 17 out of 323 instances, a non-existent phase was predicted. The highest recall - TP/(TP + FN), where TP is True Positive and FN is False Negative - was observed for BCC class (93%) which may be attributed to the considerably higher number of alloys with BCC phase in the database. The presence of FCC, BCC and IM phase was predicted with an accuracy of 91%, 97% and 83% respectively.

### 2.4. Ensemble method for exploration of alloy systems

The final ML model, used for exploration of alloy systems (§3.2–3.4), consists of an ensemble of 150 trained ANNs and model averaging is carried out to predict the final phase fraction estimates for each phase (FCC/BCC/IM). The model averaging technique was used to accommodate the high variance of ANNs. The model requires only one user input (alloy composition) and the features used by ANNs are calculated automatically by supporting scripts and elemental database. The compositional spaces in HEA systems have been explored by reducing their dimensionality through clubbing of elements into ternaries with fixed composition. For example, in Fig. 7, CrMoVNbTi alloy compositional space was broken into three components – Cr, Mo and an equiatomic ternary (VNbTi), thereby enabling representation of a slice of HEA compositional space on a ternary plot. The predictions were made at compositions separated by 1 at. % and thus, each ternary contour plot was built from 5151 unique alloy compositions. None of the alloys present in the training dataset represented these alloy systems; thus, these alloy systems were completely new to the model. Also, it must be

noted that there are regions within these plots (Figs. 5–7) that represent binary and ternary alloy systems and thus, the model accuracy in these regions is bound to suffer since the model was trained only on concentrated alloys ( $\Delta S^{\text{config}}$  distribution in training dataset: 0.8R–2.19R with average of 1.64R; where R is the universal gas constant) and the phase selection rules for such alloys can vary significantly from classic Hume-Rothery rules [32,33]. For this reason, we have marked out a line corresponding to constant  $\Delta S^{\text{config}} = 0.8R$  in each plot indicating the region outside which model predictions may not be insightful.

## 2.5. Model validation through comparison with first-principles calculations

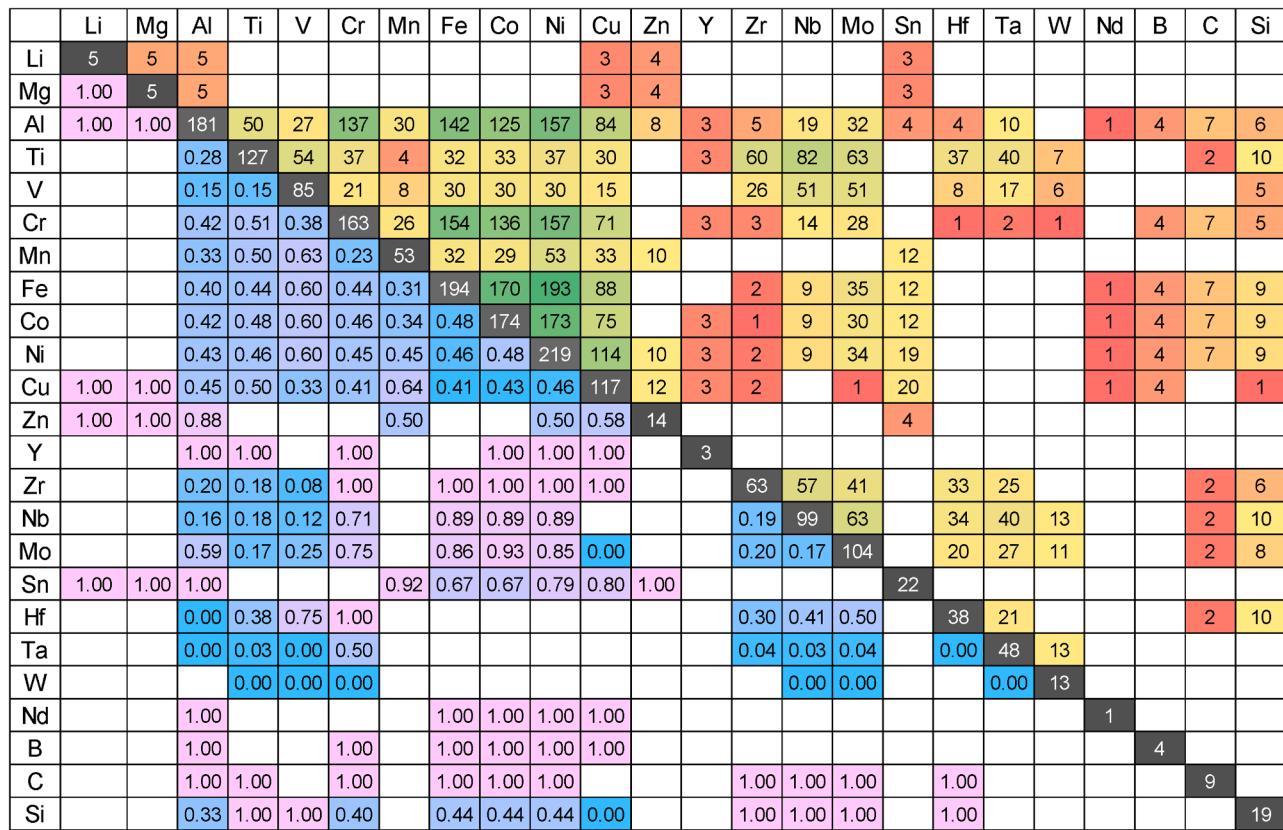
The final ensemble model was used to explore phase fractions over continuous composition variations in  $\text{Al}_x(\text{NiCoFeCr})_{1-x}$ ,  $\text{Al}_x(\text{NiCoFeCrMn})_{1-x}$  and  $\text{Mo}_x(\text{WTaTiZr})_{1-x}$  alloys as the phase stabilities in these systems have been studied previously through first-principles calculations [33–35]. The comparison is shown in Fig. S4. The ML model correctly predicts the transition from single-phase FCC to dual-phase (FCC + BCC) to single-phase BCC as Al concentration is increased in  $\text{Al}_x(\text{NiCoFeCr})_{1-x}$  and  $\text{Al}_x(\text{NiCoFeCrMn})_{1-x}$  alloy systems. The position of phase-transition boundaries also align closely with those predicted by first-principles calculations. For  $\text{Mo}_x(\text{WTaTiZr})_{1-x}$  alloy system, both ML model and DFT calculations predict BCC as the stable phase.

## 3. Results and discussion

### 3.1. Exploring insights from the dataset

#### 3.1.1. Binary tendencies and intermetallic formation

The compositional diversity in the dataset is presented in Fig. 3 as number of occurrences for each element/binary. It also depicts how often a binary is associated with the presence of IM phase, and while it would be presumptuous to portray these associations as binary tendencies, some statistical trends observed are worth mentioning – a) Mo is strongly associated with IM phase when present with (Al, Cr, Fe, Co, Ni or Hf) as compared to (Ti, V, Zr, Nb, Ta or W), b) strong association of Hf with presence of IM phase in Ti-Zr-Nb-Mo-Hf-Ta refractory HEAs, and c) strong association of Ti with IM phase when present with (Cr, Fe, Co, Ni or Cu) as compared to (Zr, Nb, Mo or Ta). Fig. S5 indicates that the binaries associated with frequent occurrence of IM phases generally exhibit high radial asymmetry (>8%) and high electronegativity difference (>0.15). This is in line with the Hume-Rothery rules and suggests that the phase selection in HEAs is not completely oblivious of binary tendencies. Further, the Ti-(Cr,Fe,Co,Ni,Cu) binary pairs (Fig. S5. c) exhibit high negative chemical enthalpy of mixing (-7 to -34 kJ/mol), significant atomic size mismatch (14 to 18%) and large Allen electronegativity difference (0.27 to 0.5); all of which point to a high tendency of IM formation. Gao et al. [36] confirmed this ordering effect of Ti experimentally as they observed a significant IM phase formation upon addition of Ti to (CrFeNi) alloy system.



**Fig. 3. Exploring the dataset comprising 323 alloys (1 ternary, 42 quaternary, 135 quinary, 108 hexanary, 33 septenary, 1 octonary and 3 nonary alloys):** The diagonal (black cells) indicates the number of times each element appears in the dataset. On right side of the diagonal (red-yellow-green color scale), each cell represents the number of times each binary appears in the dataset. On left side of the diagonal (blue-purple color scale), each cell represents the fraction of occurrences wherein a binary is associated with the presence of IM phases (1 indicates that alloy had IM phases whenever this binary was present, and 0 indicates that alloy had only SS phases whenever this binary was present).

### 3.1.2. FCC vs. BCC classification using feature maps

Guo et al. [19] proposed an empirical rule based on valence electron concentration (VEC) for establishing stability regions of FCC and BCC solid solution phases in HEAs. They observed that HEAs with  $\text{VEC} < 6.87$ ,  $6.87 \leq \text{VEC} < 8$  and  $\text{VEC} \geq 8$  comprise of BCC, (BCC + FCC) and FCC phases respectively. With the view that there might be other simple parameters that can guide FCC vs. BCC phase classification, we studied the distribution of FCC, (FCC + BCC) and BCC phases over a wide range of feature spaces (Fig. S6) and observed that cohesive energy (Coh. E) and electronegativity ( $\chi_{\text{Allen}}$ ) can also be used for classifying FCC vs. BCC stability regions. This is somewhat expected since - VEC, Coh. E and  $\chi_{\text{Allen}}$  all are related to the electron density and manifest as bonding characteristics in a crystal. Fig. 4.a. shows the distribution of SS phases over VEC, Coh. E and  $\chi_{\text{Allen}}$  feature spaces. In general, low VEC, high Coh. E and low  $\chi_{\text{Allen}}$  stabilizes BCC structure; and only BCC solid solutions were observed in the dataset when  $\text{VEC} < 6.8$  or  $\chi_{\text{Allen}} < 1.54$  or Coh. E  $> 5.25$  eV. In contrast, high VEC and low Coh. E stabilizes FCC structure and only FCC solid solutions were observed when  $\text{VEC} > 8$  or Coh. E  $< 3.55$  eV. There are some exceptions to this VEC criterion: ( $\text{Al}_8\text{Cu}_{0.5}\text{Li}_{0.5}\text{Mg}_{0.5}\text{Zn}_{0.5}$ ,  $\text{Al}_8\text{Li}_{0.5}\text{Mg}_{0.5}\text{Sn}_{0.5}\text{Zn}_{0.5}$  and  $\text{AlLi}_{0.5}\text{MgSn}_{0.2}\text{Zn}_{0.5}$ ), all of which have very low VEC and contain FCC solid-solutions, but even these get resolved with Coh. E criterion as they all have Coh. E lower than 3.55 eV. The best empirical FCC vs. BCC classification criteria we can make out from these observations are: BCC ( $\text{VEC} < 6.8$  and Coh. E  $> 3.55$  eV), FCC ( $\text{VEC} > 8$  and Coh. E  $< 5.25$  eV) and BCC + FCC ( $6.8 < \text{VEC} < 8$  and  $3.55 \text{ eV} < \text{Coh. E} < 5.25 \text{ eV}$ ).

### 3.1.3. Solid solution vs. Intermetallic classification using feature maps

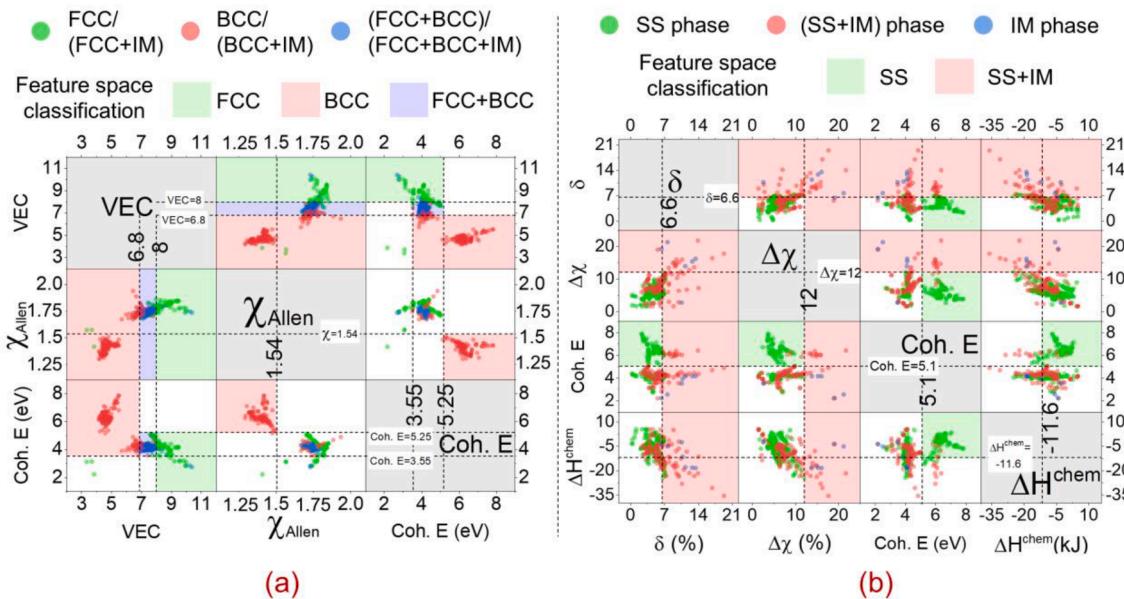
This classification has been widely studied since the initial interest in HEAs was primarily driven by a search for disordered concentrated alloys. Yang and Zhang [13] proposed an empirical rule for formation of simple solid solutions (SSSs):  $\delta < 6.6\%$  and  $\Omega > 1.1$ , where  $\delta$  is a measure of radius asymmetry that symbolizes quantum of lattice strain and  $\Omega$  is a thermodynamic parameter that compares relative strength of entropic and enthalpy contributions to alloying energetics ( $\Omega = T_m \Delta S / |\Delta H^{\text{mix}}|$ ). Guo et al. [14] proposed a slightly modified model that neglects entropy contribution and predicts formation of SSS when:  $\delta < 6.6\%$  and  $3.2 \text{ kJ/mol} > \Delta H^{\text{mix}} > -11.6 \text{ kJ/mol}$ . Despite being quite crude, primarily due to the small size of datasets that they have been derived from, these rules

align well with the Hume-Rothery rules and have guided experimental HEA design over the past decade. We tested these rules on our considerably expanded dataset and also searched for other simple parameters that can drive classification of IM phases (Fig. S7). Fig. 4.b. shows the distribution of SS and (SS + IM) phases over feature spaces comprising:  $\delta$ , chemical enthalpy of mixing ( $\Delta H^{\text{chem}}$ ), local mismatch in electronegativity ( $\Delta\chi$ ) and Coh. E. Here,  $\Delta H^{\text{chem}}$  for HEAs has been calculated through Takeuchi's extended regular solution model [27] that uses Miedema's binary enthalpies [23]. IM phases are always present whenever  $\delta > 6.6\%$  (large lattice strain) or  $\Delta\chi > 12\%$  (high ordering tendency). The ( $3.2 \text{ kJ/mol} > \Delta H^{\text{mix}} > -11.6 \text{ kJ/mol}$ ) criterion for formation of solid solution phases is not very helpful since we see a considerable presence of IM phases within this range and also an absence of IM phases in a significant number of alloys outside this range. Coh. E appears as a much more useful parameter for classifying stability region of solid solution phases. The feature space corresponding to Coh. E  $> 5.1 \text{ eV}$  and ( $\delta < 6.6\%$  or  $\Delta\chi < 12\%$ ) is almost exclusively occupied by SS phases.

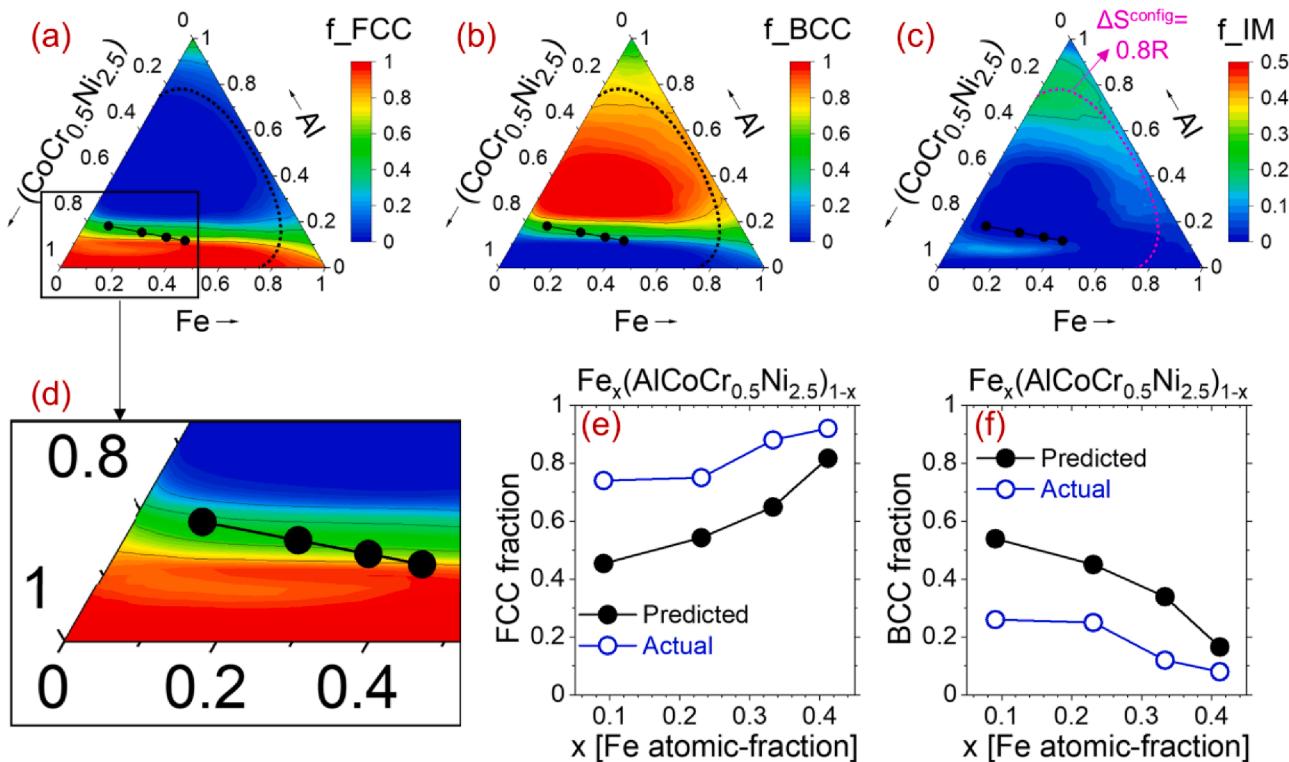
It is important to note that while such rules (§3.1.2 and 3.1.3) may be useful for cursory phase prediction, these are based only on two-dimensional feature spaces and are thus, quite simple in nature and insufficient for a robust phase classification in HEAs. With machine learning (using ANNs), we are probing a 14-feature space; therefore, the dependencies learnt are considerably more complex (often incomprehensible) and more representative of the information contained in the dataset.

### 3.2. Exploring $\text{Fe}_x\text{Al}_y(\text{CoCr}_{0.5}\text{Ni}_{2.5})_{1-x-y}$ alloy system

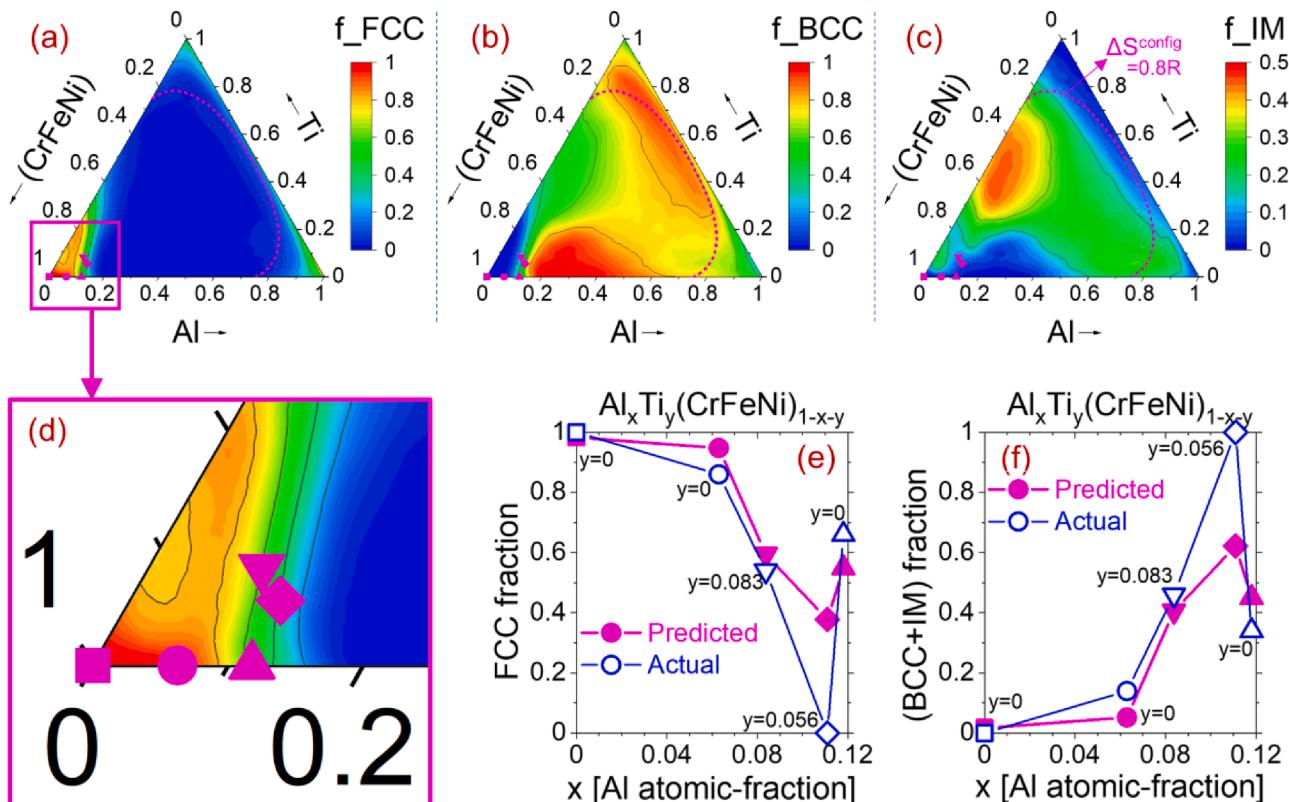
We used the ensemble model to explore the distribution of phases in  $\text{Fe}_x\text{Al}_y(\text{CoCr}_{0.5}\text{Ni}_{2.5})_{1-x-y}$  alloy system and compared the predicted and actual phase fractions (measured experimentally by Liu et al. [37]) of FCC (Fig. 5.e) and BCC (Fig. 5.f) phases in  $\text{Fe}_x(\text{AlCoCr}_{0.5}\text{Ni}_{2.5})_{1-x}$  ( $x = 0.09-0.41$ ) alloy system. Our model correctly predicts the increase in FCC phase fraction as the Fe concentration increases and also identifies the almost complete transition to the FCC phase at  $x > 0.41$ . It should be noted that the increase in iron content is accompanied by a corresponding decrease in the Al content, i.e., lower Fe content alloys have higher Al content and vice-versa. In our prior work [38], as well as



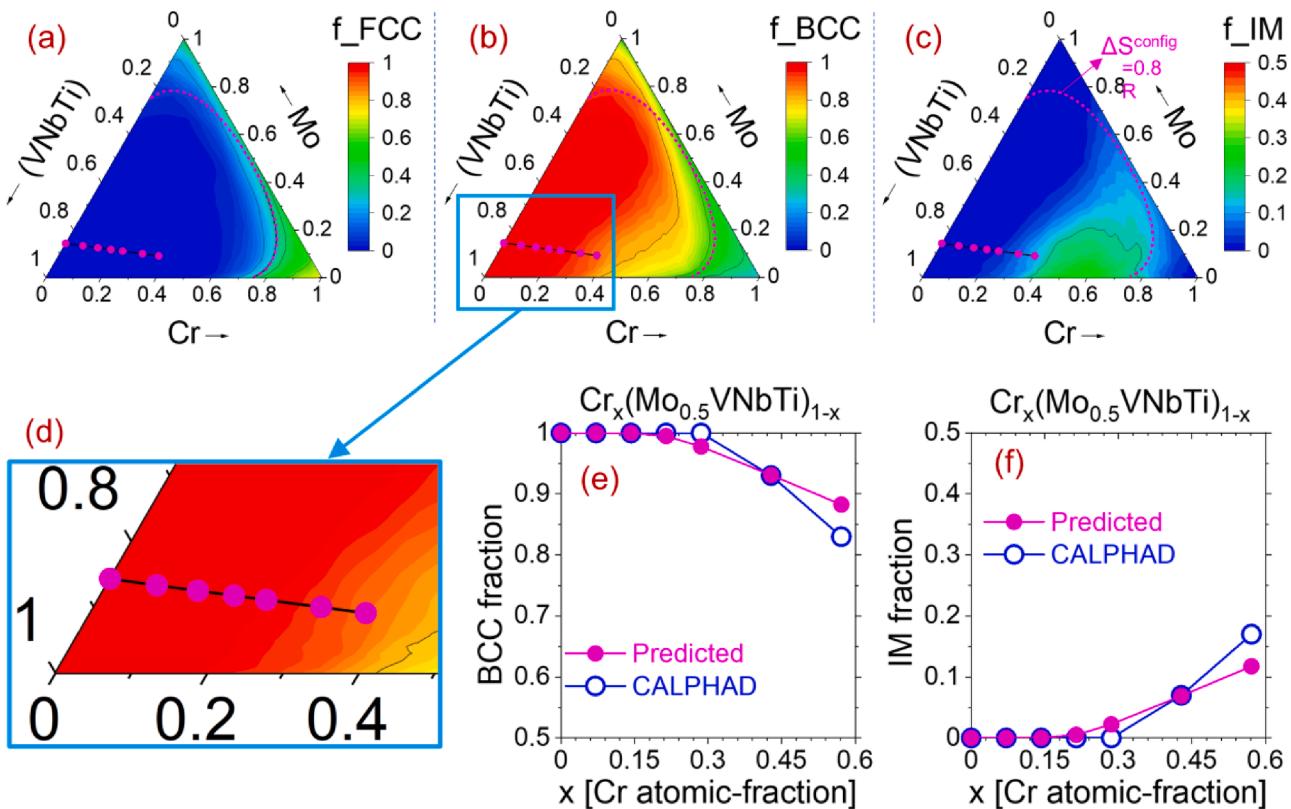
**Fig. 4. Feature maps and phase classification:** Scatter-matrix plots showing distribution of: (a) FCC and BCC solid-solution phases, and (b) Solid solution and IM phases in the HEA database as a function of different alloy feature. Each alloy has been color-coded based on the phases present and feature spaces corresponding to stability of particular phases have been classified.



**Fig. 5.** Exploring  $\text{Fe}_x\text{-Al}_y\text{-}(\text{CoCr}_{0.5}\text{Ni}_{2.5})_{1-x-y}$  alloy system: Predicted phase fraction of (a) FCC, (b) BCC and (c) IM for the entire composition range; (d) Magnified view of the composition-space containing data points where experimental results are available; Comparison of predicted and actual phase fraction of (e) FCC phase and (f) BCC phase.



**Fig. 6.** Exploring  $\text{Al}_x\text{-Ti}_y\text{-}(\text{CrFeNi})_{1-x-y}$  alloy system: Predicted phase fraction of (a) FCC, (b) BCC and (c) IM for the entire composition range; (d) Magnified view of the composition-space containing data points where experimental results are available; Comparison of predicted and actual phase fraction of (e) FCC phase and (f) (BCC + IM) phases.



**Fig. 7.** Exploring  $\text{Cr}_x\text{-Mo}_y\text{-}(\text{VNbTi})_{1-x-y}$  alloy system: Predicted phase fraction of (a) FCC, (b) BCC and (c) IM for the entire composition range; (d) Magnified view of the composition-space containing data points where CALPHAD calculations are available; Comparison of predicted and CALPHAD calculated phase fraction of (e) BCC phase and (f) IM phase.

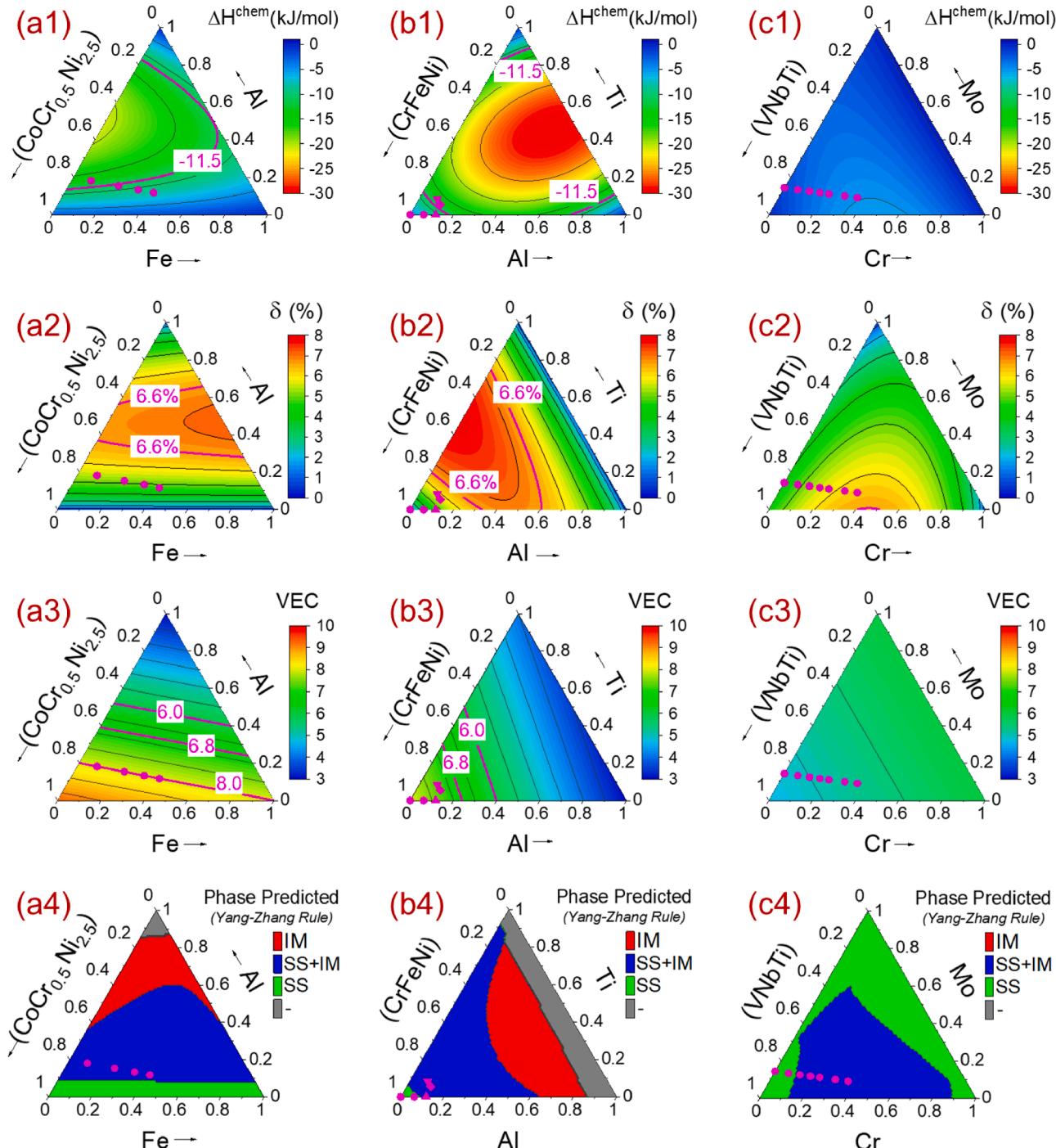
reports from other researchers [3,34], it has been shown that Al tends to promote B2 ordering and a transition to a BCC structure. This is also borne out by the high negative formation enthalpies calculated [39] (and measured [40]) for B2 IM at high Al content. In the model predictions, shown in Fig. 5.(a and b), the compositional space is neatly divided into FCC, (FCC + BCC) and BCC regions by contour lines that are almost entirely dictated by the concentration of Al. Liu et al. [37] have termed Fe as an FCC phase stabilizer in this system based on the observed increase in FCC phase fraction as Fe content was increased. Given the correlation between increasing Fe content and decreasing Al content, we feel that this assertion does not necessarily hold true in all cases. Our model, in fact, shows that if Fe content is varied keeping the Al content constant, the FCC phase fraction does not change significantly (Fig. 5.b). As we move from  $x = 0.09$  to  $0.41$ , the Fe concentration increases from 9 to 41 at.% whereas Al concentration decreases from 18 to 11.7 at.%, and our model shows that the BCC  $\rightarrow$  FCC transition is because of this decrease in Al concentration and not due to the increase in Fe concentration. Wang et al. [41] have also experimentally confirmed this FCC  $\rightarrow$  BCC transition in  $\text{Al}_x\text{CoCrFeNi}$  alloy system induced by increasing concentration of Al and observed similar phase stability boundaries with only FCC phase present when Al concentration is  $< 11$  at.% and only BCC phase present when it is  $> 19$  at.%.

### 3.3. Exploring $\text{Al}_x\text{-Ti}_y\text{-}(\text{CrFeNi})_{1-x-y}$ alloy system

For further validation, we used the ensemble model to explore  $\text{Al}_x\text{-Ti}_y\text{-}(\text{CrFeNi})_{1-x-y}$  alloy system since: a) due to the presence of both Al (that has been shown to promote FCC to BCC transition [38,41,42]) and Ti (that has a strong tendency to form IM [36,43–45]), one expects significant competition between different phases in this system, and b) some discrete compositions within this system have been studied experimentally by Zhang et al. [45]. The predicted phase fractions for

FCC (Fig. 6.a), BCC (Fig. 6.b) and IM (Fig. 6.c) indicate that: a) the addition of Al to ( $\text{CrFeNi}$ ) brings about FCC  $\rightarrow$  BCC transformation, and b) the addition of Ti to ( $\text{CrFeNi}$ ) drives formation of IM phases and also leads to FCC  $\rightarrow$  BCC transformation at higher Ti concentrations. These observations are in line with the experimental measurements [36,45] and first-principles predictions [46]. Recently, Wang et al. [43] studied the effect of Ti addition in  $\text{CoCrFeNiTi}_x$  alloy system ( $x = 0.1\text{--}0.7$ ) and observed that the structure transformed from FCC at  $x = 0.1$  to (FCC + Fe-Cr rich  $\sigma$  phase) at  $x = 0.3$  to (FCC + FeCr IM +  $\text{Ni}_2\text{Ti}$  R phase) at  $x = 0.5$ . The Co-Ni-Ti ternary phase diagram [47] shows the formation of the A1 (disordered FCC) structure at low Ti (and equiatomic Ni and Co) content. As the Ti content increases in the ternary Co-Ni-Ti system, IM compounds begin to form. A similar effect is seen in the high-entropy system as well, since increased Ti content results in Fe and Cr separating out to form the Fe-Cr  $\sigma$  phase. As the Ti content increases further, it forms IM with Ni due to a very high electronegativity difference that results in an exceptionally large negative formation enthalpy, as can be seen in Fig S5.c. Thus, Ti is now known to have a strong ordering tendency when present with ( $\text{CrFeNi}$ ) and the same is predicted by the model also (Fig. 6.c). Fig. 6.(e and f) show the comparison between the predicted and experimentally measured [45] phase fractions of FCC and (BCC + IM) phases respectively for five alloy compositions. The predictions for three compositions that do not contain Ti (i.e.,  $x = 0, 0.0625$  &  $0.117$ ;  $y = 0$ ) show that the model is correctly predicting the decrease in FCC phase fraction as Al increases. The sudden change in the phase distribution as Al increases from 11.1 to 11.7 at.% is due to the fact that the  $x = 0.111$  composition also has 5.6 at.% Ti present in it whereas  $x = 0.117$  composition does not contain any Ti. This is in agreement with earlier observations where Ti has been shown to drive IM formation thereby decreasing FCC phase fraction [36,43–45]. The experimentally observed phase distributions for Ti-containing compositions indicate that the presence of Ti can cause FCC phase to disappear at lower Al

concentrations and the same is also predicted by the ML model, as seen in Fig. 6.d where the contours representing absence of FCC phase occur at lower Al concentration as Ti increases. The model is able to closely predict the phase fractions of each phase, except for the ( $x = 0.111$ ,  $y = 0.056$ ) composition where the microstructure no longer has FCC phase, but the model predicts 37% FCC phase. At closer inspection of Fig. 6.d, it can be seen that this composition is right adjacent to the contour lines marking complete absence of FCC phase. Thus, the model is correctly predicting the complete transition from FCC to BCC phase (as observed experimentally), albeit at only a slightly higher Al concentration.



**Fig. 8. Feature distribution in alloy systems explored using ML:** Distribution of (1) Miedema's chemical enthalpy of mixing ( $\Delta H^{\text{chem}}$ ), (2) Radius asymmetry ( $\delta$ ), (3) VEC and (4) SS and IM phases predicted by empirical rule over (a)  $\text{Fe}_x\text{-Ni}_y\text{-}(\text{AlCoCr}_{0.5})_{1-x-y}$  alloy system, (b)  $\text{Al}_x\text{-Ti}_y\text{-}(\text{CrFeNi})_{1-x-y}$  alloy system, (c)  $\text{Cr}_x\text{-Mo}_y\text{-}(\text{VNbTi})_{1-x-y}$  alloy system. Contours corresponding to  $\Delta H^{\text{chem}} = -11.5 \text{ kJ/mol}$ ,  $\delta = 6.6\%$ , VEC = (6, 6.8 & 8) have been marked out in respective plots. Yang-Zhang rule refers to empirical rules (based on  $\delta$  and  $\Omega$ ) proposed by Yang and Zhang [13] and elaborated in Fig. S8.

distribution predicted by the model (Fig. 7.c) aligns well with the statistical IM formation tendencies of Mo and Cr binaries observed earlier during the exploration of database (Fig. 3), wherein Cr-(V,Nb,Ti) binaries are associated with IM phases much more frequently as compared to Mo-(V,Nb,Ti) binaries. The predicted and CALPHAD calculated phase fractions [48] of BCC and IM phases were compared for  $\text{Cr}_x(\text{Mo}_{0.5}\text{VNbTi})_{1-x}$  ( $x = 0\text{--}0.57$ ) system, as shown in Fig. 7.(d and e). The model accurately predicts the presence of only BCC phase when  $x < 0.286$  and the appearance of IM phase at  $x \geq 0.286$ . The increase in IM phase fraction at high Cr concentrations is also predicted accurately.

### 3.5. Insights into ML model decision-making process

ANNs are extremely powerful at learning linear and non-linear relationships in data for both regression and classification tasks, but their major drawback stems from the fact that it is very difficult to interpret the exact decision-making process used by them and for this reason, they are often referred to as a black-box [49]. While there are ways in which the ANN inner-workings can be glimpsed [50,51], their implementation and interpretation is an exhaustive task and remains as a future scope of this work. We have, however, constructed key feature maps ( $\Delta H^{\text{chem}}$ ,  $\delta$  and VEC) in Fig. 8, along with the phases predicted by the empirical rule [13] for alloy systems explored earlier (§3.2–3.4). The comparison of these with ML predicted phase fractions allows insights into some general trends being followed by the trained ML model.

Nonetheless, the model reinforces the empirical relation [5] between VEC and the crystal structure. The  $(\text{CoCr}_{0.5}\text{Ni}_{2.5})$  ternary has  $\text{VEC} > 8$  and the model predicts FCC phase. With addition of Fe, the VEC stays comfortably higher than 8 and model predicts FCC phase throughout  $\text{Fe}_x\text{CoCr}_{0.5}\text{Ni}_{2.5}$ . On the contrary, addition of Al decreases VEC and predicted structure transitions from FCC to  $(\text{FCC} + \text{BCC})$  to BCC. Similarly, addition of both Ti and Al to  $(\text{CrFeNi})$  decreases VEC and the ML model predicts FCC to BCC transition. On the contrary, the VEC in  $\text{Cr}_x\text{Moy}-(\text{CrFeNi})_{1-x-y}$  alloy system stays below 6 throughout the composition range and ML model predicts BCC phase. These observations indicate that the ML model is closely following the VEC dependent phase-stability rules observed by Guo et al. [19] and Singh et al. [33]. It must be noted that the ML model is looking beyond just VEC since all the discrete compositions looked at in  $\text{Fe}_x(\text{AlCoCr}_{0.5}\text{Ni}_{2.5})_{1-x}$  ( $x = 0.09\text{--}0.41$ ) have constant  $\text{VEC} = 8$  and still the crystal structure (both in experimental measurements and ML predictions) is changing from  $(\text{FCC} + \text{BCC})$  at  $x = 0.09$  to almost pure FCC at  $x = 0.41$ . While the ML model correctly predicts this transition, as seen in Fig. 5.(e and f), we are unable to identify the exact rule leading to this decision-making. This is even more puzzling since  $\chi_{\text{Allen}}$  remains constant ( $=1.79$ ) and Coh. E slightly increases ( $4.19 \rightarrow 4.22$  eV) as Fe content increases from 9 to 41 at. %, and thus, the empirical observations in §3.1.2 are unable to explain this transition.

It appears that the ML model gives more importance to  $\delta$ , as compared to  $\Delta H^{\text{chem}}$ , for predicting formation of IM since a high  $\delta$  is almost always associated with presence of IM phase.  $(\text{CrFeNi})$  has a negligible size difference ( $\delta = 0.4\%$ ) but the addition of both Al and Ti increases  $\delta$  and the increase is much steeper with Ti addition as compared to Al. Similarly,  $\Delta H^{\text{chem}}$  decreases (becomes more negative) with addition of both Al and Ti to  $(\text{CrFeNi})$  but the decrease is much steeper with Ti addition. Both of these point to a considerable higher IM formation tendency associated with Ti and the same is predicted by the ML model and also observed in experimental observation. In principle, this decision-making is in agreement with existing empirical rules [13,14] but the boundaries drawn by ML model are much more versatile.

The decision boundaries in ML model are much more flexible as compared to those obtained from empirical rules [13,14], which is to be expected as they are built on a 14-feature space, as compared to only 2-feature space for most empirical models, and use infinitely more fitting parameters along with non-linear activations. The usefulness of higher-

dimensional classification boundaries of our ML model becomes clear when we compare these (Figs. 5–7) with phase predictions from Yang-Zhang empirical rule (Fig. 8.a-d.4, Fig. S8). The simple empirical rules assign large compositional spaces as containing (SS + IM) phases but give no information about relative changes in the amount of SS or IM phases and this makes them insufficient for targeted alloy design. But the ML model predicts the variation in relative amounts of each phase as composition changes. The ML model also allows prediction of relative amount of FCC vs BCC rather than just a qualitative information obtained from empirical rules based on VEC [19]. Also, the empirical rules indicate presence of IM phases in  $\text{Fe}_x(\text{AlCoCr}_{0.5}\text{Ni}_{2.5})_{1-x}$  ( $x = 0.09\text{--}0.41$ ) alloy compositions but the ML model predicts only SS phases which matches experimental observation. Similarly, in  $\text{Cr}_x(\text{Mo}_{0.5}\text{VNbTi})$  alloys, empirical rules indicate the appearance of IM phases at much lower concentrations of Cr as compared to CALPHAD calculations and ML predictions.

Our investigations on the three alloy systems in §3.2–3.4 also bring out another interesting trend – while the trends in BCC and FCC fractions are predicted reasonably well, numerical agreement between the model predictions and experimental results improves when the Al content is low. In Fig. 5.(e and f), it can be seen that as the Al content decreases (i.e., the Fe content increases), the numerical agreement improves. In Fig. 6.(e and f), we observe that at  $x = 0$  (i.e., absence of Al), we get excellent numerical agreement between the model and experiments, while at somewhat higher Al content, the agreement though reasonable, does have some error associated with it. An excellent agreement was obtained in case of the refractory Cr-Mo-V-Nb-Ti system which does not contain Al. First principles calculations by Singh et al. on Al containing HEAs [34,46] indicate very high interchange energies for pairs containing Al in the B2 phase which drives the Warren-Cowley short-range order parameter. We surmise that in Al containing systems, the difference between experimental measurements and our model may be due to the tendency for Al to promote short-range ordering which may not be well-captured by the features selected in this work. There are possibly two other factors that may contribute to the difference between the experimental and predicted phase fractions. Firstly, the model is trained on the presence/absence of each phase and the phase fractions are extracted indirectly from the probabilities of occurrence of phases. Since occurrence probabilities are not a perfect representation of phase fractions, some errors are inadvertently introduced with this assumption and thus, even though the relative changes (trends) are captured effectively, some numerical disagreement will appear in the predictions. We see this in Fig. 5(e and f) where a large discrepancy is observed between experimental and predicted phase fractions at high Al but the trend is nonetheless captured. Secondly, the ML model is trained on only 323 discrete compositions which span a multitude of alloy systems, and thus the non-linear decision boundaries identified are susceptible to be translated from their actual position. We see this in Fig. 5(e and f), where predicted FCC fraction increases as Al decreases from 18 to 15.4 at.% but the actual FCC fraction remains almost same, and in Fig. 6(e and f), where the complete transition from FCC to BCC is predicted at a slightly higher Al concentration as compared to experimental observation.

## 4. Conclusions

A machine-learning model was developed for prediction of phases and relative phase-fractions based on a dataset of 323 HEAs. Our approach leverages a model-averaging technique on an ensemble of 150 ANNs to predict the presence of FCC solid solution, BCC solid solution, IM compounds or a combination of these phases along with the phase fractions. The final description of the alloy phases is achieved through a three-element vector which is converted into a probabilistic prediction of relative phase-fractions through an activation function (as described in Fig. 1.c). In addition to phase selection predictions and trends in a variety of test alloys, we also demonstrate the ensemble model to have

excellent quantitative agreement with the phase fractions in three systems:  $\text{Fe}_x\text{Ni}_y\text{-}(\text{AlCoCr}_{0.5})_{1-x-y}$ ,  $\text{Al}_x\text{Ti}_y\text{-}(\text{CrFeNi})_{1-x-y}$  and  $\text{Cr}_x\text{Mo}_y\text{-}(\text{VNbTi})_{1-x-y}$ . The model predictions were compared with experimental results when available ( $\text{Fe}_x\text{Ni}_y\text{-}(\text{AlCoCr}_{0.5})_{1-x-y}$  and  $\text{Al}_x\text{Ti}_y\text{-}(\text{CrFeNi})_{1-x-y}$ ) as well as with CALPHAD calculations ( $\text{Cr}_x\text{Mo}_y\text{-}(\text{VNbTi})_{1-x-y}$ ). In general, it was observed that the model performs better in the absence of strong ordering drivers such as Al. The model developed here offers the potential for high-throughput assessment of phase selection and microstructure assemblage in high-entropy systems, which we have demonstrated in this paper with examples in diverse systems. Specifically, we have shown the model to correctly predict –

- FCC → BCC phase transition accompanying the increase in Al concentration in (CrFeNi) and formation of IM induced by addition of Ti in (CrFeNi).
- Phase fractions of FCC and (BCC + IM) phases for different compositions in  $\text{Al}_x\text{Ti}_y\text{-}(\text{CrFeNi})_{1-x-y}$ .
- BCC → FCC phase transition in  $\text{Fe}_x\text{AlCoCr}_{0.5}\text{Ni}_{2.5}$  as Fe concentration increases since it indirectly leads to a decrease in overall Al concentration.
- Presence of only BCC phase in  $\text{Cr}_x(\text{Mo}_{0.5}\text{NbTi})_{1-x}$  at low Cr concentrations ( $x < 0.286$ ) and formation of IM at high Cr concentrations ( $x \geq 0.286$ ).

When averaged over 320 alloys, the model was shown to have an accuracy of 84% for prediction of individual combinations of FCC, BCC, IM, FCC + BCC, FCC + IM, BCC + IM and FCC + BCC + IM. To our knowledge, this model is the first of its kind that enables not only classification of phases, but efficient and accurate predictions of phase fractions as well.

#### CRediT authorship contribution statement

**D. Beniwal:** Data Curation, Formal Analysis, Investigation, Methodology, Visualization, Writing – original draft, Writing – review and editing. **P.K.Ray:** Conceptualization, Formal Analysis, Funding Acquisition, Project Administration, Supervision, Writing – original draft, Writing – review and editing

#### Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

#### Acknowledgements

One of the authors (Pratik K. Ray) would like to acknowledge seed funding (ISIRD Phase-I) from IIT Ropar made available to new faculty members for research initiation.

#### Data availability statement

The processed dataset, along with the necessary code to train ANN models, required to reproduce these findings are available to download from [https://github.com/Dishant1389/IDEAs-ML-HEAs-Phase\\_Predictions.git](https://github.com/Dishant1389/IDEAs-ML-HEAs-Phase_Predictions.git).

Additional data related to the work is available from the authors upon request.

#### Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.commatsci.2021.110647>.

#### References

- B. Cantor, I.T.H. Chang, P. Knight, A.J.B. Vincent, Microstructural development in equiatomic multicomponent alloys, *Mater. Sci. Eng., A* 375–377 (2004) 213–218, <https://doi.org/10.1016/j.msea.2003.10.257>.
- J.-W. Yeh, S.-K. Chen, S.-J. Lin, J.-Y. Gan, T.-S. Chin, T.-T. Shun, C.-H. Tsau, S.-Y. Chang, Nanostructured High-entropy alloys with multiple principal elements: novel alloy design concepts and outcomes, *Adv. Eng. Mater.* 6 (5) (2004) 299–303, [https://doi.org/10.1002/\(ISSN\)1527-264810.1002/adem.v6:510.1002/adem.200300567](https://doi.org/10.1002/(ISSN)1527-264810.1002/adem.v6:510.1002/adem.200300567).
- S. Singh, N. Wanderka, B.S. Murty, U. Glatzel, J. Banhart, Decomposition in multi-component AlCoCrCuFeNi high-entropy alloy, *Acta Mater.* 59 (1) (2011) 182–190.
- E. Ma, X. Wu, Tailoring heterogeneities in high-entropy alloys to promote strength–ductility synergy, *Nat. Commun.* 10 (2019) 5623, <https://doi.org/10.1038/s41467-019-13311-1>.
- Y. Zhang, X. Yang, P.K. Liaw, Alloy design and properties optimization of high-entropy alloys, *JOM* 64 (7) (2012) 830–838.
- Y. Zhang, T.T. Zuo, Z. Tang, M.C. Gao, K.A. Dahmen, P.K. Liaw, Z.P. Lu, Microstructures and properties of high-entropy alloys, *Prog. Mater. Sci.* 61 (2014) 1–93, <https://doi.org/10.1016/j.pmatsci.2013.10.001>.
- A. Roy, T. Babuska, B. Krick, G. Balasubramanian, Machine learned feature identification for predicting phase and Young's modulus of low-, medium- and high-entropy alloys, *Scr. Mater.* 185 (2020) 152–158, <https://doi.org/10.1016/j.scriptamat.2020.04.016>.
- Y. Li, W. Guo, Machine-learning model for predicting phase formations of high-entropy alloys, *Phys. Rev. Mater.* 3 (2019), 095005, <https://doi.org/10.1103/PhysRevMaterials.3.095005>.
- Y. Zhang, C. Wen, C. Wang, S. Antonov, D. Xue, Y. Bai, Y. Su, Phase prediction in high entropy alloys with a rational selection of materials descriptors and machine learning models, *Acta Mater.* 185 (2020) 528–539, <https://doi.org/10.1016/j.actamat.2019.11.067>.
- Z. Zhou, Y. Zhou, Q. He, Z. Ding, F. Li, Y. Yang, Machine learning guided appraisal and exploration of phase design for high entropy alloys, *NPJ Comput. Mater.* 5 (2019) 1–9, <https://doi.org/10.1038/s41524-019-0265-1>.
- W. Huang, P. Martin, H.L. Zhuang, Machine-learning phase prediction of high-entropy alloys, *Acta Mater.* 169 (2019) 225–236, <https://doi.org/10.1016/j.actamat.2019.03.012>.
- E. Menou, I. Toda-Caraballo, P.E.J. Rivera-Díaz-del-Castillo, C. Pineau, E. Bertrand, G. Ramstein, F. Tancret, Evolutionary design of strong and stable high entropy alloys using multi-objective optimisation based on physical models, statistics and thermodynamics, *Mater. Des.* 143 (2018) 185–195, <https://doi.org/10.1016/j.matdes.2018.01.045>.
- X. Yang, Y. Zhang, Prediction of high-entropy stabilized solid-solution in multi-component alloys, *Mater. Chem. Phys.* 132 (2–3) (2012) 233–238, <https://doi.org/10.1016/j.matchemphys.2011.11.021>.
- S. Guo, Q. Hu, C. Ng, C.T. Liu, More than entropy in high-entropy alloys: Forming solid solutions or amorphous phase, *Intermetallics* 41 (2013) 96–103, <https://doi.org/10.1016/j.intermet.2013.05.002>.
- A.K. Singh, N. Kumar, A. Dwivedi, A. Subramaniam, A geometrical parameter for the formation of disordered solid solutions in multi-component alloys, *Intermetallics* 53 (2014) 112–119, <https://doi.org/10.1016/j.intermet.2014.04.019>.
- Z. Wang, Y. Huang, Y. Yang, J. Wang, C.T. Liu, Atomic-size effect and solid solubility of multicomponent alloys, *Scr. Mater.* 94 (2015) 28–31, <https://doi.org/10.1016/j.scriptamat.2014.09.010>.
- Y.F. Ye, Q. Wang, J. Lu, C.T. Liu, Y. Yang, The generalized thermodynamic rule for phase selection in multicomponent alloys, *Intermetallics* 59 (2015) 75–80, <https://doi.org/10.1016/j.intermet.2014.12.011>.
- O.N. Senkov, D.B. Miracle, A new thermodynamic parameter to predict formation of solid solution or intermetallic phases in high entropy alloys, *J. Alloy. Compd.* 658 (2016) 603–607, <https://doi.org/10.1016/j.jallcom.2015.10.279>.
- S. Guo, C. Ng, J. Lu, C.T. Liu, Effect of valence electron concentration on stability of fcc or bcc phase in high entropy alloys, *J. Appl. Phys.* 109 (10) (2011) 103505, <https://doi.org/10.1063/1.3587228>.
- S.Y. Lee, S. Byeon, H.S. Kim, H. Jin, S. Lee, Deep learning-based phase prediction of high-entropy alloys: Optimization, generation, and explanation, *Mater. Des.* 197 (2021) 109260, <https://doi.org/10.1016/j.matdes.2020.109260>.
- S. Gorse, M.H. Nguyen, O.N. Senkov, D.B. Miracle, Database on the mechanical properties of high entropy alloys and complex concentrated alloys, *Data in Brief* 21 (2018) 2664–2678, <https://doi.org/10.1016/j.dib.2018.11.111>.
- J.-P. Couzinié, O.N. Senkov, D.B. Miracle, G. Dirras, Comprehensive data compilation on the mechanical properties of refractory high-entropy alloys, *Data in Brief* 21 (2018) 1622–1641, <https://doi.org/10.1016/j.dib.2018.10.071>.
- A.R. Miedema, P.F. de Châtel, F.R. de Boer, Cohesion in alloys — fundamentals of a semi-empirical model, *Physica B+C* 100 (1) (1980) 1–28, [https://doi.org/10.1016/0378-4363\(80\)90054-6](https://doi.org/10.1016/0378-4363(80)90054-6).
- R.F. Zhang, S.H. Zhang, Z.J. He, J. Jing, S.H. Sheng, Miedema Calculator: A thermodynamic platform for predicting formation enthalpies of alloys within framework of Miedema's Theory, *Comput. Phys. Commun.* 209 (2016) 58–69, <https://doi.org/10.1016/j.cpc.2016.08.013>.
- J.D. Eshelby, The Continuum Theory of Lattice Defects, in: F. Seitz, D. Turnbull (Eds.), *Solid State Physics*, Academic Press, 1956, pp. 79–144, [https://doi.org/10.1016/S0081-1947\(08\)60132-0](https://doi.org/10.1016/S0081-1947(08)60132-0).
- J. Friedel, Electronic structure of primary solid solutions in metals, *Adv. Phys.* 3 (12) (1954) 446–507, <https://doi.org/10.1080/00018735400101233>.

- [27] A. Takeuchi, A. Inoue, Calculations of Mixing Enthalpy and Mismatch Entropy for Ternary Amorphous Alloys, *Mater. Trans.* 41 (2000) 1372–1378.
- [28] D.G. Pettifor, A Quantum-Mechanical Critique of the Miedema Rules for Alloy Formation, in: H. Ehrenreich, D. Turnbull (Eds.), *Solid State Physics*, Academic Press, 1987, pp. 43–92, [https://doi.org/10.1016/S0081-1947\(08\)60690-6](https://doi.org/10.1016/S0081-1947(08)60690-6).
- [29] L. Breiman, Random Forests, *Machine Learn.* 45 (2001) 5–32, <https://doi.org/10.1023/A:1010933404324>.
- [30] K. Team, Keras documentation: RMSprop, (n.d.). <https://keras.io/api/optimizers/rmsprop/>, (accessed April 19, 2021).
- [31] K. Team, Keras documentation: Probabilistic losses, (n.d.). [https://keras.io/api/losses/probabilistic\\_losses/#binarycrossentropy-class](https://keras.io/api/losses/probabilistic_losses/#binarycrossentropy-class), (accessed April 19, 2021).
- [32] U. Mizutani, Hume-Rothery rules for structurally complex alloy phases, *MRS Bull.* 37 (2012) 169, <https://doi.org/10.1557/mrs.2012.45>.
- [33] P. Singh, A. Sharma, A.V. Smirnov, M.S. Diallo, P.K. Ray, G. Balasubramanian, D. D. Johnson, Design of high-strength refractory complex solid-solution alloys, *NPJ Comput. Mater.* 4 (2018) 1–8, <https://doi.org/10.1038/s41524-018-0072-0>.
- [34] P. Singh, A.V. Smirnov, D.D. Johnson, Atomic short-range order and incipient long-range order in high-entropy alloys, *Phys. Rev. B* 91 (2015), 224204.
- [35] X. Sun, H. Zhang, S. Lu, X. Ding, Y. Wang, L. Vitos, Phase selection rule for Al-doped CrMnFeCoNi high-entropy alloys from first-principles, *Acta Mater.* 140 (2017) 366–374, <https://doi.org/10.1016/j.actamat.2017.08.045>.
- [36] S. Gao, T. Kong, M. Zhang, X. Chen, Y.W. Sui, Y.J. Ren, J.Q. Qi, F.X. Wei, Y.Z. He, Q.K. Meng, Z. Sun, Effects of titanium addition on microstructure and mechanical properties of CrFeNiTi<sub>x</sub> (x = 0.2–0.6) compositionally complex alloys, *J. Mater. Res.* 34 (5) (2019) 819–828, <https://doi.org/10.1557/jmr.2019.40>.
- [37] M. Liu, W. Xu, S. Zhang, Z. Wang, Z. Wang, B. Wang, D. Wang, F. Li, Microstructures and hardnesses of AlCoCr0.5FeNi2.5 high entropy alloys with equal valence electron concentration, *J. Alloy. Compd.* 824 (2020) 153881, <https://doi.org/10.1016/j.jallcom.2020.153881>.
- [38] P. Singh, S. Gupta, S. Thimmaiah, B. Thoeny, P.K. Ray, A.V. Smirnov, D. D. Johnson, M.J. Kramer, Vacancy-mediated complex phase selection in high entropy alloys, *Acta Mater.* 194 (2020) 540–546.
- [39] P.K. Ray, M. Akinc, M.J. Kramer, Applications of an extended Miedema's model for ternary alloys, *J. Alloy. Compd.* 489 (2) (2010) 357–361.
- [40] R. Hu, H.-N. Su, P. Nash, Enthalpies of formation and lattice parameters of B2 phases in Al-Ni-X systems, *Pure Appl. Chem.* 79 (2007) 1653–1673.
- [41] W.-R. Wang, W.-L. Wang, S.-C. Wang, Y.-C. Tsai, C.-H. Lai, J.-W. Yeh, Effects of Al addition on the microstructure and mechanical property of Al<sub>x</sub>CoCrFeNi high-entropy alloys, *Intermetallics* 26 (2012) 44–51, <https://doi.org/10.1016/j.intermet.2012.03.005>.
- [42] J.Y. He, W.H. Liu, H. Wang, Y. Wu, X.J. Liu, T.G. Nieh, Z.P. Lu, Effects of Al addition on structural evolution and tensile properties of the FeCoNiCrMn high-entropy alloy system, *Acta Mater.* 62 (2014) 105–113, <https://doi.org/10.1016/j.actamat.2013.09.037>.
- [43] X. Wang, Q. Liu, Y. Huang, L.u. Xie, Q. Xu, T. Zhao, Effect of Ti Content on the Microstructure and Corrosion Resistance of CoCrFeNiTi<sub>x</sub> High Entropy Alloys Prepared by Laser Cladding, *Materials (Basel.)* 13 (10) (2020) 2209, <https://doi.org/10.3390/ma13102209>.
- [44] A. Roy, P. Sreeramagiri, T. Babuska, B. Krick, P.K. Ray, G. Balasubramanian, Lattice distortion as an estimator of solid solution strengthening in high-entropy alloys, *Mater. Charact.* 172 (2021) 110877, <https://doi.org/10.1016/j.matchar.2021.110877>.
- [45] M. Zhang, Y. Ma, W. Dong, X. Liu, Y.e. Lu, Y. Zhang, R. Li, Y. Wang, P. Yu, Y. Gao, G. Li, Phase evolution, microstructure, and mechanical behaviors of the CrFeNiAl<sub>x</sub>Ti<sub>y</sub> medium-entropy alloys, *Mater. Sci. Eng., A* 771 (2020) 138566, <https://doi.org/10.1016/j.msea.2019.138566>.
- [46] P. Singh, A.V. Smirnov, A. Alam, D.D. Johnson, First-principles prediction of incipient order in arbitrary high-entropy alloys: exemplified in Ti0.25CrFeNiAl<sub>x</sub>, *Acta Mater.* 189 (2020) 248–254, <https://doi.org/10.1016/j.actamat.2020.02.063>.
- [47] K.P. Gupta, The Co-Ni-Ti system (cobalt-nickel-titanium), *Journal of Phase Equilibria*, 20 (1) (1999) 65–72.
- [48] C. Xiang, H.M. Fu, Z.M. Zhang, E.-H. Han, H.F. Zhang, J.Q. Wang, G.D. Hu, Effect of Cr content on microstructure and properties of Mo0.5VNbTiCr<sub>x</sub> high-entropy alloys, *J. Alloy. Compd.* 818 (2020) 153352, <https://doi.org/10.1016/j.jallcom.2019.153352>.
- [49] F. Despagne, D.L. Massart, Neural networks in multivariate calibration, *Analyst* 123 (1998) 157R–178R, <https://doi.org/10.1039/A805562I>.
- [50] G. Montavon, W. Samek, K.-R. Müller, Methods for interpreting and understanding deep neural networks, *Digital Sig. Process.* 73 (2018) 1–15, <https://doi.org/10.1016/j.dsp.2017.10.011>.
- [51] J.M. Benítez, J.L. Castro, I. Requena, Are artificial neural networks black boxes? *IEEE Trans. Neural Networks* 8 (5) (1997) 1156–1164, <https://doi.org/10.1109/TNN.7210.1109/72.623216>.