Interpretable Machine Learning Design and Discovery of Complex-composition Metallic Materials

Jie Qi
Xinzhou, Shanxi, China

Bachelor of Science, Hong Kong University of Science and Technology, 2012

A Dissertation Presented to the Graduate Faculty
of the University of Virginia in Candidacy for the Degree of
Doctor of Philosophy

Department of Physics

University of Virginia
September 2022

Prof. S. Joseph Poon, Chair

Prof. Gia-Wei Chern

Prof. Utpal Chatterjee

Prof. John R. Scully

Interpretable Machine Learning Design and Discovery of Complex-composition Metallic Materials

Jie Qi

Abstract

High-entropy alloy (HEA), also known as Complex Concentrated Alloy (CCA) or Multi-principal Elements Alloys (MPEA), is a type of metallic material with multiple principal elements. Compared to conventional alloys, HEA engenders vast opportunities for designing new materials with desirable structural and functional properties such as mechanical properties, thermal/electric conductivity, corrosion resistance, super-conductivity, radiation absorption, and hydrogen storage. However, the HEA compositional space is exceedingly large due to the nature of possessing multiple principal elements. Therefore, fundamental challenges arise in efficiently exploiting compositions with exceptional features. HEA compositions and processing methods control the formation of HEA phases, including solid-solution phases (SS) and intermetallic phases (IM). Phases of HEA determine the properties and need to be carefully designed. In this dissertation, we will discuss Machine Learning (ML) based HEA phase prediction methods, interpretation of HEA phase formation, HEA properties prediction, high-throughput HEA design methodology, and experimental synthesis of HEA with desired properties.

As HEA phase prediction methods have evolved from single physic-based parameters to first-principles calculations and ML approaches, the accuracy and capability of the HEA phase prediction methods are continuously improving. However, the prediction of HEA phases, especially the IM, is still underdeveloped due to the expensive computing power needed for first-principles calculations, or the lack of appropriate ML features and the limited dataset needed for ML approaches. To address these issues, we developed novel ML models with detailed phase classification and high accuracies, where nine phase categories can be predicted with accuracies close to 90 %. This model utilizes an innovative set of phenomenological ML features mined from binary phase diagrams and the feature engineering technique. 86 new HEAs were synthesized to validate the model's accuracy. The HEA phase formation interpretation has significant scientific

importance. ML normally provide accurate predictions without disclosing the science behind it. Therefore, we identified and interpreted the key scientific factors controlling IM formation, guiding the HEA design and in-depth studies on phase formation.

The prediction of HEA properties includes melting temperature, density, cost, and mechanical properties (e.g., hardness, tensile/compression yield strength, and fracture strain). Various prediction methods will be introduced for different properties. Based on the phase and property prediction models, a comprehensive high-throughput HEA design method is developed to search for compositions with desired phase and properties. Through the use of this method, a series of HEAs have been designed specifically for marine environments, demonstrating the effectiveness of the method.

Acknowledgements

Over the past twenty-eight years, I have been learning and attempting to make a small corner of science accessible to the world. During this process, I was fortunate to receive significant support from talented and professional people. It is my pleasure to express my gratitude to them and specifically acknowledge a few individuals here.

Firstly, my sincere gratitude goes to my advisor, Prof. S. Joseph Poon, for his constant support and guidance. Critical thinking, enthusiasm, diligence, creativity, leadership, and collaboration are what I learned over these years, and I will keep them in mind as I move forward.

I want to express my gratitude to Professors John R. Scully, Sean R. Agnew, and Peter K. Liaw for sharing their knowledge and experience during collaboration. In addition, I would like to thank the members of my research and dissertation committees, Professors Simonetta Liuti, Gia-Wei Chern, Utpal Chatterjee, and John R. Scully, for their insightful comments and inspiring questions during the committee meeting and thesis defense.

It has been a pleasure to share our laboratory with my colleagues Wei Zhou, Dr. Sheng Gao, Dr. Xixiao Hu, Dr. Chung Ting Ma, Diego Ibarra Hoyos, and Dr. Mousumi Mitra. The beautiful plasma in the vacuum chamber, the cleanroom we built together, and the shiny metal samples we synthesized will always be cherished memories for me. Moreover, I would like to thank Mark A. Wischhusen, Samuel B. Inman, Xuesong Fan, and Dr. Rui Feng for the projects on which we collaborated and produced fascinating results. And last but not least, thank you, Dr. Andrew M. Cheung, for being my very close friend and helping me with alloy synthesis and manuscript development.

Over the years, my friendships with members of the Physics department, basketball team, pilot community, and other communities have been valuable to me. Despite not being able to mention everyone here, please accept my sincere gratitude.

Most importantly, I want to express my gratitude to my family, Dehua Qi, Yindi Zhang, and Yiyuan Qi. Throughout my life, they have always encouraged me to pursue my dreams without fear. And a special thank you to my girlfriend Dr. Xiaohan Du. Having your love, companionship, and emotional support is the most valuable thing in life. They will always be in my heart.

(Support: This work is supported by the Office of Naval Research grant N00014-18-1-2621.)

Content

List of Figures

List of Tables

Chapter 1 Introduction

1.1 Introduction to high entropy alloys

Metallic material has been utilized and optimized by humans since the discovery of bronze. Most alloys were developed based on one principal element with small amounts of other alloying elements. Examples are stainless steel, aluminum-based alloys, and nickel-based superalloys. The attempts to make multi-principle element alloys (MPEA) can be dated back to the last century, while the outbreak in this field started in 2004 when Dr. Brian Cantor[1] and Dr. Jien-Wei Yeh[2] independently found two MPEAs in single face-centered cubic structure. The nature of multiple principal elements results in more thermodynamic microscopic configurations and a higher configurational entropy. Therefore, this type of alloy is also known as High Entropy Alloy (HEA) or Complex Component Alloy (CCA). We will use “HEA” to represent this type of alloys in this thesis.

The early definition of HEA is that the alloy composes at least five elements, and the concentration of each element is between 5 and 35 at.%[3]. Subsequently, these limitations were lifted and HEA now represents any alloy with multiple principal elements[3].

Four “Core Effects” of HEA were proposed by Yeh[4] as the high-entropy effect, the lattice distortion effect, the sluggish diffusion effect, and the “cocktail” effect as the four core effects for HEA. HEA’s high entropy can suppress ordered intermetallic phases (IM) formation and stabilize disordered solid solution phases (SS). Compared to the conventional alloy with one principal element, HEA generally has an increased lattice distortion due to the atomic size mismatch among the different elements. The severe lattice distortion leads to effects such as broadened peaks in X-ray diffraction (XRD) patterns, increased material hardness and strength, and higher electrical/thermal resistance. The sluggish diffusion reflects the slow atomic diffusion in HEA, which requires a longer time or a higher temperature to reach the thermal equilibrium state. Finally, the ‘cocktail’ effect generally reminds people of the unexpected excellent properties obtained from

HEA. This encourages people to search the unconventional elemental combinations for unexpected outcomes.

1.2 High entropy alloy phases

Phases (crystal structures) of HEA can be classified in a variety of ways. The first classification system is based on the number of phases. HEA phases can be classified as amorphous phase (AM), single phase (SP), and mixed phase (MP). AM corresponds to the alloy without long-range ordered crystal structure, such as the metallic glass. SP or MP corresponds to alloys with single or multiple phases. The HEA phases in the second classification system can be classified as SS and IM. SS corresponds to the phases where atoms randomly occupy the lattice sites. Examples are disordered face-centered cubic (FCC), disordered body-centered cubic (BCC), and disordered hexagonal-close-packed structure (HCP). IM corresponds to the phases with long-range ordering, where certain atoms tend to occupy specific sites on the lattice. Common ordered IM in HEA include B2, $L1_2$, $L2_1$ (or Heusler), Laves (C14, C15, C15b, and C36), Sigma ($D8_b$), χ(A12), and η ($D0_{24}$) phases.

HEA phases are important in determining structural and functional properties. For example, FCC phases are normally ductile. Heusler[5] and B2 phases can improve the HEA structural and functional properties[6–10]. Heusler phases are known for their wide range of multifunctional properties, including magneto-optical, magnetocaloric, and spintronic properties[5]. In addition, the Heusler phase is reported to have a superior creep resistance[11,12], and its presence in a HEA SS host can improve the mechanical properties[8,13,14]. The B2 phase in HEAs generally consists of two types: AlNi type[9] and Al-X-Y type (X and Y are specific groups of refractory elements)[15]. The AlNi type is widely used as a strengthening precipitate in HEAs[9], while the Al-X-Y type can improve high-temperature mechanical properties, lower physical density and cost, and enhance oxidization-resistance over the traditional disordered BCC refractory HEAs[15]. On the other hand, some IM, such as Sigma and Laves phases, are well-known for their embrittling effects[16,17].

The composition of HEA plays a major role in determining phase formation. Nevertheless, different phases can be formed by the same alloy composition. On the one hand, the equilibrium phases at different solidification temperatures are normally different. Therefore, the annealing temperature should be controlled to form the correct phase. On the other hand, HEA synthesis may result in meta-stable instead of equilibrium phases due to rapid solidification in fast cooling. Other processing techniques, such as rolling and hot pressing, can also adjust phase morphologies. Therefore, it is essential to design compositions and processing techniques carefully to obtain HEA with the desired phase.

1.3 High entropy alloy properties and applications

HEA properties include structural properties and functional properties.

The structural properties include, but are not limited to, compression or tensile strain and stress, hardness, and toughness at cryogenic, room, or high temperatures. Figure 1.1 shows that some HEAs have better tensile mechanical properties than conventional alloys. With the formation of certain IM (e.g., Sigma and Laves), HEA Vickers hardness can be improved to the values of greater than 600 Hv from 100 - 200 Hv with a single FCC phase[18,19]. In addition, refractory HEAs possess excellent high-temperature functional properties. According to the Figure 1.2, some refractory HEAs retain high compressive yield strength up to 1000 ℃, making them the ideal materials in the gas turbine industry.

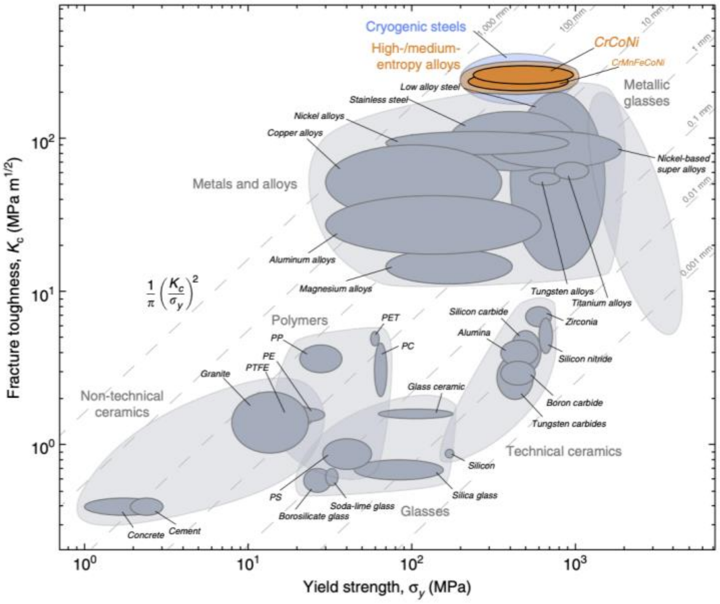


Figure 1.1: Ashby plot comparing the tensile yield strength and fracture toughness among different material systems. High entropy alloys generally show better properties. Figure from Gludovatz met al.[20]

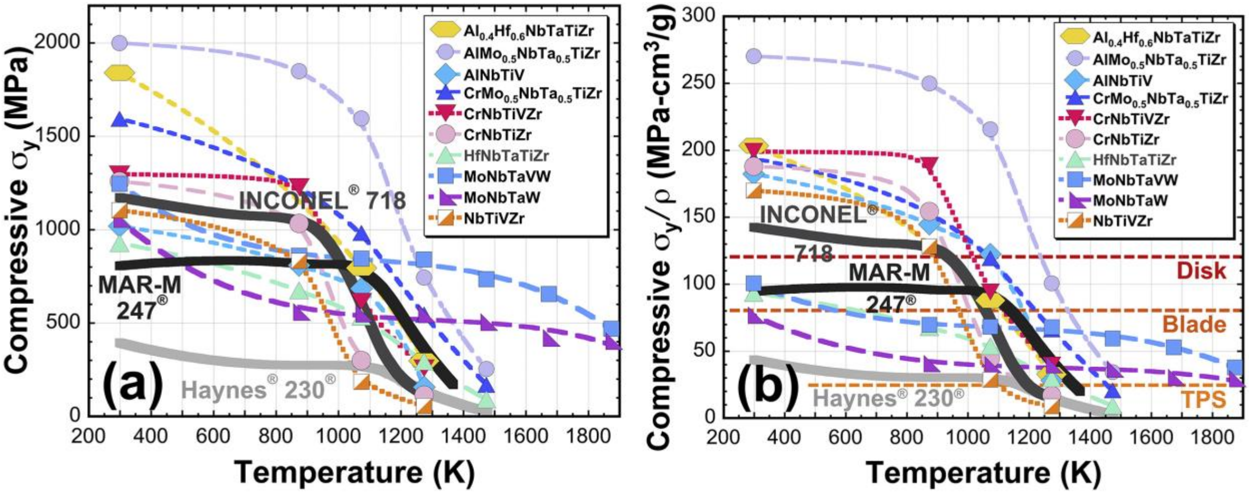


Figure 1.2: (a) Compression yield strength, (b) Compression yield strength normalized by density versus the measurement temperature. Haynes 230, INCONEL 718, and MAR-M 247 are commercial super alloys with their tensile data shown as comparison. Figures from D.B. Miracle and O.N. Senkov[3].

HEAs also have great functional properties, such as the corrosion resistance[21–23], irradiation resistance[24], thermal[25], electrical[26,27], and magnetic[27–29] properties. In addition, HEAs' applications as superconducting material[30], hydrogen storage material[31–33], catalysts[34], and thermoelectric[35–37] material have also been explored extensively.

Other fundamental properties, such as density and cost, are also critical in material selection. Alloy design is an art where multiple materials properties should be optimized simultaneously. Exploration, exploitation, optimization, and trade-off are the cores of this art.

1.4 Challenges in high entropy alloy research

The greatest challenge in HEA is efficiently exploiting the exceedingly large compositional space. For example, using only the 20 common metallic elements on the periodic tables can form over 20,000 ternary, quaternary, and quinary element combinations. The number of compositions can amount to trillion possibilities when considering the atomic percentage variation or higher-order alloy systems. However, over the past 20 years, only ~1,000 HEA systems have been experimentally studied[38]. Therefore, the trial-and-error method is outdated for finding the desired HEA in the vast compositional space. The high-throughput simulation and high-throughput experimentation may solve this issue.

To save experimental efforts, high-throughput simulation method designs HEAs by predicting phases and properties. Simulation tools such as Monte Carlo (MC), ab-intio simulation, Density Function Theorem (DFT), Calculation of Phase Diagrams (CALPHAD), and Machine learning (ML) are widely used. They are, however, challenged by problems such as the unsatisfactory accuracy, a large amount of computation power, and the absence of thermodynamic databases. In this thesis, we will introduce various phase and properties prediction methods. They will be integrated as a high-throughput HEA design model.

High-throughput experimentation synthesizes and characterizes the gradient of compositions during the experiment. Chemical vapor depositions, sputtering, and additive manufacturing are the common synthesis techniques. These methods can quickly enlarge the experimental database and may identify the composition with desired properties. However, the large-scale alloy synthesis in high-throughput experiment may behave differently from the with traditional casting methods.

1.5 Structure of this thesis

This thesis will review the methodologies for predicting the HEA phase and properties, and then introduce our methods. The major HEA phase prediction methods over the past 20 years are reviewed in Chapter 2, including the single parameters, first-principles and thermodynamic calculations, and ML methods. Our ML phase prediction models are introduced in Chapter 3, which covers the creation of ML features, the construction of feature engineering algorithms, the interpretation of features, and the experimental validation. The prediction methods for various HEA properties are reviewed in Chapter 4. Chapter 5 presents a high-throughput HEA design method that can search for HEAs with desired phase and properties. The experimental techniques used in our research are introduced in Chapter 6, along with some experimental results for the HEAs we designed. Finally, Chapter 7 is the summary and outlook.

Chapter 2 Review on High Entropy Alloys Phase Prediction Methods

2.1 Background

The formation of high-entropy phases is primarily controlled by thermodynamic and kinetic factors. For understanding the growing number of HEAs, empirical methods that utilized atomistic and thermodynamic parameters were introduced to investigate HEA compositional regions [3,39,40]. The empirical approaches were later complemented by first-principles calculations [41] and Calculation of Phase Diagrams (CALPHAD) [42,43] to shed light on the thermodynamic factor of HEA formation. Monte Carlo simulations showed promising results in predicting the formation of intermetallic phases and the evolution of phase structures with temperature [44]. Despite progress in understanding the formation trend of HEAs, much of the alloy design for HEAs still relies on trial-and-error experimentation. Recently, there have been increasing efforts in employing data-driven methods to exploit the growing data set of HEAs. Some initial methods included the utilization of statistical models complemented with thermodynamics[45], and as well as high-throughput experimentation [46,47] designed to underpin the HEA phase formation trend. The thin film deposition method employed in the high-throughput study tended to result in the metastable HEA phases. Meanwhile, the utilization of machine learning (ML) models has demonstrated some initial promise in phase prediction and property design (e.g., high hardness)[48,49]. Most ML has focused on supervised learning with different models, such as the support vector machine (SVM)[48–53] and artificial neural network (ANN)[50,53,54]. The ML models were trained using atomistic and thermodynamic parameters. Despite some success in categorizing the compositional regions of certain solid solution phases (SS) and intermetallic phases (IM), the predictions often fell short of differentiating between the specific phases. On the other hand, a ML model that utilized phenomenological features obtained from binary phase diagrams was found to achieve high accuracy in categorizing specific phase formation [55].

The various computation and experimentation methodologies will continue to develop towards advancing the science and design of high-entropy alloys. In view of this promising development, this chapter serves as a timely report of the status of progress in harnessing the high-dimensional composition space of HEAs as a requisite for designing the desired properties.

In this chapter, Section 2.2 will provide a comprehensive review of the various empirical parameters relevant to phase formation. Section 2.3 will introduce several typical first-principles calculation methods in phase prediction. Section 2.4 will provide a comprehensive review of the ML models for HEA phase prediction. Section 2.5 will compare the three types of methods.

2.2 Empirical parameters

Empirical parameters can reflect the HEA phase formation tendency from aspects of the minimization of Gibbs Free Energy, Hume-Rothery rules, electronic configuration, and lattice strain. These empirical parameters, defined and discussed in Sections 0-2.2.3, were used individually or in conjunction to determine HEA phases. The correlations between these parameters and the HEA phases formed are analyzed in Section 2.2.4.

2.2.1 Free energy parameters

The phases favored during the solidification of a HEA possess the lowest Gibbs free energy of mixing (ΔG_{mix}). Parameters associated with ΔG_{mix} are the mixing entropy (ΔS_{mix}), the mixing enthalpy (ΔH_{mix}), and labeled parameters Ω, ϕ, Φ, η, and k_1^{cr}, defined in Eqn. 2.2 - 2.8. ΔG_{mix} is defined as

$$\Delta G_{mix} = \Delta H_{mix} - T\Delta S_{mix} \tag{2.1}$$

where T is the phase formation temperature. The ΔS_{mix} for forming a single or multiple SS is approximated as the configurational entropy[56] (ΔS_C) and is calculated according to Boltzmann's hypothesis (Eqn. 2.2):

$$\Delta S_{mix} \approx \Delta S_C = -R\sum_{i=1}^{N} c_i \ln(c_i) \tag{2.2}$$

where R is the gas constant and c_i is the atomic percentage of the i-th element for a N-component system. The definitions of c_i and N are the same throughout this chapter. A SS formation is energetically favored over an IM formation when the ΔS_{mix} term is larger.

The ΔH_{mix} term represents the chemical compatibility among the elements in HEAs [57]. For HEAs, the ΔH_{mix} for forming a SS is typically calculated from Miedema's model [58] (Eqn. 2.3):

$$\Delta H_{mix} = \sum_{i=1, i\neq j}^{N} 4\, \Delta H_{i,j}^{mix}\, c_i c_j \qquad 2.3$$

where $\Delta H_{I,j}^{mix}$ is the binary mixing enthalpy of an i-j elemental pair. An increase in the negativity of ΔH_{mix} increases the probability of forming an IM. A positive ΔH_{mix} indicates an immiscibility among certain elements, which could lead to phase separation. As shown in Figure 2.3a, $-16\frac{kJ}{mol} < \Delta H_{mix} < +5\frac{kJ}{mol}$ is the criterion proposed for forming a single SS [39].

ΔG_{mix} is determined by the entropy and enthalpy terms. Whether SS or IM formation is favored is dependent on the interplay of these two terms. The parameters Ω and ϕ are used to compare the magnitudes of the entropy and enthalpy terms. Zhang et al. [59] defined the Ω-parameter as

$$\Omega = \frac{T_m \Delta S_{mix}}{|\Delta H_{mix}|} \qquad 2.4$$

where $T_m = \sum_{i=1}^{N} c_i\, T_{m_i}$ is the HEA melting temperature and T_{m_i} is the melting temperature of the i-th element. A large $T_m \Delta S_{mix}$ or $|\Delta H_{mix}|$ term stabilizes SS or IM formation, respectively. When $\Omega > 1$ [59], as shown in Figure 2.3b, a SS formation is favored.

Ye et al. [60] defined the ϕ-parameter. The total configurational entropy of mixing (ΔS_T) is defined as $\Delta S_T = \Delta S_C + \Delta S_E$, where ΔS_C is the configurational entropy of mixing for an ideal gas and ΔS_E is the excessive entropy of mixing [61]. ΔS_T deviates from the approximation of $\Delta S_T \approx \Delta S_C$ due to the influence of factors such as differences in atomic size and the packing fraction. ΔS_E, usually negative, is introduced to represent this deviation and ΔS_T is adjusted by its absolute magnitude, $\Delta S_T = \Delta S_C - |\Delta S_E|$. The parameter ϕ defined as

$$\phi = \frac{\Delta S_C - \left|\frac{\Delta H_{mix}}{T_m}\right|}{|\Delta S_E|} > 1 \qquad 2.5$$

is the result of combining ΔS_T and Ω. Based on existing values of known HEAs, shown in Figure 2.3c, $\phi > 7$ is the proposed range for SS formation [39].

Instead of only comparing the enthalpy and entropy terms for predicting the formation of SS or IM, the parameters Φ, η, and k_1^{cr} were defined by examining difference in ΔG_{mix}. King et al.

[62] defined the Φ-parameter to compare the ΔG_{mix} for forming a fully disordered SS (ΔG_{SS}) with the ΔG_{mix} for IM formation or phase segregation (ΔG_{max}). The Φ-parameter is defined as

$$\Phi = \frac{\Delta G_{SS}}{-|G_{max}|} \qquad 2.6$$

where $|\Delta G_{max}|$ represents the absolute magnitude of the larger of the following two values: the lowest possible negative ΔG_{mix} when the strongest binary compound forms, or the highest possible positive ΔG_{mix} when a phase is segregated due to the positive mixing enthalpy between two specific constituent elements. When $\Phi > 1$, then SS formation is favored.

Troparevsky et al. [63] defined a parameter, later labeled by others as η, that is a first order approximation used to compare the ΔG_{mix} for forming SS and IM. The ΔH_{mix} for SS formation and the ΔS_{mix} for forming IM are usually small. Thus, the entropy contribution $-T_{ann}\Delta S_{mix}$, where T_{ann} is the annealing temperature of a HEA, is used to approximate the ΔG_{mix} for forming SS. The enthalpy of formation (ΔH_f), the most negative binary mixing enthalpy for IM formation among the constituent element pairs derived from density functional theory (DFT) calculations, is used to approximate the ΔG_{mix} for forming IM. η is then defined as

$$\eta = \frac{-T_{ann}\Delta S_{mix}}{|\Delta H_f|} \qquad 2.7$$

where an increasingly larger value of the parameter indicates a favorability for forming a SS. $\eta > 0.19$ [39] is the proposed lower boundary for the region of SS formation as seen in Figure 2.3c.

Similar to the formulation of η, Senkov and Miracle [64] developed parameters to compare the ΔG_{mix} for forming SS and IM. Their approach was less approximate. Their criterion for forming SS is $\Delta H_{mix} - T\Delta S_{mix} < \Delta H_{IM} - T\Delta S_{IM}$, where ΔH_{mix} and ΔH_{IM} are the mixing enthalpies, and ΔS_{mix} and ΔS_{IM} are the mixing entropies for forming SS and IM, respectively. ΔS_{IM} for the IM is approximated to be $0.6\Delta S_{mix}$. This relation of a simple thermodynamic criterion can be expressed as

$$k_1^{cr} = 1 - \frac{0.4T\,\Delta S_{mix}}{\Delta H_{mix}} > \frac{\Delta H_{IM}}{\Delta H_{mix}} \qquad 2.8$$

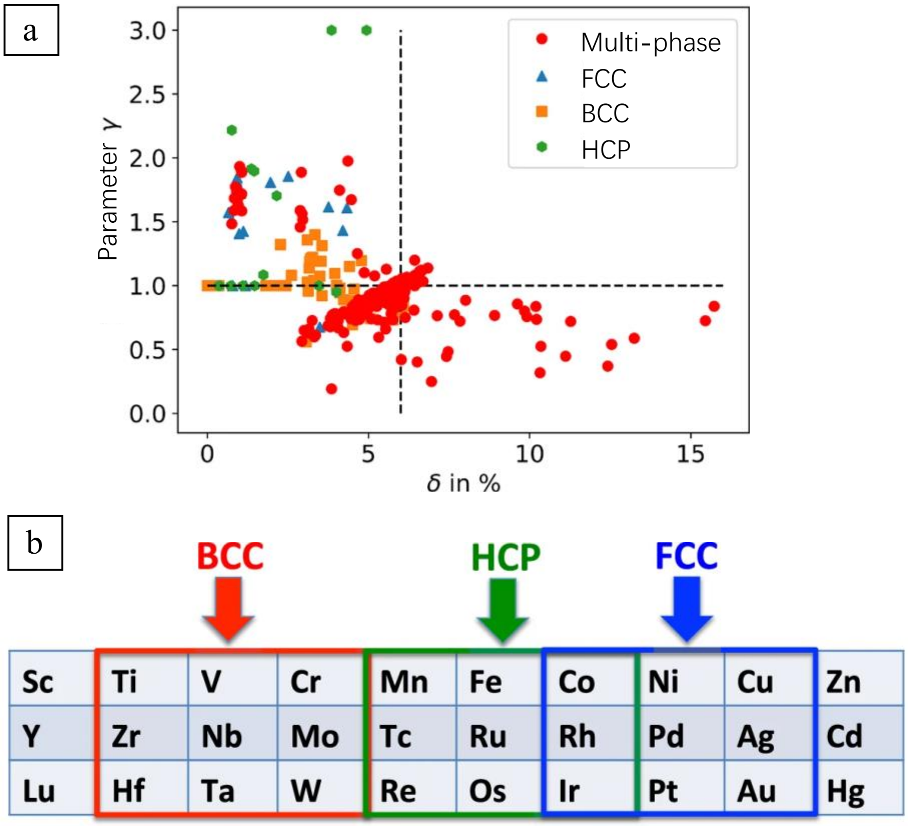
The parameters k_1^{cr} and $\frac{\Delta H_{IM}}{\Delta H_{mix}}$ are plotted in Figure 2.3d. When $k_1^{cr} > \frac{\Delta H_{IM}}{\Delta H_{mix}}$, a SS formation is favored.

Pei et al. [65] defined the parameter γ to compare the ΔG_{mix} for forming single phases and multi-phases. In γ's calculation, $\Delta G_{mix} = \Delta H_{mix} - \alpha T_m \Delta S_c$, where α is a scaling parameter. The

ΔH_{mix} definition varies from other methods by using a combination of the formation enthalpy calculated based on the Lennard-Jones potential, and the strain-induced energy calculated from the Kanzaki force [66]. For any given composition, ΔH_{mix} was calculated for FCC, BCC, HCP, and simple cubic structures. The minimum ΔH_{mix} was adopted to calculate ΔG_{mix}. For the entropy term, the real system entropy was typically smaller than ΔS_c, and the real temperature when the SS was stable could be below T_m. Consequently, α is a scale-down factor for the entropy contribution from the ideal to the real conditions. Optimum phase separation occurred at a value of $\alpha = 0.25$. For a N-component HEA, the ΔG_{mix} was calculated for all the constituent binaries (ΔG_2) and the HEA (ΔG_N). The uniform SS formation ability depended on ΔG_N and the smallest ΔG_2 value, $\min(\Delta G_2)$. Thus, γ was defined as

$$\gamma = \begin{cases} \dfrac{\Delta G_N}{\min(\Delta G_2)}, & \text{if} \min(\Delta G_2) < 0; \\ -\dfrac{\Delta G_N}{\min(\Delta G_2)}, & \text{if } \Delta G_N < 0 \text{ and} \min(\Delta G_2) > 0; \end{cases} \qquad 2.9$$

The criterion for forming the SS was $\gamma \geq 1$. γ was used to test 296 existing HEAs in BCC, FCC, HCP, and multi-phases. While 73 % were classified correctly, when jointly using γ and the radius mismatch ($\delta < 6\,\%$, defined in Section 2.2.2), 81 % consistency was obtained, as demonstrated in Figure 2.1a. The validity of γ was further confirmed by using CALPHAD. γ was calculated for each of the 1,146 equimolar HEAs. The compositions were selected, as seen in Figure 2.1b, from three 9-element blocks in the periodic table. From these blocks, 266 single SS HEAs with 74 BCC, 145 HCP, and 47 FCC phases were obtained. Of the 266 predicted HEAs, only 77 could be validated with CALPHAD due to the limitations of the thermal databases. However, the 77 had a high validation consistency of 94 %.



Sc	Ti	V	Cr	Mn	Fe	Co	Ni	Cu	Zn
Y	Zr	Nb	Mo	Tc	Ru	Rh	Pd	Ag	Cd
Lu	Hf	Ta	W	Re	Os	Ir	Pt	Au	Hg

Figure 2.1: (a) Plot of the distribution of FCC, BCC, HCP, and multi-phase HEAs for parameters γ and δ. The criteria for forming a single SS phase are bounded by $\gamma \geq 1$ and $\delta < 6\,\%$. (b) Equimolar HEAs with BCC, HCP, and FCC predicted phases were obtained from three 9-element blocks in the periodic table. Figures from Pei et al. [65]

2.2.2 Parameters from Hume-Rothery rules

According to the Hume-Rothery rules [67], the formation of a SS is influenced by radius mismatch, electronegativity mismatch, and electron concentration among the constituent elements. Based on this theory, to study HEAs, several parameters that can influence the phase formation were defined.

The intrinsic residual strain, caused by the radius difference, makes the multi-phase formation possible. Parameters δ, $\sqrt{<\varepsilon^2>}$, $\frac{E_2}{E_0}$, s_m, and K_m, defined below, relate HEA phase formation to intrinsic strain.

Small radii differences between constituent elements, equivalent to small lattice distortions, favor the formation of the SS. The radius mismatch of an alloy (δ) [57] is calculated by

$$\delta = \sqrt{\sum_{i=1}^{N} c_i \left[1 - \frac{r_i}{\sum_{j=1}^{N} c_j\, r_j}\right]^2} \tag{2.10}$$

where r_i is the atomic radius of the i-th element. $\delta < 6\ \%$ [39] is the region for SS formation, as seen in Figure 2.3a.

Ye et al. [68] developed a geometric model to calculate the root-mean-square residual strain $\sqrt{<\varepsilon^2>}$ from other parameters such as the atomic percentage, atomic size, and packing density. The mean-square is defined as residual strain $<\varepsilon^2> = \sum_{i=1}^{N} c_i {\varepsilon_i}^2$, where ε_i is the residual strain of the i-th element in a N-component system. After derivation, ε_i can be expressed as

$$\varepsilon_i = \frac{\sum_{j=1}^{N} \omega_{ij} c_j}{\sum_{k=1}^{N} A_{ik} c_k} - \frac{4\pi\eta_{ideal}}{N_i \sum_{k=1}^{N} A_{ik} c_k} \tag{2.11}$$

where ω_{ij} is the solid angle subtended by j-th element around the i-th element with $\omega_{ij} = 2\pi\left[1 - \frac{\sqrt{r_i(r_i+2r_j)}}{r_i+2r_j}\right]$, r_j is the atomic radius of the j-th element, A_{ik} is a dimensionless constant with $A_{ik} = \frac{2\pi x_{ik}}{(x_{ik}+1)^2\sqrt{x_{ik}(x_{ik}+2)}}$ for i-th and k-th elements, η_{ideal} is the ideal atomic packing fraction and is computed by $\eta_{ideal} = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} c_j c_i N_i \left[1 - \frac{\sqrt{x_{ij}(x_{ij}+2)}}{x_{ij}+1}\right]$, x_{ij}=r_i/r_j is the atomic radius ratio, and N_i is the coordinate number of the i-th atom. A significantly large $\sqrt{<\varepsilon^2>}$ leads to lattice distortions that disrupt single-phase lattices and form multi-phase lattices.

Wang et al. [69] defined another parameter $\frac{E_2}{E_0}$ related to the intrinsic elastic strain energy. In an ideal N-component uniform HEA lattice, the average atomic radius is $\bar{r} = \sum_{i=1}^{N} c_i\, r_i$. In a real lattice, atoms are displaced from $\bar{r}$. The dimensionless strain is calculated to be $\Delta d = \frac{|r_i+r_j-2\bar{r}|}{2\bar{r}}$. The dimensionless parameter $\frac{E_2}{E_0}$ is defined as

$$\frac{E_2}{E_0} \propto (\Delta d)^2 = \sum_{j\geq i}^{N} \frac{c_i c_j \left|r_i + r_j - 2\bar{r}\right|^2}{(2\bar{r})^2} \tag{2.12}$$

where a low value of $\frac{E_2}{E_0}$, similar to small values of $\sqrt{< \varepsilon^2 >}$, favors the SS formation. Figure 2.3e is a plot of $\frac{E_2}{E_0}$ and $\sqrt{< \varepsilon^2 >}$ for HEAs with different phases. It demonstrates that when $\frac{E_2}{E_0} < 13.6 \times 10^{-4}$ and $\sqrt{< \varepsilon^2 >} < 6.1$ %, single-phase HEAs tend to form [39].

Interatomic spacing mismatch (s_m) and the bulk modulus mismatch (K_m) were developed by Toda-Caraballo et al. [40,70] with

$$s_m = \sum_{i=1}^{N}\sum_{j=1}^{N} c_i c_j \left|1 - \frac{s_{ij}^d}{s_{lat}}\right| \qquad 2.13$$

and

$$K_m = \sum_{i=1}^{N}\sum_{j=1}^{N} c_i c_j \left|1 - \frac{K_{ij}^d}{K_{lat}}\right| \qquad 2.14$$

where s_{ij}^d and K_{ij}^d are two matrices representing the interatomic spacing and bulk modulus for i-j atom pairs, K_{lat} is the bulk modulus of the lattice, and s_{lat} is the mean interatomic distance across the lattice. Figure 2.2 shows the HEA phase separation based on parameters s_m and K_m. In the plot, SS tend to from when s_m is small. This result, again, implies that small lattice distortion prompts the SS formation. As for the influence of K_m, the FCC phase tends to form when $K_m < 4$, while BCC phase forms when $K_m > 4$, implying that the different forces acting on atoms in a FCC lattice are closer to being homogeneous than those acting on a BCC lattice.

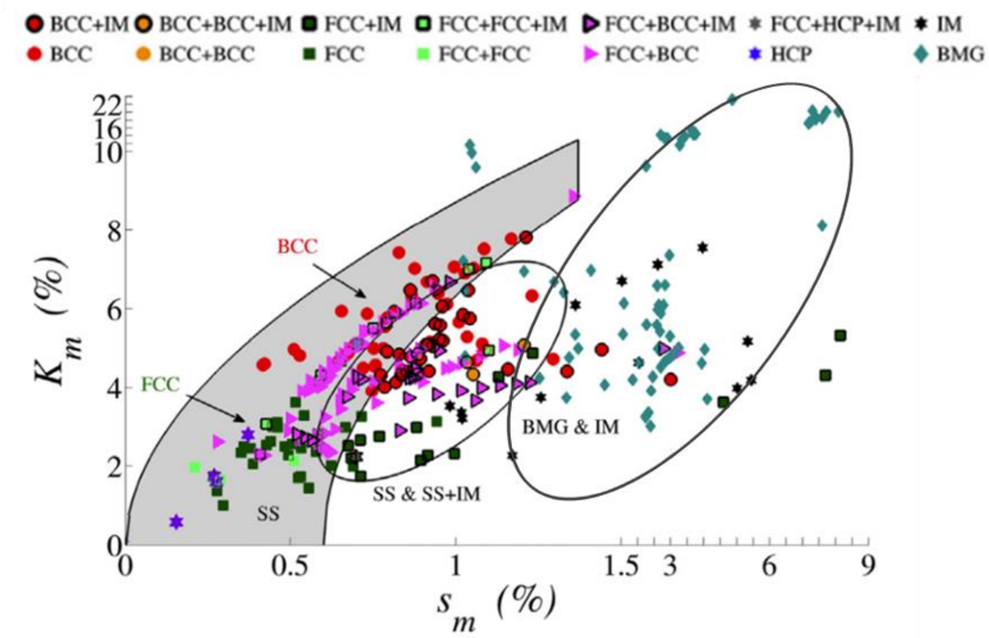


Figure 2.2: The plot of the distribution of HEA phases for parameter s_m and K_m. Figure from Toda-Caraballo and Rivera-Díaz-Del-Castillo [40].

In addition to the effects from the intrinsic strain, electronegativity difference ($\Delta\chi$) and electron configuration are also considered. A small $\Delta\chi$ has been shown to promote SS formation[71]. $\Delta\chi$ is defined as

$$\Delta\chi = \sqrt{\sum_{i=1}^{N} c_i \left[\chi_i - \sum_{j=1}^{N} c_j \chi_j\right]^2} \qquad 2.15$$

where χ_i is the i-th HEA element electronegativity. Dong et al. [72] showed that the formation of the Topological Close-Packed (TCP) phases such as Sigma, Laves, and μ phases can be influenced by $\Delta\chi$ when $\Delta\chi > 0.133$.

Another parameter is the electron concentration, which has two definitions according to Guo et al [73]. The first one is $\frac{e}{a}$ which is the average number of itinerant electrons per atom:

$$\frac{e}{a} = \sum_{i=1}^{N} c_i \left(\frac{e}{a}\right)_i \qquad 2.16$$

where $\left(\frac{e}{a}\right)_{\mathrm{i}}$ is the itinerant electrons per atom of the i-th element. The second one is the valence electron concentration (VEC) [73–75] which is the total number of electrons including the *d*-electrons held in the valence band. VEC is defined as

$$\mathrm{VEC} = \sum_{\mathrm{i}=1}^{\mathrm{N}} \mathrm{c_i}\ \mathrm{VEC_i} \tag{2.17}$$

where $\mathrm{VEC_i}$ is the VEC of the i-th element. VEC was found to be superior to $\frac{e}{a}$ in predicting HEA phases. Figure 2.3f shows that BCC phases form when $\mathrm{VEC} < 6$, FCC phases form when $\mathrm{VEC} > 7.8$, and mixed FCC-BCC phases form when $6 < \mathrm{VEC} < 7.8$ [39]. Tsai et al. [75] applied VEC to a study on σ phase formation. They discovered that $6.88 < \mathrm{VEC} < 7.84$ is the σ-prone formation region for HEAs containing Cr or V.

2.2.3 Other parameters

The average value of a d-orbital energy level ($\overline{\mathrm{Md}}$) was proposed by Lu et al. [76] $\overline{\mathrm{Md}}$ is related to the electronegativity and metallic atomic radii. The TCP formation is influenced by $\overline{\mathrm{Md}}$. TCP phases form when $\overline{\mathrm{Md}} > 1.09$ and no TCP formation occurs when $\overline{\mathrm{Md}} < 0.95$.

Poletti et al.[77] proposed the parameter $\mu = \mathrm{T_m}/\mathrm{T_{SC}}$, where $\mathrm{T_{SC}}$ is the spinodal decomposition temperature. A large gap between $\mathrm{T_m}$ and $\mathrm{T_{SC}}$ can prompt the single SS formation at high temperatures. As a result, $\mu > 1.5$ is the proposed region for single SS formation.

2.2.4 Correlation between the parameters and phase formation

Gao et al.[39] compared the effectiveness of these empirical parameters by coupling and plotting them in Figure 2.3. Although the correlation between the parameters and the phase formation exists, precise phase predictions based solely on pairs of these parameters is challenging. Different phases on the plots overlap with ambiguous separation. Additionally, specific phase content in certain categories, such as “multi-phase” and “IM,” were not included. However, these empirical parameters provide fundamental ideas for applying ML to HEA phase formation research. These parameters are related to different aspects of phase formation. Many ML methods utilize these parameters to consider all the phase formation factors comprehensively and this results in improved predictions. Further details are found in Section 2.4.

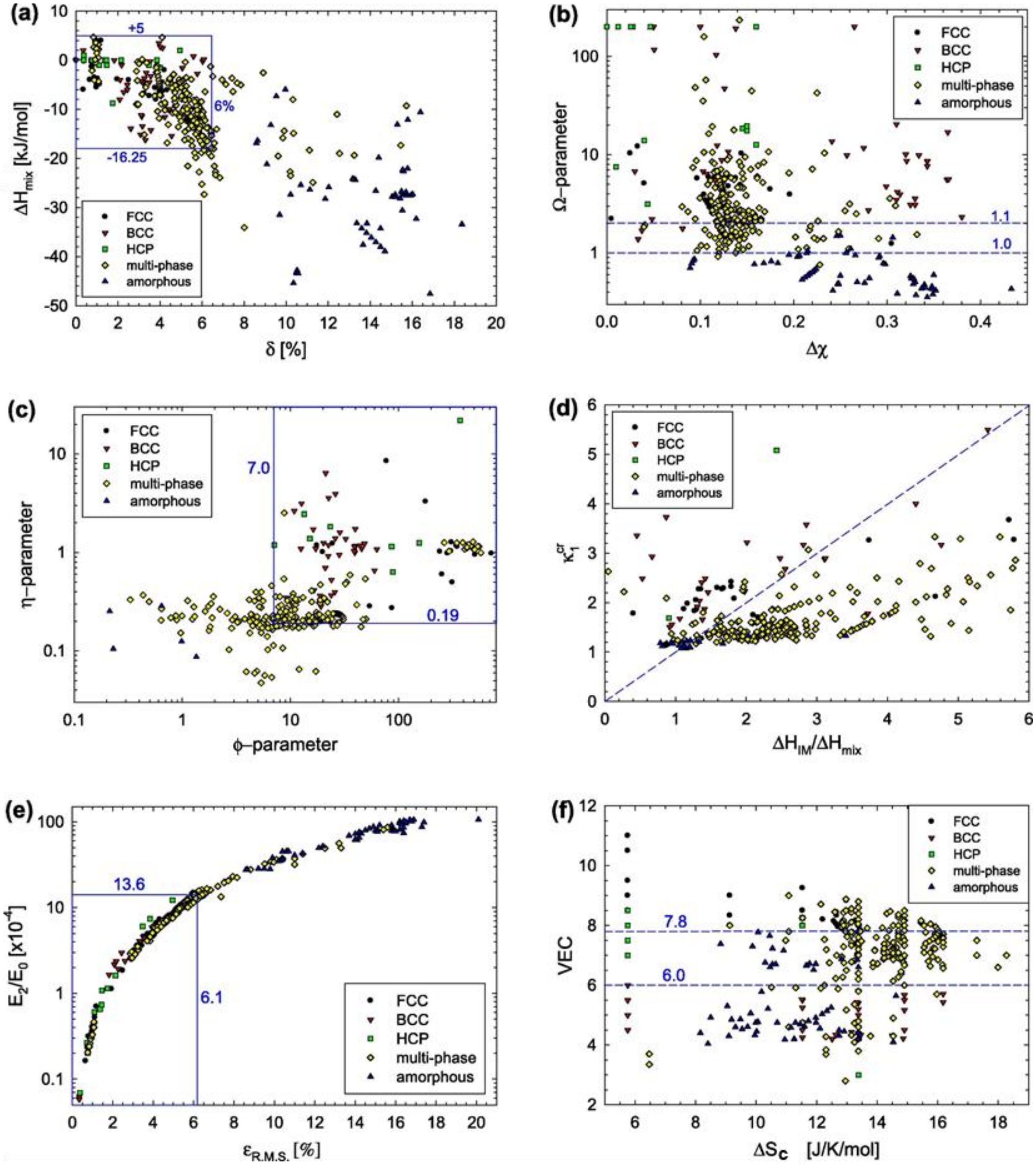


Figure 2.3: Plots comparing different empirical parameter effects in separating HEA phases. Figure from Gao et al [39].

2.3 Thermodynamic and first-principles calculations

Thermodynamic methods, such as the CALPAHD and Monte Carlo (MC) simulations, simulate the phase formation by minimizing the Gibbs free energy of mixing and adopting thermodynamic database obtained from experiments or other fundamental simulations. The more fundamental first-principles calculations methods, such as the ab-initio calculations and DFT, predict the phase formation by simulating the interactions between fundamental particles, such as the electrons and nuclei, using the well-established physical principles. This section highlights a few studies on HEA phase formation prediction using CALPHAD, and ab-initio calculations/DFT.

2.3.1 CALPHAD

CALPHAD [78–80] is a direct method for determining HEA phase formation. It is a powerful methodology that is employed to predict phase formation and thermodynamic properties such as the composition or temperature boundaries for phase transformations, and precipitation nucleation barrier [81]. Thermodynamic databases, which are the core of CALPHAD, are obtained either from experimental data or DFT results [78]. In order to have accurate HEA phase predictions, the database should ideally include thermodynamic data from lower-order binary or ternary systems [42] that can be extrapolated [82] to simulate higher-order multi-component HEA systems.

As a prediction method that gives detailed phase information, CALPHAD is used widely [23,42,43,74,78,83–91] in exploring vast compositional spaces. However, due to the vast compositional space in which HEAs lie, a limitation for the use of CALPHAD is the availability of sufficiently complete binary and ternary thermodynamic databases [42,92]. Current multi-component alloy databases are designed for traditional alloy systems based primarily on elements such as Al, Fe, Ni and Ti. Complete data for a multitude of ternary systems have yet to be acquired. Without this data, the predictions will not be fully accurate. Additionally, the veracity of the CALPHAD phase predictions drops when miscibility gaps or IM are present in the phase diagrams [81].

2.3.2 Ab-initio simulations and density functional theory

Ab-initio simulations [41] predict the thermodynamic and mechanical properties [93–103] of HEAs. These properties are determined from simulated electron density which was found using DFT. DFT provides an approximated solution to the Schrodinger equation of a simulated alloy system. An advantage of ab-initio calculations is that they rely on the fundamental quantum mechanical properties of the system, and no experimental nor empirical inputs are needed. However, ab-initio simulations can be computationally intensive methods. Furthermore, in practical use, they are usually used collaboratively with experimental results or other simulation methods like CALPHAD.

DFT calculations can provide binary phase formation information such as the formation energy [63] or bonding strengths [87–89]. Strong binary bonding indicates IM formation during solidification. A large positive binary formation energy indicates a potential phase separation. These empirical values can be used as a guide for finding systems with homogeneous binary pairs that tend to form SS. However, these methods usually do not distinguish between different SS phases such as the FCC, BCC, or HCP phases.

Ab-initio simulations can calculate ΔG_{mix} for forming a specific phase. These results have improved accuracy over results using only empirical parameters ΔH_{mix} and ΔS_{mix}. For a given composition and phase, the contributions to ΔG_{mix} from the electronic energy, magnetic free energy, atomic vibration free energy, and the configurational entropy can all be computed with certain approximations at various temperatures [104–106]. Ideally, the ΔG_{mix} selected from the phase formation determination process is the most negative of all possible phase configurations calculated. In practice, determining all the phase configurations is not possible or computationally exorbitant.

Alternatively, specific strategies have been used to expedite the simulation process [41]. The first approach is the combination of ab-initio and existing experimental results. For example, the experiment results from AlCoCrFeNi-type HEAs reveal that the FCC, BCC, and FCC+BCC phases are the phases that can form [107]. Based on these experimental results, Tian et al. [104] used ab-initio calculations to determine the ΔG_{mix} for forming FCC and BCC phases in AlCoCrFeNi-type HEAs. They then inferred the theoretical compositional space to form FCC, BCC, and mixed FCC+BCC phases. Another approach is enumerating the most probable phases. Wang et al. [108]

used this approach to study the phase formation of MoNbTaVW. They computed the ΔG_{mix} at given temperatures for 178 phases. The calculated ΔG_{mix} values were analyzed to determine the phase stabilities, phase separation tendencies, and order-disorder transitions. The third approach is studying binary phase diagrams [109,110]. For example, Rogal et al. [109] from inspection of constituent binary phase diagrams selected the $D0_{19}$ and HCP phases as candidate phases for the HEA $Al_{15}Hf_{25}Sc_{10}Tr_{25}Zr_{25}$. DFT calculations show that the HCP to $D0_{19}$ phase transition occurs at 1230 K, which is in agreement with the experimental result.

Computations of long-range order (LRO) and short-range order (SRO) in HEAs provide valuable information. Ab-initio calculations, coupled with MC or molecular dynamics (MD) simulations, are used for chemical order studies to determine order-disorder transition temperatures [97,99,108,109,111–122]. Santodonato et al. [44] used MC simulations with inputted DFT results [63] to determine, in the HEA $Al_xCoCrFeNi$ system with variations in Al content, the change in the phase transformation temperature for the precipitation of B2 phase from the BCC phase. Lederer et al. [123] developed a high-throughput ab-initio method to search for potential disordered SS HEAs from 1,240 candidates. The ab-initio results from AFLOW[124], a software framework for high-throughput calculations of crystal structure properties, are then incorporated with a generalized quasi-chemical approximation model[125], generating a temperature-dependent HEA order parameter. The order-disorder transition temperatures were estimated based on the change of this parameter. Furthermore, for a HEA the comparison between the order-disorder transition and melting temperatures was indicative of the predicted disordered SS formation tendency. The accuracy of the model was corroborated with MC simulations, experimental data, and CALPHAD, showing high agreement in both SS system predictions and transition temperature predictions.

2.4 Machine learning and statistical studies

Since the discovery of HEAs in 2004 [1,56], a large number of HEAs and their phases have been reported. The rapidly expanding database, in recent years, made it possible to bring ML into this field [45,46,50–54,65,98,126–135]. ML, in general, is capable of extracting non-linear correlation between input and output data. When applied to HEAs it can be utilized to discover patterns in the large amounts of existing HEA data. For ML, each HEA datum includes values of features and a class. The class is the HEA phase. The features are correlated with the phase formation and are utilized to make phase predictions. ML algorithms are methods capable of identifying patterns between the input features and the HEA phases. Based on these connections, phase predictions for new HEAs are given. In ML, choosing informative, discriminating, and independent features is crucial to the training process of an algorithm. The algorithms examine feature-phase relationships in a portion of the whole database called a training set. After that, ML makes and verifies the predictions for HEAs in the remaining database called a test set. A prediction success rate is generated from the ability of the training set to predict the test set correctly. The current HEA ML prediction methods can differ by three different aspects: (1) the features used, (2) the algorithms used to analyze the training database, or (3) the HEA phase classifications.

Tancret et al. [45] combined empirical parameters ($\Delta\chi$, VEC, K_m, ΔH_{mix}, δ, μ, e/a, Ω, and S_m), Gaussian processes (GP) ML algorithm, and CALPHAD to find a robust method of identifying single-phase HEAs in a database with 322 HEAs. The use solely of empirical parameters or CALPAHD is not reliable in phase prediction. Nevertheless, the combination of the two with GP can be useful. GP with empirical features first returns the probability for a HEA being a single-phase SS. When this probability for a HEA was higher than 0.59, CALPHAD would be applied to predict the phase. All HEAs were found to be single-phase SS when both predictions agree. However, many single-phase SS HEAs were misidentified as mixed phase HEAs by this method, this led to the absence of potential useful HEAs from the predictions.

GP was also used by Pei et al. [65] in classifying alloys as a multi-phase or a single-phase. Single-phase alloys were further classified as a BCC, FCC, or HCP phase. The database included 1,252 alloys ranging from binary alloys to multi-component HEAs. The database was partitioned into 627 multi-phase alloys and 625 in the single-phase alloys. The atomic percentage weighted averages of 85 elemental properties composed the features pool. Initially, ten features were selected based on their relevance for making GP phase classification decisions. Then different combinations of features were tested until only the optimum features remained. This method returned a prediction accuracy of 93 %. Molar volume, bulk modulus, electronegativity, melting temperature, valence, vaporization heat, and thermal conductivity were determined to be the most relevant features. GP returned a probability for each alloy indicating its tendency to form a single phase. This GP probability was plotted against δ in Figure 2.4, where GP probability > 0.5 and $\delta < 6$ % were the criteria for forming a single phase.

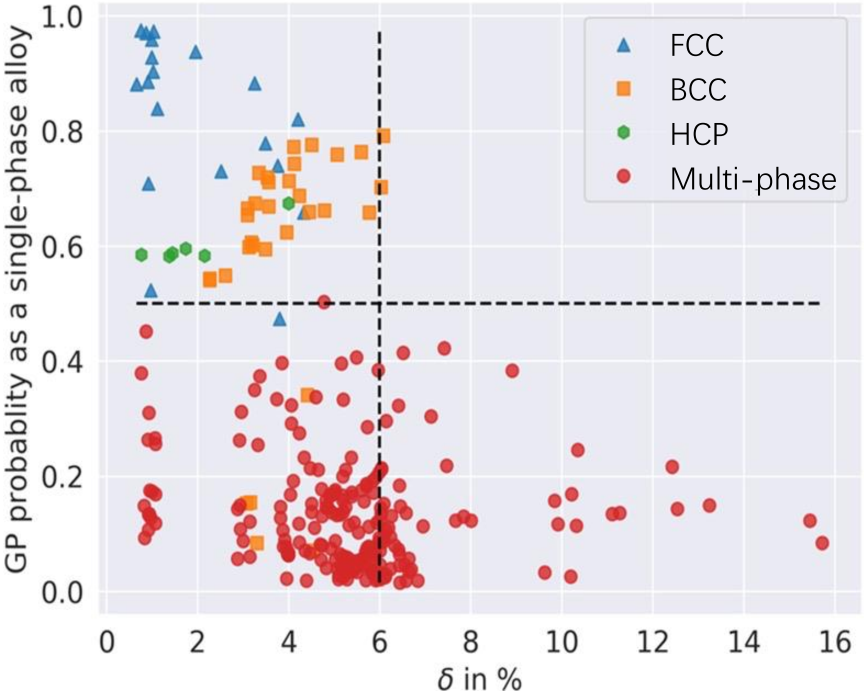


Figure 2.4: GP probability as a single-phase alloy plotted against radius mismatch, δ to separate alloys into different phases. The criteria for forming a single-phase alloy are represented with dashed lines. Figure from Pei et al [65].

Islam et al. [50] used empirical parameters as the ML features as well, but with an ANN ML algorithm. The 118 HEAs used were classified as 64 SS, 21 IM, and 33 amorphous (AM) phases. ANN predictions showed that the relevance with the phase formation between different features increases in the following order: ΔS_{mix}, δ, ΔH_{mix}, Δχ, and VEC. The ANN prediction accuracy was 83 %. Like prior mentioned methods, detailed phase content is still not predictable with this method.

Huang et al. [51] used five empirical parameters VEC, Δχ, ΔH_{mix}, ΔS_{mix}, and δ as ML features. The 401 HEAs used were classified as 174 SS, 173 SS+IM, and 54 IM HEAs. Three kinds of ML algorithms were used; (1) k-nearest neighbors (KNN) returned a prediction accuracy no larger than 68.6 %, (2) SVM returned an accuracy of 64.3 %, and (3) supervised multi-layer feed-forward neural network (MLFFNN) returned an accuracy of 74.3 %. Binary classifications of phases between SS and IM, SS and SS+IM, as well as IM and SS+IM, were also conducted with MLFFNN, returning accuracies of 86.7 %, 78.9 %, and 94.3 %, respectively. Of the empirical parameters used, δ and VEC were of greater importance than the others. According to the authors, additional features will improve the accuracy.

Similarly, Li and Guo [52] used SVM with empirical parameters to classify a database containing 18 BCC, 43 FCC, and 261 other phases called NSP (not forming single SS) HEAs. Different combinations of ML features among candidates $\{\mathrm{VEC}, \delta, \Delta\mathrm{H}_{\mathrm{mix}}, \Delta\mathrm{S}_{\mathrm{mix}}, \Delta\chi, \Delta\mathrm{H}_{f}, \mathrm{T}_{\mathrm{m}}\}$ were used. The feature combination of $\{\mathrm{VEC}, \delta, \Delta\mathrm{H}_{\mathrm{mix}}, \Delta S_{\mathrm{mix}}, \mathrm{T}_{\mathrm{m}}\}$ was found to give the best prediction accuracy of 90.69 %. The test accuracy increases by > 5 % when the training data set percentage increased from 50 % to 90 %. This indicates that the performance of the model can be further increased by including additional future experimental data. When the training set percentage was 90 %, the accuracies for BCC, FCC, and NSP phase predictions were 60 %, 75 %, and 97.79 %, respectively. The error in the performance occurred due to missed predictions between FCC and NSP, or BCC and NSP phases. The method excelled at separating BCC and FCC HEAs.

Agarwal and Prasada Rao [54] used an adaptive neuro-fuzzy interface system (ANFIS), a hybrid method using an ANN ML algorithm and fuzzy logic, to predict HEAs with BCC, FCC, and multi-phases. Two sets of input features were used. The first used compositions of HEAs and returned a prediction accuracy of 84.21 %. The second used empirical parameters VEC, δ, $\Delta\mathrm{H}_{\mathrm{mix}}$, $\Delta\mathrm{S}_{\mathrm{mix}}$, ϕ, and $\sqrt{<\varepsilon^2>}$ and returned a prediction accuracy of 80 %. In the second model, the importance of each feature was ranked by removing one individual empirical feature and calculating the prediction accuracy drop due to the absence. The ranking of importance of the empirical features was determined to be $\sqrt{<\varepsilon^2>} > \mathrm{VEC} > \delta > \phi > \Delta\mathrm{H}_{\mathrm{mix}} = \Delta S_{\mathrm{mix}}$. By systemically changing the value of one feature while keeping the other features unaltered, the phase prediction results from ANFIS could reflect how the change of each feature affects the phase formation. For example, BCC phase formation is favored over FCC phase formation when δ increases.

Zhou et al. [53] used three algorithms to study the phase formation rules. They were ANN, one-dimensional convolutional neural network (CNN), and SVM. The database used was composed of 13 empirical features and 601 as-cast binary, ternary, quaternary, and higher-order alloys. The ML model studied the appearance of SS, IM, and AM phases in HEAs. Multiple positive phase predictions would indicate a combination of those predicted phases. The testing accuracies of three algorithms on predicting the appearances of the SS, IM, and AM phases were all near or above 95 %. Correlations between features and the appearance of a phase were examined by a compound transformation function. It was derived from linear transformation matrices and biases among the input, hidden, and output layers in the ANN model. As can be seen in Figure 2.5, certain features were found to be strongly correlated to specific phase appearances. For example, large values for T_m and the standard deviation of binary ΔH_{mix} ($\sigma_{\Delta H}$) promote the formation of the AM phase while suppressing the formation of the IM. And while ΔS_{mix} promotes, δ suppresses the SS formation. Experimental results using $(FeCrNi)_{10-x}(ZrCu)_x$ further validated this model, however, the results are cooling rate sensitive.

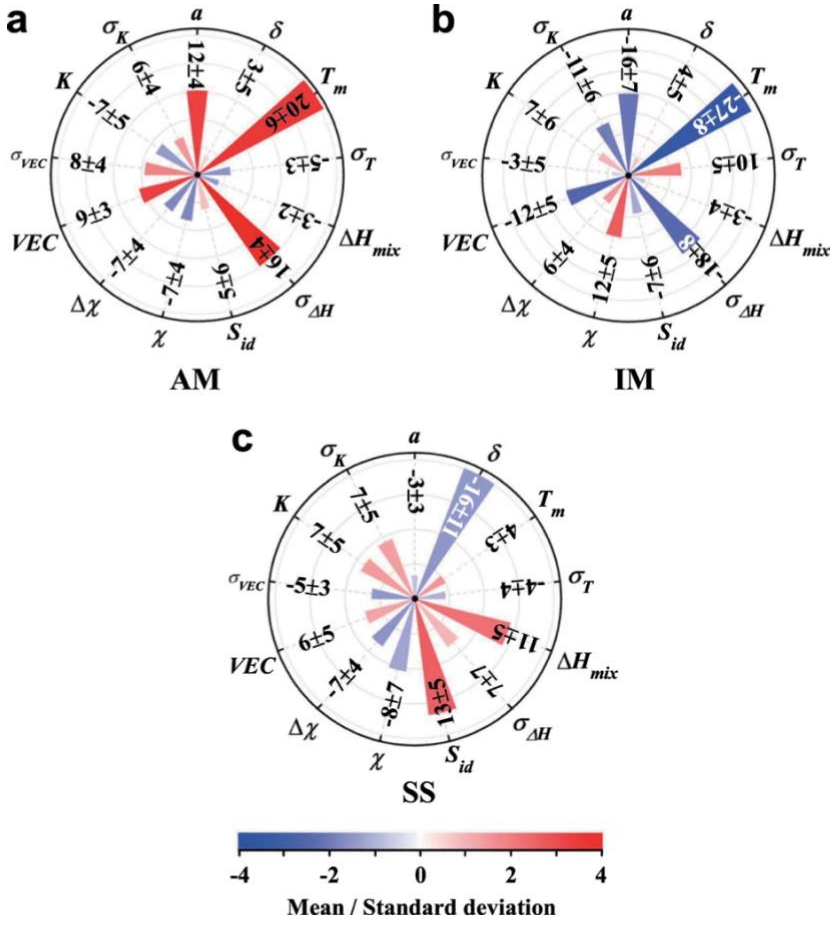


Figure 2.5: Correlations between features and the appearance of (a) AM, (b) IM, and (c) SS phases. Red and blue colors represent positive and negative correlations, respectively. Figure from Zhou et al [53].

Recently, there has been an increase in studies utilizing feature engineering and active learning [98,99,129,136,137]. Feature engineering is an approach to mathematically expanding the ML feature pool. A large number of features are synthesized from a limited basis set of features. Then a down-selected combination of features is chosen, which returns the highest accuracy. Active learning is an approach that can experimentally expand the database under the guidance of ML and improve prediction accuracy. They have been applied in the following works.

Zhang et al. [129] utilized the genetic algorithm (GA) method to select the best combinations of ML features and models systemically. In their classification I, a 550 HEA database was classified into SS and non-solid solution (NSS) phases. In their classification II, the SS HEAs were further divided into FCC, BCC, and dual phases (DP), a combination of FCC and BCC phases. A flowchart for this work is shown in Figure 2.6. The pool of features and models was composed of 70 features and nine common ML classification algorithms. The results from their computations showed that only a minimum of four features were required to produce an accurate prediction. The Pearson correlation coefficients [138], which measures the statistical relationships between two continuous variables, were calculated among the 70 features to remove the redundant features. For each ML algorithm, the GA systemically changed the four features used and determined the feature combination returning the best prediction accuracy. All nine algorithms were optimized in the same manner, and the best model was selected. Classification I and II eventually had accuracies of 88.7 % and 91.3 %, respectively. Active learning was further employed to refine their predictions. Ten new HEAs, whose predicted phases had high uncertainties from ML, were experimentally prepared and measured for their phases. After adding the new data into the database, the ML prediction accuracy increased. This implies that iterating the active learning steps can improve the ML accuracy.

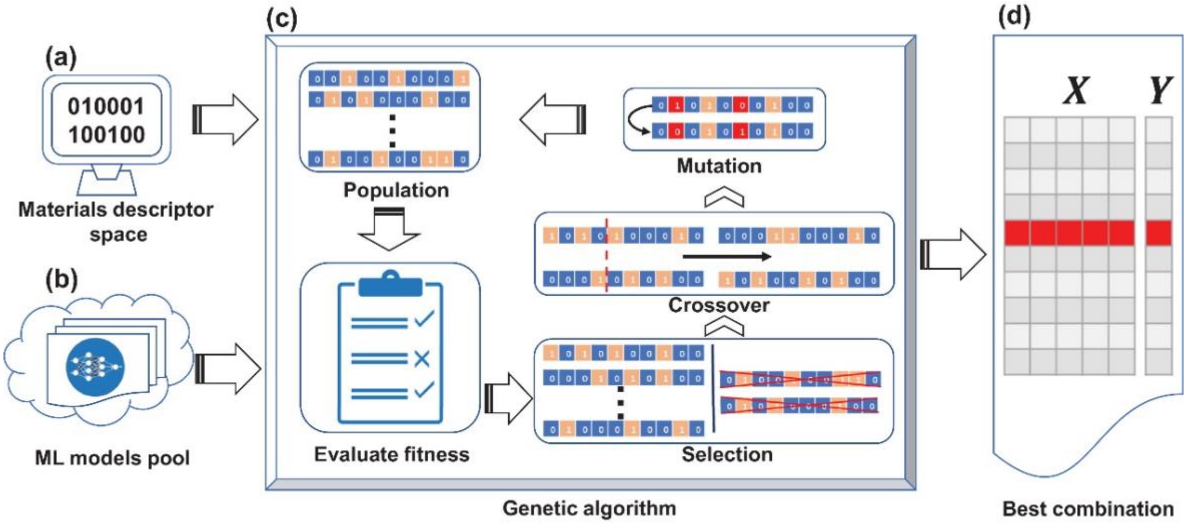


Figure 2.6: Flowchart from Zhang et al. [129] describing the GA method to best select ML algorithms and features for HEA selection. First, (a) is the ML feature pool with 70 features that can influence the phase formation. Second, (b) is the commonly used ML algorithms pool. Third, (c) is the GA process, where the optima combinations of features and ML algorithm is obtained. And lastly, (d) is the optimized output.

Dai et al. [130] explored feature engineering with empirical features to improve the prediction ability. First, specific features were highly correlated with each other based on their Pearson correlation coefficients. Feature pairs such as $\sqrt{<\varepsilon^2>}$ and δ had large Pearson correlation coefficients and one redundant feature should be removed. With this method, 14 candidate features were down-selected to nine features. Second, from the nine features, a pool of 30,450 non-linear features was generated. These non-linear features were calculated by the following relationships: $\sqrt{|x|}$, x^2, x^3, and log(1+|x|) for feature x, or multiplying two or three of these features together. A recursive feature elimination method was used to eliminate the irrelevant features, and 20 features were left. To compare the two feature pools, a database with 407 HEAs and a simple linear regression ML algorithm were employed to classify the HEAs into BCC, FCC, HCP, multi-phase, and AM phase categories. The highest ML prediction accuracies obtained from using the original nine features and the 20 engineered features were 75 % and 86 %. The constructed non-linear features outperformed the original features. This work shows that feature engineering can improve the feature-phase relationship to increase the accuracy of predictions.

Recently, Lee et al[139] applied ML techniques, ensemble learning, post hoc model interpretability of black-box models, and clustering analysis to interpret the HEA phase formation. Lee et al. identified the most important features and tuned the parameter values to show how the phase formation is influenced. The work is valuable in that it reveals how different features can influence phase formation.

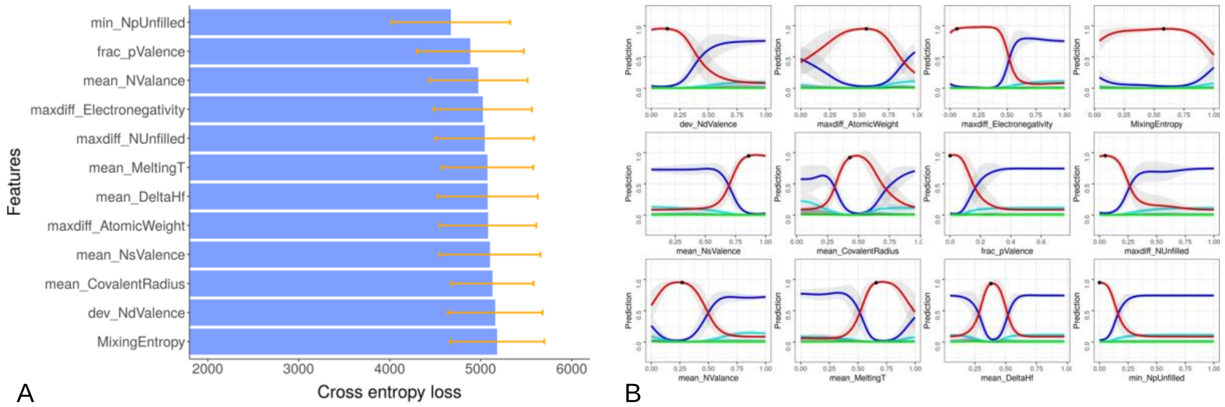


Figure 2.7: (A) Cross entropy loss corresponding to each ML feature on the vertical axis. Large cross entropy loss represents higher importance in ML. (B) Ceteris Paribus profile for NbTaTiV. Plots demonstrate how feature values influence the phase prediction predicted by ML. Horizontal axes are the features' value. Vertical axes are the ML predicted phase content of NbTaTiV. Line colors represent different phases: blue, Multiphase; violet, Amorphous; cyan, FCC; orange, BCC+FCC; light blue, HCP; red, BCC; green, IM. Figures from Lee et al[139].

The works mentioned above show that the empirical parameters are frequently selected as features for ML methods. The following ML methods limit features to only the HEA compositions or their elemental components.

Wu et al. [131] designed a eutectic HEA system, AlCoCrFeNi, using ANN. Their database contained 311 eutectic HEAs predicted by CALPHAD and 10 HEAs discovered by experiments. Their data was divided into training and test sets of 75 % and 25 %, respectively. The input features consisted of the atomic percentages of the five elements in each HEA. The output was a normalized number between -1 and +1: a negative value represented the formation of a hypereutectic B2 primary HEA, 0 was the formation of an eutectic HEA, and a positive value represented the formation of a hypoeutectic FCC HEA. As shown in Figure 2.8a, the predicted values and the target values are in agreement. The ML showed high phase constitution prediction ability, and 400 new near-eutectic HEAs were predicted. Figure 2.8(b-c) are the element content distributions of different elements for the new near-eutectic HEAs. In Figure 2.8b, the Al and Cr content distributions are clustered in certain ranges, while the Co, Ni, and Fe content distributions are near evenly distributed. Figure 2.8c shows correlations between different element content distributions. A strong content correlation for Al-Cr is noted. Additional findings also seen from Figure 2.8: Al was identified as the most relevant element to determine the phase constitution; Cr is associated with Al content to influence the eutectic formation; Co, Ni, and Fe are miscible elements influencing the eutectic formation by tuning their average VEC value; and a high VEC favors a FCC phase formation while a low VEC favors a B2 phase formation. From their work, an effective eutectic HEA design pathway was presented, and several HEAs exhibiting strengths of ~1300 MPa and elongation of ~20 % were made.

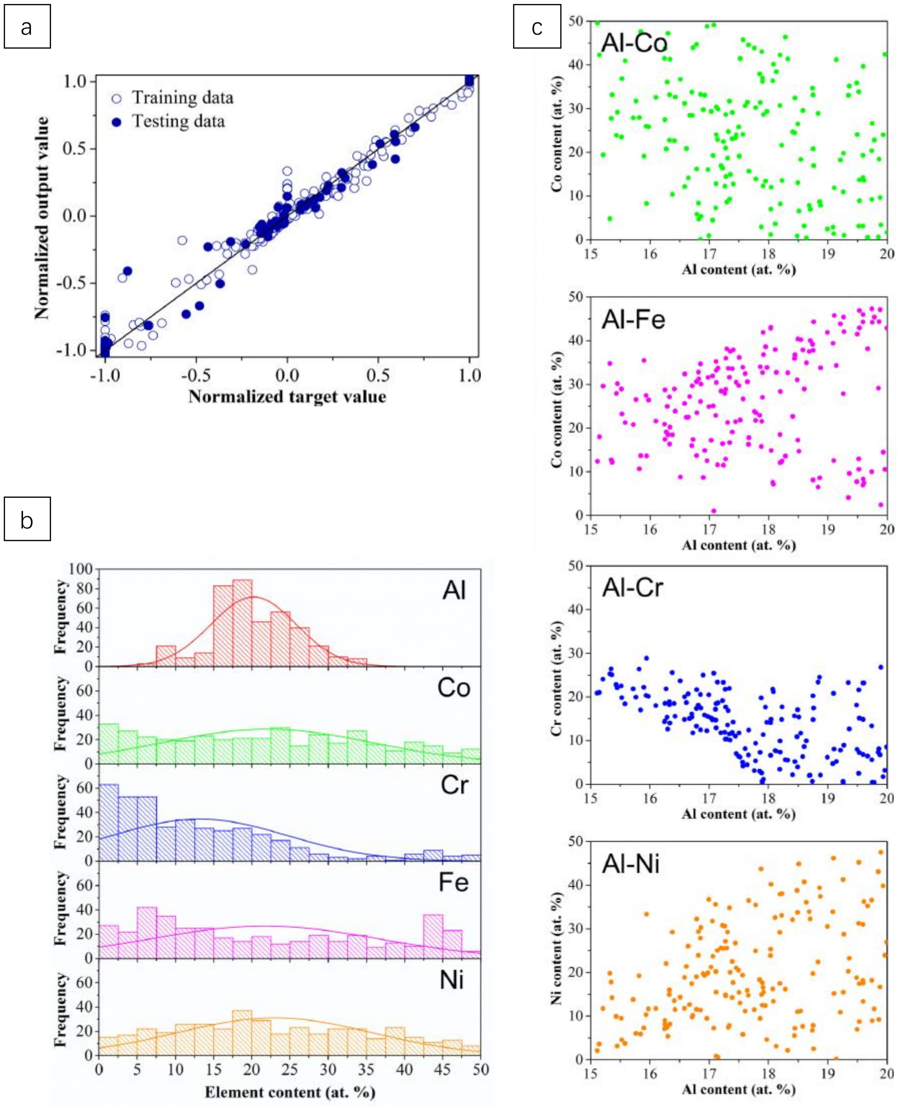


Figure 2.8: (a) Regression analysis of the training and testing data. The predicted values and the target (actual phase constitution) values are the y and x-axes. (b)-(c) Element content distributions of different elements for the 400 predicted near-eutectic HEAs. Figures from Wu et al [131].

Kube et al. [46] used a linear ordinal logistic regression method to predict the HEA phases based on their compositions. After optimization, the values were assigned to elements representing their strength in stabilizing BCC or FCC phases. The average of these stabilizing effects, denoted as solid solution selection index (SSSI), determined HEA phase formation tendencies. As shown in Figure 2.9, this method separates the BCC and FCC HEAs, but not the mixed BCC+FCC HEAs. The results showed that certain elements have an influence on stabilizing specific phases. There are three particular limitations to this method. First, the HEA database used for training sets was produced from high-throughput sputter depositions, which as a method can extend the compositional range of SS formation due to the rapid quench nature of sputtering as opposed to phase formation from traditional alloying methods. Second, the elements involved in this study were confined to Al, Cr, Mn, Fe, Co, Ni, and Cu with the effect of other elements on phase formation needing further study. And final limitation is that the phases predicted by this method are only BCC, FCC, and their mixtures while other phases are not accounted for.

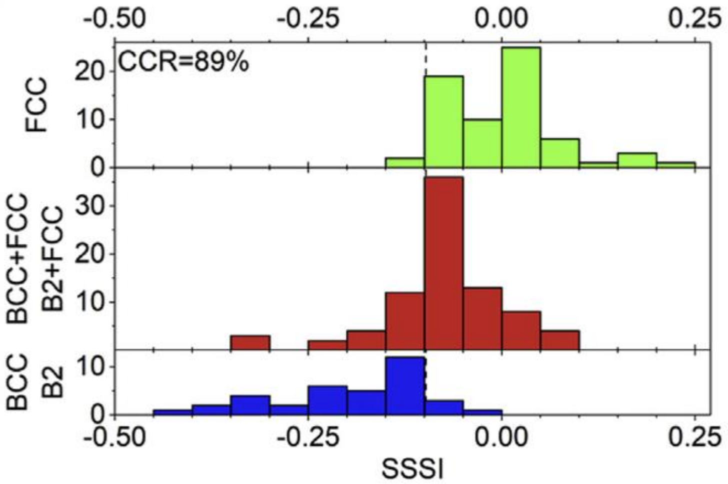


Figure 2.9: Prediction performance of Kube's method. SSSI is the parameter representing a phase stabilizing effect. Negative SSSI favors BCC/B2 phase formation, while positive SSSI favors a FCC phase formation. Blue (bottom), red (middle), and green (top) histogram bar plots correspond to BCC/B2, BCC+FCC/B2+FCC, and FCC phases, respectively. BCC/B2 and FCC HEAs are mostly separated by their SSSI values, while the mixed BCC/FCC phases still mix with other phases. The overall correct classification rate (CCR) is 89 %. Figure from Kube et al [46].

The prior mentioned statistical and ML methods are summarized in Table 2.1. For ML, the empirical parameters are the most frequently used features. ML combinations of these features yield an improved robust prediction ability as opposed to any of the features used individually.

Figure 2.10 provides a summary on: (A) the number of ML HEA phase prediction publications each year; (B) the phase classifications used in these works. Two issues exist in the current ML phase classification models, namely, a low number of phase categories, and in some cases, a low level of detail within a classified category; that is, the categories are general instead of specific. As discussed in more detail below, many models only classify HEA phases into no more than three categories, because as additional categories are included, there is an increase in the complexity and the challenge of attaining high accuracy increases. Only FCC, BCC, FCC+BCC, and HCP categories in these models (Figure 2.10B) represent specific phases. More general categories, such as SS, AM, IM, and single/multi-phase, correspond to unspecified structural phase groups. The low level of categorization detail gives limited guidance for HEA design, i.e., when a HEA is categorized as IM, it can be either B2 (ordered BCC), Laves, Sigma, or Mu phase. This limit may result from the limited database availability of less common phases.

The two challenges mentioned above still remain unsolved in the literatures: (1) Can we achieve a more specific/detailed IM classification? (2) Can we predict more phase categories simultaneously and accurately to guide HEA design? They will be addressed in Chapter 3.

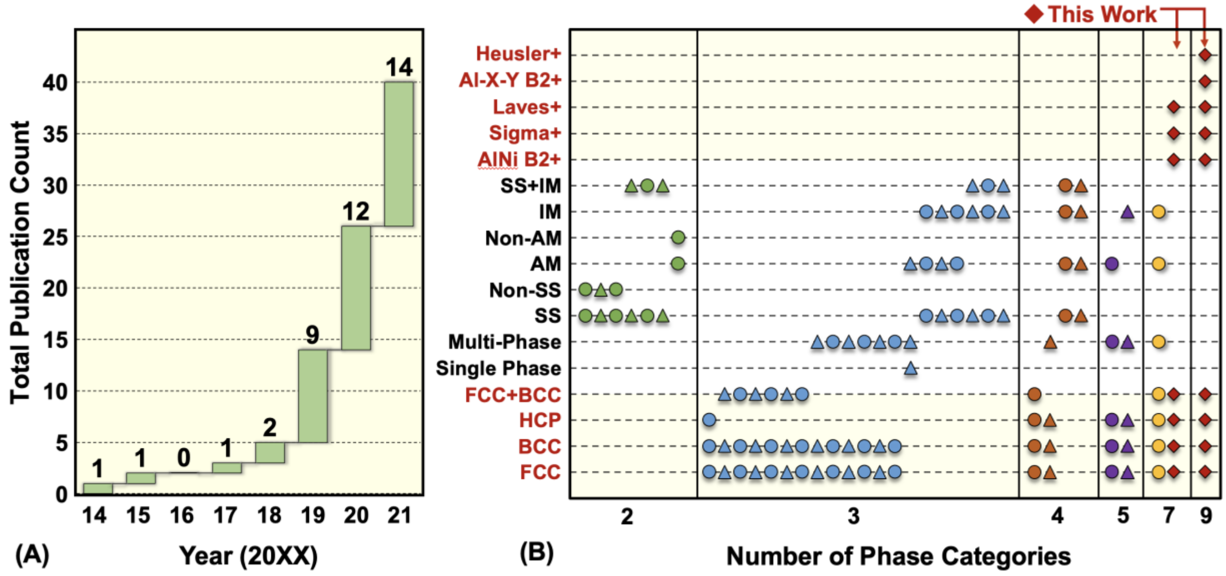


Figure 2.10: (A). The number of publications on ML-based HEA phase prediction versus year. Individual years' publication subtotals are listed on the plot. (B). Summary of how current ML-based HEA phase prediction models classify phases. Symbols aligned vertically represent phase categories in one model. There are 36 models in the plot. Models that predict specific numbers of phase categories are grouped in columns. Phase categories are on the vertical axis. Categories with yellow background and red text represent specific phase structures; Categories with white background and black text represent different groups of unspecified structural phases. The last two columns of red diamonds represent the models in this work, with higher level-of-detail phase categories. The column of seven red diamonds represents the multi-phase prediction model that classifies seven phases. The classification method is further expanded to nine categories by four high-accuracy sub-models, represented by the column of nine red diamonds. The phase categories are labeled by abbreviations: FCC+BCC: Mixed FCC-BCC phase; SS: Solid solution phase; AM: Amorphous; IM: Intermetallic phase; + sign represents that other phases may form simultaneously with the specific IM. B2 phase generally has two types in HEA: AlNi[9] and Al-X-Y[15]. The two columns labeled with "This Work" will be presented in Section 3.3.

Table 2.1: Summary of the current statistical and ML methods for predicting HEA phases. Abbreviations of the algorithms are: ANFIS (Adaptive Neuro-Fuzzy Interface System), ANN (Artificial Neural Network), CALPHAD (Calculation of Phase Diagrams), CNN (Convolutional Neural Network), GA (Genetic Algorithm), GP (Gaussian Processes), KNN (K-nearest Neighbors), MLFFNN (Supervised Multi-layer Feed-forward Neural Network), and SVM (Support Vector Machine). The ML classification algorithms, the ML features, the phase categories defined in each ML method, the prediction success rates, and the references to the work are listed. In the phase

categories column, the total count of HEAs in each phase category is listed, if the information was available.

Classification Algorithms	Features	Phase Categories (count reported)	Overall Success Rate	Ref.
ANFIS	HEA Compositions	HEA: BCC, FCC, and Multi-phase	84.21 %	Agarwal and Prasada Rao[54]
ANFIS	VEC, δ, ΔH_{mix}, ΔS_{mix}, ϕ, and $\sqrt{< \varepsilon^2 >}$		80 %	
ANN	HEA Compositions	HEA (321): Hypereutectic, Eutectic, and Hypoeutectic	N/A	Wu et al.[131]
ANN	ΔS_{mix}, δ, ΔH_{mix}, $\Delta\chi$, and VEC	HEA: SS (64), IM (21), and AM (33)	83 %	Islam et al.[50]
ANN, CNN, and SVM	13 Empirical Parameters	HEA (601): SS, IM, and AM	> 95 %	Zhou et al.[53]
Feature Engineering + Simple Linear Regression	20 Features engineered from 14 Empirical Parameters	HEA: BCC (43), FCC (48), HCP (16), Multi-phase (237), and AM (63)	86 %	Dai et al.[130]
GA + Active Learning	4 Features downselected from 70 Features by Feature Engineering	HEA (550): SS, and NSS	88.7 %	Zhang et al.[129]
		SS HEAs: FCC, BCC, and DP	91.3 %	
GP + CALPHAD	$\Delta\chi$, VEC, K_m, ΔH_{mix}, δ, μ, e/a, Ω, and S_m	HEA (322): Single SS, and Other phases	63 % to 80 % (single SS, CALPHAD database dependent)	Tancret et al.[45]
GP	Atomic Percentage Weighted Averages of 85 Elemental Properties	HEA & Non-HEA (1,252): BCC, FCC, HCP, and Multi-phase	93 %	Pei et al.[65]
KNN, MLFFNN, SVM	ΔS_{mix}, δ, ΔH_{mix}, $\Delta\chi$, and VEC	HEA: SS (174), SS+IM (173), and IM (54)	74.3 % (MLFFNN), 68.6 % (KNN), 64.3 % (SVM)	Huang et al.[51]
SVM	ΔS_{mix}, δ, ΔH_{mix}, T_m, and VEC	HEA: BCC (18), FCC (43), and non-single-phase (261)	60 % (BCC), 75 % (FCC), 97.79 % (NSP)	Li and Guo [52]
Linear Ordinal Logistic Regression	HEA Compositions (generated from sputtering deposition)	HEA: BCC or B2 (762), FCC (553), and Mixed BCC or B2 + FCC (446)	89 %	Kube et al.[46]

2.5 Methods comparison and synergetic use

Empirical parameters are ineffective in predicting HEA phases as shown in Section 2.2.4. The thermodynamic methods normally provide accurate and detailed phase formation prediction, such as the phase content, volume fractions, transformation temperatures, and element distribution among phases. However, the time and computational power required by these methods are enormous. This makes high-throughput screening for promising HEAs time-consuming. On the other hand, ML is not computationally intensive and gives accurate phase prediction results. ML is the ideal method for high-throughput composition screening.

At the current stage, the synergetic use of these methods should be considered. For example, ML models could conduct high-throughput screening in the vast compositional space and search for compositions with desired phases and properties; first-principles and thermodynamic methods then conduct fine adjustments on these compositions for further optimizations.

Chapter 3 Machine Learning Phase Prediction with Engineered Phase Diagrams-based, Thermodynamic, and Hume-Rothery Rule Features

Recall that machine learning (ML) is a process that applies algorithms to quantify the correlation between features and results (HEA phases, in this case). The chapter will focus on improving two crucial parts of ML: features and algorithms.

Firstly, we innovatively developed a set of phenomenological ML features based on binary alloy phase diagrams (Section 0). Then, we further improved the prediction capability using the feature engineering technique (Section 3.3). Furthermore, we interpreted the ML results and identified the most important scientific parameters controlling each phase's formation (Section 3.4). Finally, all the ML prediction models are experimentally validated by newly synthesized HEAs (Section 3.5). This model will be the basis of further HEA design work.

3.1 Database introduction

Currently, over 1,100 reported HEAs are in the updated version of the prior reported comprehensive database [55]. Only a subset of 828 of the HEAs was used. They fulfilled two requirements: (1) they were in either as-cast or high-temperature annealed states, and (2) they include the common HEA phases with broad homogeneity regions, e.g., FCC, BCC, HCP, B2, Heusler ($L2_1$), Sigma, or Laves phases. HEAs were classified based on their phase content into nine categories (category label and data counts in parentheses):

1. Single disordered FCC (A1, 132),
2. Single disordered BCC (A2, 178),
3. Single disordered HCP (A3, 14),
4. Mixed disordered FCC+BCC, including the commingling of A1s, A2s, or the coexistence of A1s and A2s (A1+A2, 72),
5. AlNi type B2 mixed with other disordered SS (AlNi type B2+, 291),

6. Al-X-Y type B2 mixed with other phases (Al-X-Y type B2+, 52),
7. Heusler mixed with other phases (Heusler+, 77),
8. Sigma mixed with other phases (Sigma+, 52),
9. Laves mixed with other phases (Laves+, 96).

Two things needed to be addressed. Firstly, the database is based on experiments where certain errors must exist. The actual alloy compositions can be different from the nominal compositions due to elements vaporization and contamination during alloy preparation. The actual compositions are normally not disclosed in the literature, so the nominal values are commonly used in ML feature computation. Besides, phase characterization methods and instruments differ among the research groups, inducing error again. X-Ray Diffraction (XRD) is commonly used for phase characterization. However, the background noise may mask the diffraction peaks from the minor phases, limiting their identification. Although transmission electron microscopy (TEM) can provide more reliable phase detection, it is not frequently used due to the significant experimental effort. These above-mentioned factors may lead to conflicting experimental results for the same alloy from different groups. In that case, we will keep the results from more reliable experiment works (e.g., using TEM results instead of XRD results if both are available). Secondly, the HEA database is constantly expanding. Generally, including more training data can increase ML prediction capability. However, considering the already high ML accuracies shown in this chapter, any accuracy improvement from further expanding the training dataset will likely be marginal. We believe the current database size is enough for predicting the nine categories. However, it is still important to include more data for a certain type of alloys of interest and increase the prediction capability.

3.2 Phase diagram-based parameters

3.2.1 Overview

As shown in Table 2.1, HEA compositions and empirical parameters are commonly used as ML features. Departing from these features, we developed phenomenological features using binary phase diagrams to predict the HEA phases. The advantage of using binary phase diagrams to assess phase stability is that they can readily provide direct and realistic information about the roles of individual elemental components on phase formation. Experimentally determined binary phase diagrams contain information on the crystal structure, elemental mixing, and phase separation over temperature and composition ranges. They have encoded within them the information for equilibrium binary phase formation. Information for the prediction of the phase of a HEA can be extrapolated from the set of all possible constituent binary phase diagrams.

Here, the effectiveness of the proposed method will be demonstrated by introducing physically meaningful phenomenological parameters that can be conveniently accessed from binary phase diagrams. These parameters are used to demarcate the phases forming regions for HEAs. The seven phase categories studied here are those with homogeneity ranges in the phase diagrams: FCC (A1), BCC (A2), HCP (A3), mixed FCC-BCC (A1+A2), AlNi type B2+, Laves+, and Sigma+. Al-X-Y type B2+ and Heusler+ HEAs are not studied in this section because they normally do not appear on binary phase diagrams. They will be addressed in the next section, Section 3.3.

3.2.2 Melting temperature and phase formation temperature

Alloy homogenized phases are known to be different under different temperatures. Alloy starts solidification when cooling down from the melting temperature T_m, and the phase will keep evolving at high temperature due to the large atomic kinetic energies. At certain temperature, the rapid phase evolution ceases, and the phases formed are retained at lower temperatures. Such temperature is defined as the phase formation temperature (T_{pf}).

For the annealed HEAs, the phases formed during annealing are locked in with rapid quenching, and T_{pf} is assigned the final annealing temperature. The phenomenological parameters that controlled the phase formation were calculated based on the T_{pf} value. For the as-cast HEAs, the undercooling temperature extends to 0.8 T_m. T_{pf} was approximated as T_{pf} = 0.8 T_m, which optimized the results for the ML predictions [55] using these parameters.

The T_m is calculated from the liquidus temperatures in binary phase diagrams. c_i and c_j are the atomic percentages of the elements i and j. For the binary pair i-j, binary liquidus temperatures T_{i-j} can be found at the composition where i and j element relative ratio is $c_i : c_j$. T_m of the whole system will be calculated by:

$$T_m = \frac{\sum_{i\neq j} T_{i-j} \times c_i \times c_j}{\sum_{i\neq j} c_i \times c_j} \qquad 3.1$$

An example of calculating T_{i-j} is given in Figure 3.1. $T_{Cr\text{-}Ni}$ is found to be 1700K on Cr-Ni binary phase diagram when Cr % : Ni % = 6 : 4. The technique of image processing is to scan all pixels on the diagram along the binary compositional line from the highest temperature. The first pixel with a color is the liquids point.

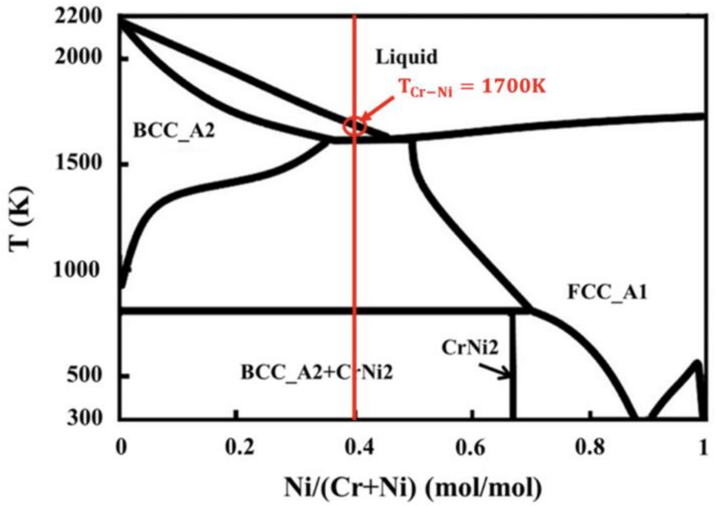


Figure 3.1: Example of finding binary liquidus temperature ($T_{Cr\text{-}Ni}$ = 1700K) on Cr-Ni binary phase diagram at Cr % : Ni % = 6 : 4.

3.2.3 Phase field parameters and phase separation parameter

When the temperature is above or equal to T_{pf}, atoms are free to exchange neighbors due to high atomic mobility. The neighbors of each atom are random. The alloy mixture is essentially ergodic and local atoms have nearly equal probabilities of sampling any binary configurations on the relevant phase diagrams. As such, the probability of forming a phase X locally for i-j elements can be determined by the binary phase field percentage of phase X on an i-j phase diagram and is denoted as X_{i-j}.

The local probabilities of forming a specific phase from all atomic pairs can be integrated to yield an overall probability. The probability of forming a phase X for the HEA is the Phase Field Parameter (PFP_X), and it is calculated as the weighted average of all constituent X_{i-j} by Eqn. 3.2.

$$\mathrm{PFP_X} = \frac{\sum_{i \neq j} X_{i-j} \times c_i \times c_j}{\sum_{i \neq j} c_i \times c_j} \div 100\ \% \qquad 3.2$$

In this method, the $\mathrm{PFP_X}$ values have been calculated for the targeted phases, and they are denoted as $\mathrm{PFP_{A1}}$, $\mathrm{PFP_{A2}}$, $\mathrm{PFP_{B2}}$, $\mathrm{PFP_{A3}}$, $\mathrm{PFP_{Laves}}$, and $\mathrm{PFP_{Sigma}}$.

HEA $Al_2CoCrCuNi$ is presented as an example to determine the phase field percentages used to calculate PFPX. This HEA has a predicted T_m = 1569 K and the phases are assumed to be locked at T_{pf} = 1255 K. In Figure 3.2, it is seen that high concentrations of Cr favor BCC phase formation, while high concentrations of Ni favor FCC phase formation. Under the assumption of equally sampling all binary configurations, the probability of Cr-Ni binary favoring BCC phase formation locally is the binary phase field percentage of the BCC phase. This percentage is the line segment between the two intersection points of an isotherm at T_{pf} and the compositional boundary of the BCC phase. In this case, it is approximately 5 % for the BCC phase and approximately 44 % for the FCC phase.

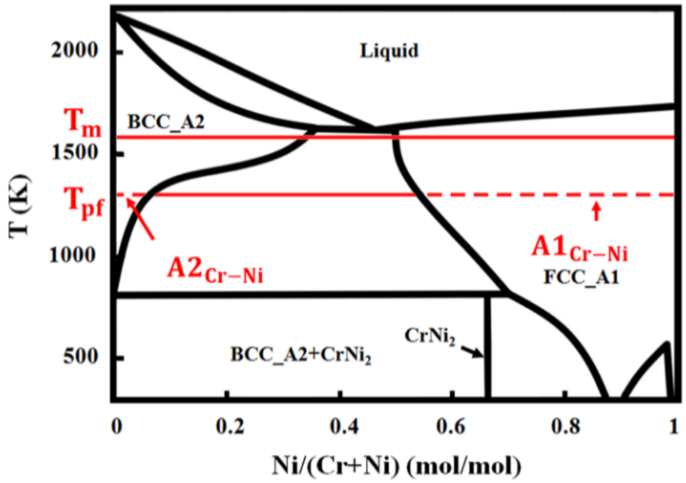


Figure 3.2: Demonstration of the binary phase field percentage calculation. The binary phase diagram Cr-Ni is used to determine the fractions of BCC and FCC phases for the HEA $Al_2CoCrCuNi$. The binary phase field percentages of BCC and FCC phases are represented as $A2_{Cr-Ni}$ and $A1_{Cr-Ni}$, respectively. Figure from Qi et al [55].

These binary phase field percentages are then used to calculate $\mathrm{PFP_X}$, which will be used to visualize separations in the HEA phase space. The above calculation method can be applied to any phase diagram and for any type of phase.

If a miscibility gap exists in phase diagrams, this interatomic repulsion can lead to phase separation in HEAs [140,141] and the formation of multiple coexisting phases such as FCC+BCC. The binary phase separation percentage on the binary phase diagram represents the probability of the two elements being separated into two different phases in the HEA. For a given phase diagram, an isothermal line drawn at T_{pf} is composed of two parts. The first is the binary phase separation percentage denoted as $\mathrm{Separation}_{i-j}$ and the remainder of the line is defined as the elemental mixing denoted as Mixing_{i-j} for an i-j binary system. If the phase separation is absent from a phase diagram, then $\mathrm{Separation}_{i-j} = 0\ \%$. To calculate the Phase Separation Parameter (PSP) for a HEA the following equation is used

$$\mathrm{PSP} = \frac{\sum_{i\neq j} \mathrm{Separation}_{i-j} \times c_i \times c_j}{\sum_{i\neq j} \mathrm{Mixing}_{i-j} \times c_i \times c_j} \qquad 3.3$$

where the $\mathrm{Separation}_{i-j}$ and Mixing_{i-j} are used from the HEA constituent binary systems.

$\mathrm{Separation}_{i-j}$ and Mixing_{i-j} are illustrated using the same HEA as used to calculate binary phases field for PFP_X. Figure 3.3 shows two binary phase diagrams of the $Al_2CoCrCuNi$ HEA with different separation effects. In Figure 3.3a, the large positive ΔH_{mix} of the Cr-Cu binary prevents them from having a mixing effect. In HEAs, Cu and Cr tend to reside in the different phases. In Figure 3.3b, a large separation effect exists for the Co-Cu binary due to the positive ΔH_{mix} with a small mixing effect occurring at high temperatures. The Cr-Cu binary phase diagram at T_{pf} has a $\mathrm{Separation}_{Cr-Cu} = 100\ \%$ and $\mathrm{Mixing}_{Cr-Cu} = 0\ \%$, and in the Co-Cu binary phase diagram $\mathrm{Separation}_{Co-Cu} = 92\ \%$ and $\mathrm{Mixing}_{Co-Cu} = 8\ \%$.

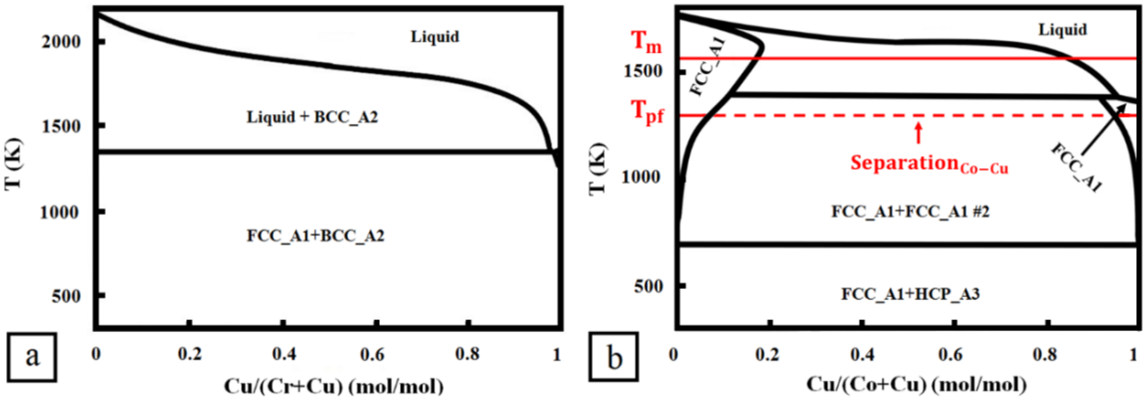


Figure 3.3: Two binary phase diagrams used to determine the binary phase separation percentage for HEA $Al_2CoCrCuNi$. (a) Phase diagram of Cr-Cu to show a complete phase separation effect.

(b) Overlay of the Co-Cu phase diagram illustrating the line segment method to determine the $\mathrm{Separation}_{\mathrm{Co-Cu}}$ for the HEA $Al_2CoCrCuNi$. Figure from Qi et al [55].

3.2.4 Visualization of the phase fields in parameter space

PFP_X parameters were calculated for six phases A1, A2, A3, AlNi type B2, Sigma, and Laves. The addition of PSP parameter makes seven parameters in total: PFP_{A1}, PFP_{A2}, PFP_{B2}, PFP_{A3}, PFP_{Laves}, PFP_{Sigma}, and PSP. A 7-dimension space with parameter axes was constructed to visualize the distribution of HEA phases. To visualize the position of a HEA in this 7D space, several projections in 2D and 3D space were selected. These plots show the partitioning of phase regions for two results based on (1) SS and (2) IM.

The HEAs in categories A1, A2, A1+A2, AlNi type B2+, and A3 were plotted for various combinations of the phase parameters. Figure 3.4 shows different plotted views highlighting HEA phase region separations. These views were selected based on the three parameters which best highlighted distinct HEA phase region separations. Figure 3.4a is a combination of the A1, A2, A1+A2, and AlNi type B2+ HEA phase regions. In Figure 3.4b, A1 and A2 HEAs are separated into high PFP_{A1} or PFP_{A2} regions. A high PFP_{A1} or PFP_{A2} value stabilizes A1 or A2 phase formation, respectively. The A1+A2 HEAs, in Figure 3.4c-d, are mostly in a region where neither PFP_{A1} nor PFP_{A2} is dominant. In Figure 3.4d, the higher PSP values for A1+A2 HEAs result in separation from the A2 HEAs. PSP prompts the formation of multiple phases due to the elemental repulsion. In Figure 3.4e-g, the phase regions of AlNi type B2+ HEAs are plotted against phase regions of A1, A2, and A1+A2 HEAs, respectively. PFP_{B2} is used to predict AlNi type B2 formation. AlNi type B2+ HEAs are all located in a region with relatively higher PFP_{B2} values. This indicates that having a high PFP_{B2} value corresponds to having a high probability of forming the B2 phase in a HEA. Figure 3.4f shows that all A3 HEAs are separated from the other phases because of high PFP_{A3} values indicating a higher chance of forming A3.

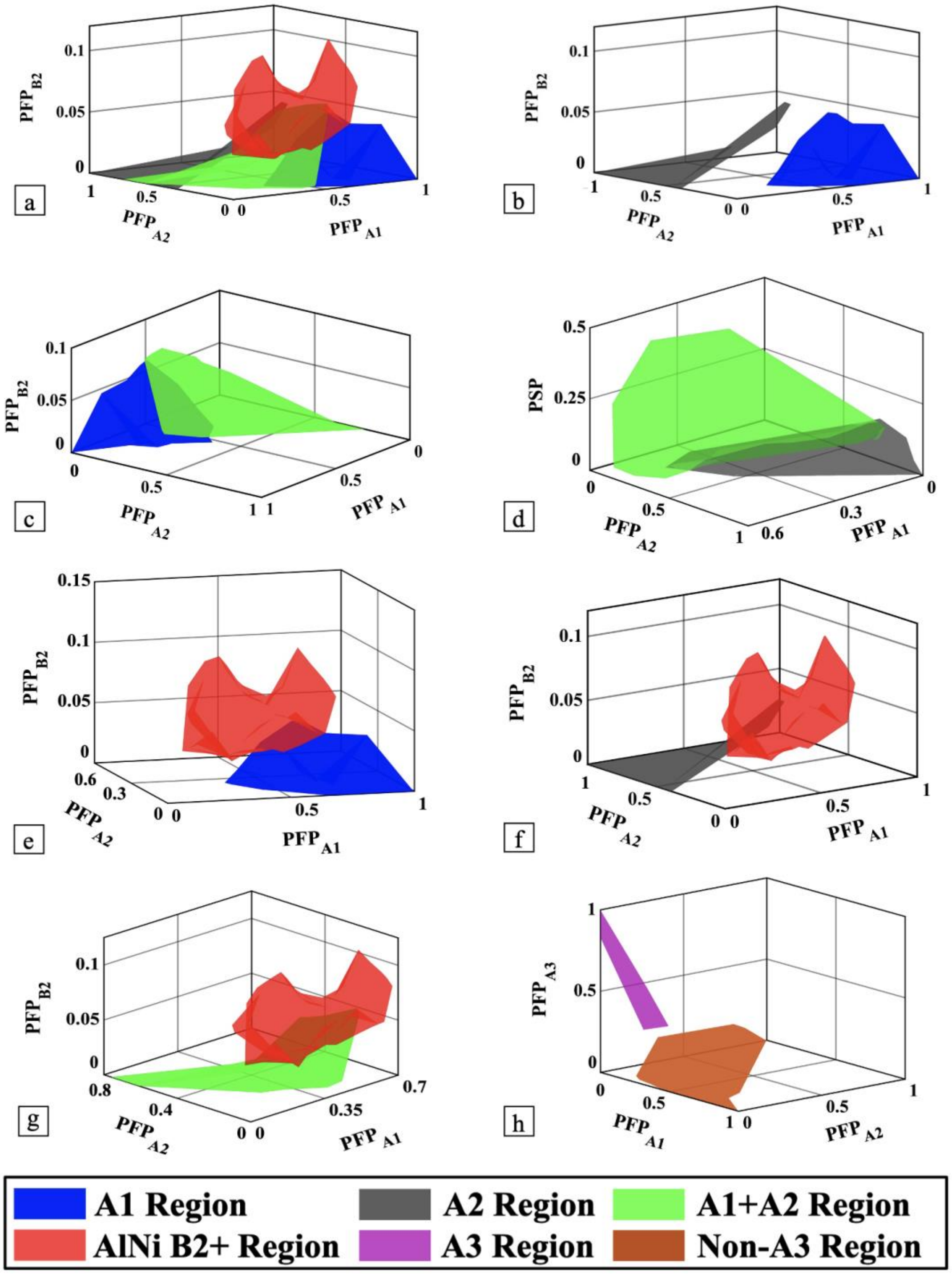


Figure 3.4: Visualizations of partitions among phases A1, A2, and A1+A2, AlNi type B2+, and A3. (a) PFP_{A1}, PFP_{A2}, and PFP_{B2} are plotted for A1, A2, A1+A2, and AlNi type B2+HEAs; (b) PFP_{A1}, PFP_{A2}, and PFP_{B2} are plotted for A1 and A2 HEAs; (c) PFP_{A1}, PFP_{A2}, and PFP_{B2} are plotted for phase regions of A1 and A1+A2 HEAs; (d) PFP_{A1}, PFP_{A2}, and PSP are plotted for phase regions of A2 and A1+A2 HEAs; (e)-(h) PFP_{A1}, PFP_{A2}, and PFP_{B2} are plotted to highlight the AlNi type

B2+ phase region relative to the A1, A2, and A1+A2 phase regions; and (h) PFP_{A1}, PFP_{A2}, and PFP_{A3} are plotted for phase regions of A3 and Non-A3 (not forming A3 phase) HEAs.

Sigma and Laves phases are the two predominant intermetallic phases present in HEAs, based on intermetallic phases present in the HEA database. In Figure 3.5, we group HEAs with Laves and Sigma phase as category IM+. HEAs without IM formation (Non-IM) HEAs were plotted with IM+ HEAs on a plot with axes PFP_{Sigma} and PFP_{Laves}. Although there is overlap between the phase distribution regions of the IM+ and Non-IM HEAs, IM+ HEAs exist in a region with large PFP_{Sigma} or PFP_{Laves} values.

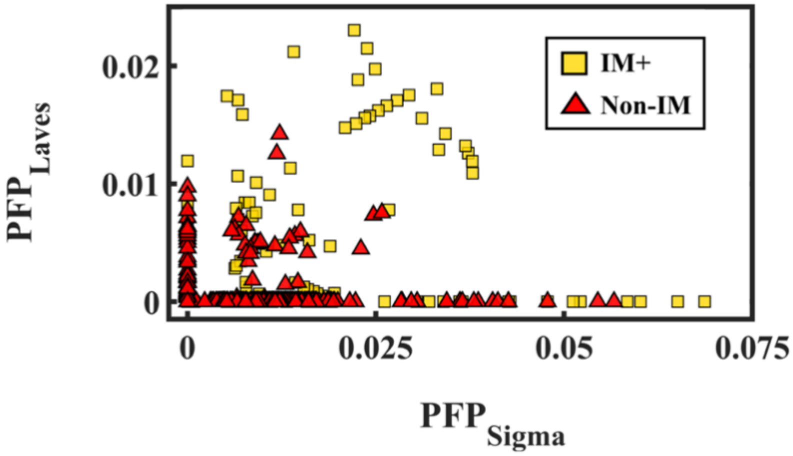


Figure 3.5: Parameters PFP_{Sigma} and PFP_{Laves} plotted for IM+ and Non-IM HEAs, where Non-IM includes A1, A2, A3, A1+A2, and AlNi type B2+. Figure from Qi et al [55].

The visualization methods for IM and SS HEA phases show that a large PFP_X value generally coincides with the formation of phase X. A large PSP value corresponds with phase separation and multiple phase formations. Different phase regions have overlaps on these plots. Due to the inherent limitations of visualizing seven parameters in 3D space, a better method was needed. Next, a ML model built based on PSP and PFP will be presented to give a quantitative evaluation of the effectiveness of these parameters.

3.2.5 Machine learning based on phenomenological features

The prior defined seven phenomenological phase-diagram based parameters were fed into a ML method as features and they were jointly used to make phase predictions. The ML algorithm called Random Forest was used. The HEA database used was divided into training and test sets, with training set percentages ranging from 10 % to 90 %. Test sets were composed of the remainder of the database.

The phase prediction success rates are shown in Figure 3.6. The HEA phase categories are A1, A2, A3, A1+A2, AlNi type B2+, and IM+ (by grouping Sigma+ and Laves+ HEAs). With the training set percentage being 90 %, the overall prediction success rate is 83 %. The prediction accuracy was generally higher for the single-phase A1, A2, and A3 and the ordered AlNi type B2 phase HEAs. The prediction accuracy of the A1+A2 mixed phase is relatively low. The features are closely correlated to the phase formation. The prediction accuracy decreases only slightly when the training set percentage decreases from 90 % to 50 %. The prediction is accurate even with a small training set. Including new HEAs may only marginally increase the accuracy for these features. When the training set percentage is below 50 %, the success rates drop due to the small training dataset size.

Overall, the phenomenological features are shown to be effective in predicting the phases with large homogeneity regions on phase diagrams. In next section, we will introduce a ML technique, feature engineering, that will further increase the ML accuracy and predict other phases not presenting on the binary phase diagrams.

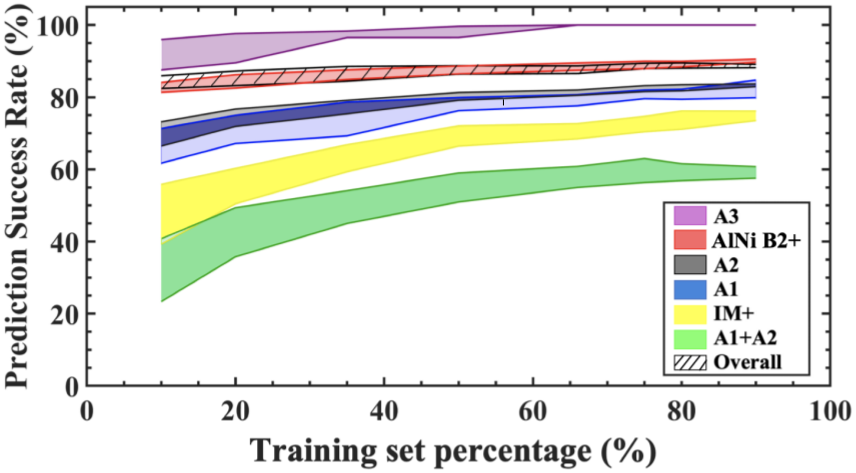


Figure 3.6: ML prediction success rates for HEAs in different phases are plotted. The shaded regions are the confidence bands for prediction success rates for different training set percentages. The width of the confidence bands represents one sigma deviation from the average prediction success rate. Figure from Qi et al[38].

3.3 Feature engineering and IM prediction

3.3.1 Overview

At the end of Section 2.4, we discussed the ML HEA phase prediction models in the literatures and identify two problems: (1) Can we achieve a more specific/detailed IM classification? (2) Can we predict more phase categories simultaneously and accurately to guide HEA design?

Detailed IM classification and prediction are certainly important in advancing the ML design of HEA beyond the common phases. Current knowledge indicates that Laves, Sigma, B2, and Heusler ($L2_1$) phases are four of the most common IM in HEA[142]. As discussed in Section 1.2, B2 and Heusler can be beneficial to HEA properties, while Sigma and Laves phases are normally unfavored due to their embrittling effects[16,17]. The need to achieve the predictive formation of beneficial IM while avoiding the unfavorable IM has led us to develop a more accurate and interpretable phase prediction method.

Feature engineering (FE), which has been underused in data science-driven materials research, has been successfully adapted to formulating superconducting critical temperature equations[143] and designing HEA[130]. HEA phase prediction is a complex problem that may not be efficiently executed by using individual features alone. Rather, features should interact with each other to expand the feature pool and transform the feature space through which the classification error is reduced. Thus, FE is applied in our work.

In this section, we will describe a FE strategy that synergistically blends phase diagram-based (PD) features (Section 3.2.3), thermodynamic (Thermo) features (Section 0), and Hume-Rothery rule (HR) features (Section 2.2.2) to interpret and predict the formation of different HEA phases. The feature engineering-based method enhances the accuracy of the ML models to near 90 % for nine HEA phases categories: FCC(A1), BCC(A2), HCP(A3), mixed FCC-BCC (A1+A2), AlNi type B2+, Laves+, Sigma+, Al-X-Y type B2+ and Heusler+ HEAs. As such, the present method predicts more phase categories with a higher level of specificity and accuracy than other reported methods to date.

The HEA phase classification methodology utilizes a two-layer method, as illustrated in Figure 3.7. The first layer corresponds with the multi-phase prediction model for SS (FCC, BCC, HCP, FCC+BCC) and common IM (AlNi type B2+, Laves+, and Sigma+). The model has an overall accuracy of 84 % in classifying specific phases in 835 HEAs. In particular, the accuracy for AlNi type B2+ is high at 90% while the accuracies for Laves+ and Sigma+ are lowered by ~10% from previous results in Section 3.2.5 where only PD features are used without FE. Accordingly, the second layer consists of four models that are grouped into two pairs for IM prediction. If a HEA is predicted as one of the commonly occurring Laves+ or Sigma+ in the first layer, then the verification from two models in the second layer, as shown on the left in Figure 3.7, will result in accuracies above 90 % for both phases. On the other hand, if the multi-phase prediction model predicts no Laves+ or Sigma+ formation, two other models will evaluate whether the HEA can form IM Heusler or Al-X-Y type B2 phases, with accuracies of 92 % and 80 %, respectively. In other words, the two-layer method herein can predict single-phase HEAs as well as HEA composites comprising specific phases with high accuracy.

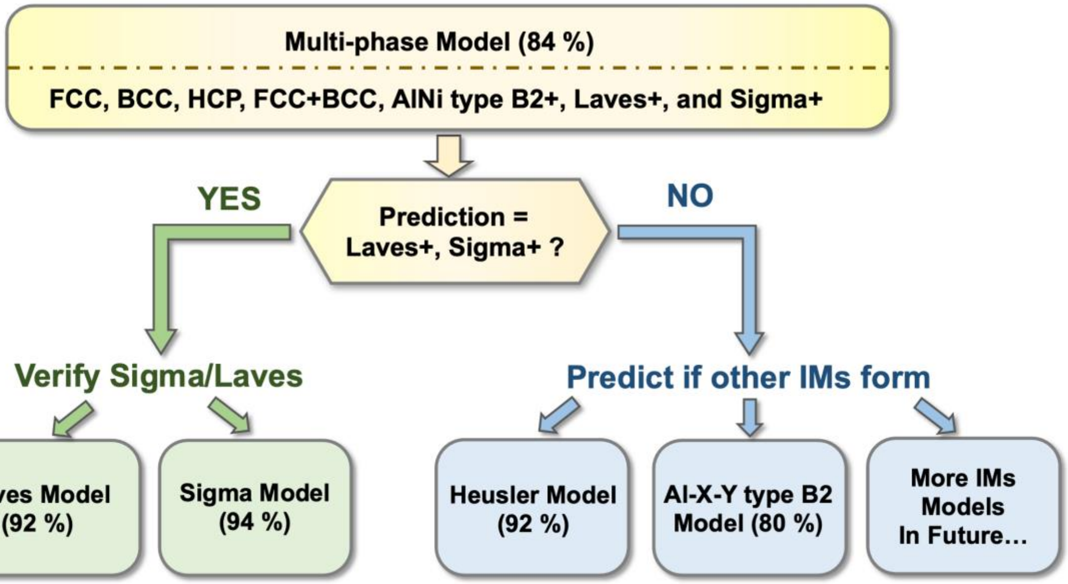


Figure 3.7: Two-layer method for predicting HEA phases. Pathways of modeling, with resulting classification accuracy (parentheses) for each model, are shown.

3.3.2 Feature engineering method

Feature engineering is a technique for developing and identifying the best math variations of raw features. The process includes feature construction, transformation, reduction, and selection.

The feature construction process collects the individual raw physical features that may influence phase formation. All relevant raw features are included regardless of the degree of importance they possess in determining the phase. Unimportant features are filtered out in the later steps.

The feature transformation process (Figure 3.8A) transforms the raw features by first constructing mathematical variations x^2, x^{-1}, $\sqrt{x}$, $\ln(x)$, and e^x for each feature X. The different expressions can mathematically change how features influence the phase prediction in ML algorithms. For example, $\ln(x)$ or e^x may reduce or inflate the effect of the outliers compared to using feature X. Then, the feature pool is further expanded by grouping any two math variations, A and B, using operations A+B, A-B, A/B, and AB. This step creates some synergetic effects from multiple features. For example, the comparison effects (A-B, A/B) or joint effects (A+B, AB) may bring new insights into phase prediction. At this point, the feature transformation constructs a huge feature pool, which potentially includes engineered features more qualified for phase prediction than the raw features. The current work expands 17 raw features to ~ 25,000 engineered features. Then to select the best features from the pool, a systematic method including feature reduction and selection is provided below. The feature reduction and selection methods contain filtering, intrinsic, and wrapper methods:

1. Filtering method:

The Pearson Correlation Coefficient (PCC) between two features indicates their linear correlation strength. As shown in Figure 3.8B, PCC values approaching +1, -1, or 0 indicates a strong positive, strong negative, or no linear correlation. Strongly correlated features are considered to be inter-substitutable in ML. Therefore, only one feature is kept from any pair with |PCC| > 0.9 in this work.

2. Intrinsic method:

Direct feature selection from the filtered-out features is computationally expensive and unnecessary as many features are irrelevant to phase formation. Therefore, a rapid ML method, logistic regression (LR) with L1 (or Lasso) regularization, is used to remove the irrelevant features (Figure 3.8C). This algorithm will minimize the total prediction cost as follows:

$$J(\overrightarrow{W}) = \frac{1}{m}\sum_{j=1}^{m} \mathrm{Cost}\left[h_{\overrightarrow{w}}\left(\vec{F}^{(j)}\right), y^{(j)}\right] + \gamma \sum_{i=1}^{n} |w_i| \qquad 3.4$$

Herein, $J(\overrightarrow{W})$ is the prediction cost with feature weight vector $\overrightarrow{W} = [w_{1,} w_{2,} \ldots, w_n]$. The first term is LR prediction cost $\mathrm{Cost}\left[h_{\overrightarrow{w}}\left(\vec{F}^{(j)}\right), y^{(j)}\right]$, which is directly related to the classification error, wherein the cost function of predicting the j-th sample as $h_{\overrightarrow{w}}\left(\vec{F}^{(j)}\right)$ while the correct category is $y^{(j)}$. $h_{\overrightarrow{w}}\left(\vec{F}^{(j)}\right)$ is obtained based on feature weights $\overrightarrow{W}$ and feature values $\vec{F}^{(j)}$. m is the total sample count in the dataset. The second term is the regularization cost. n is the number of features. γ is the regularization strength. w_i is the i-th feature's weight in $\overrightarrow{W}$. To reduce $J(\overrightarrow{W})$, the first term tends to use more features to reduce the prediction error, while the second term tends to invalidate more features by zeroing their weight w_i. The trade-off between the two terms will activate the minimum number of essential features in ML. Tuning γ changes the regularization strength and regulates the number of selected/activated features. After this step, about 100 features are retained.

3. Wrapper method:

Sequential learning (SL), shown in Figure 3.8D, selects the best features iteratively from ~100 features. ML models built with different combinations of features are evaluated by the average *f1* error from thirty rounds of 5-fold cross-validations with different random seeds. SL starts with an empty feature set in the first round, tests each feature independently, and picks the feature with the lowest classification error. In the subsequent rounds, each unselected feature is tested combinatorically with the previously picked ones. Finally, the best feature combination to minimize the classification error is constructed.

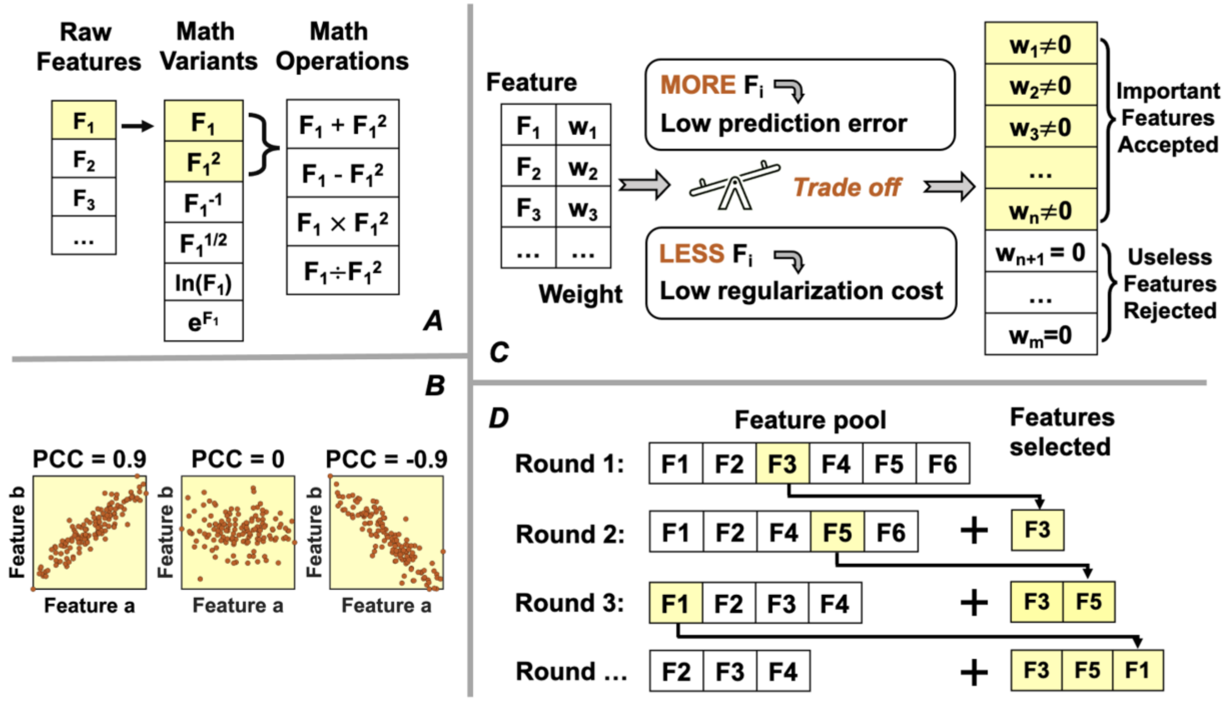


Figure 3.8: Process of feature engineering: (A) Feature expansion method; (B) PCC values reflect the linear correlations between two features; (C) Intrinsic method: LR with L1 regularization to eliminate features irrelevant to phase formation; (D) Wrapper method: SL selecting several best features for ML.

3.3.3 Results and discussion

A total of 17 thermodynamic (Thermo), Hume-Rothery rule (HR), and phase diagram-based (PD) features are used as the raw features in this work. They are: PD features PFP_{A1}, PFP_{A2}, PFP_{A3}, PFP_{B2}, PFP_{Laves}, PFP_{Sigma}, and PSP, Thermo features ΔS_{mix}[2], ΔH_{mix}[57], Ω[59], Φ[62], η[63], and k_1^{cr}[64], and HR features δ[57], $\frac{E_2}{E_0}$[69], Δχ[71], and VEC[73,74,144] are used.

Part 1: Multi-phase Prediction Model

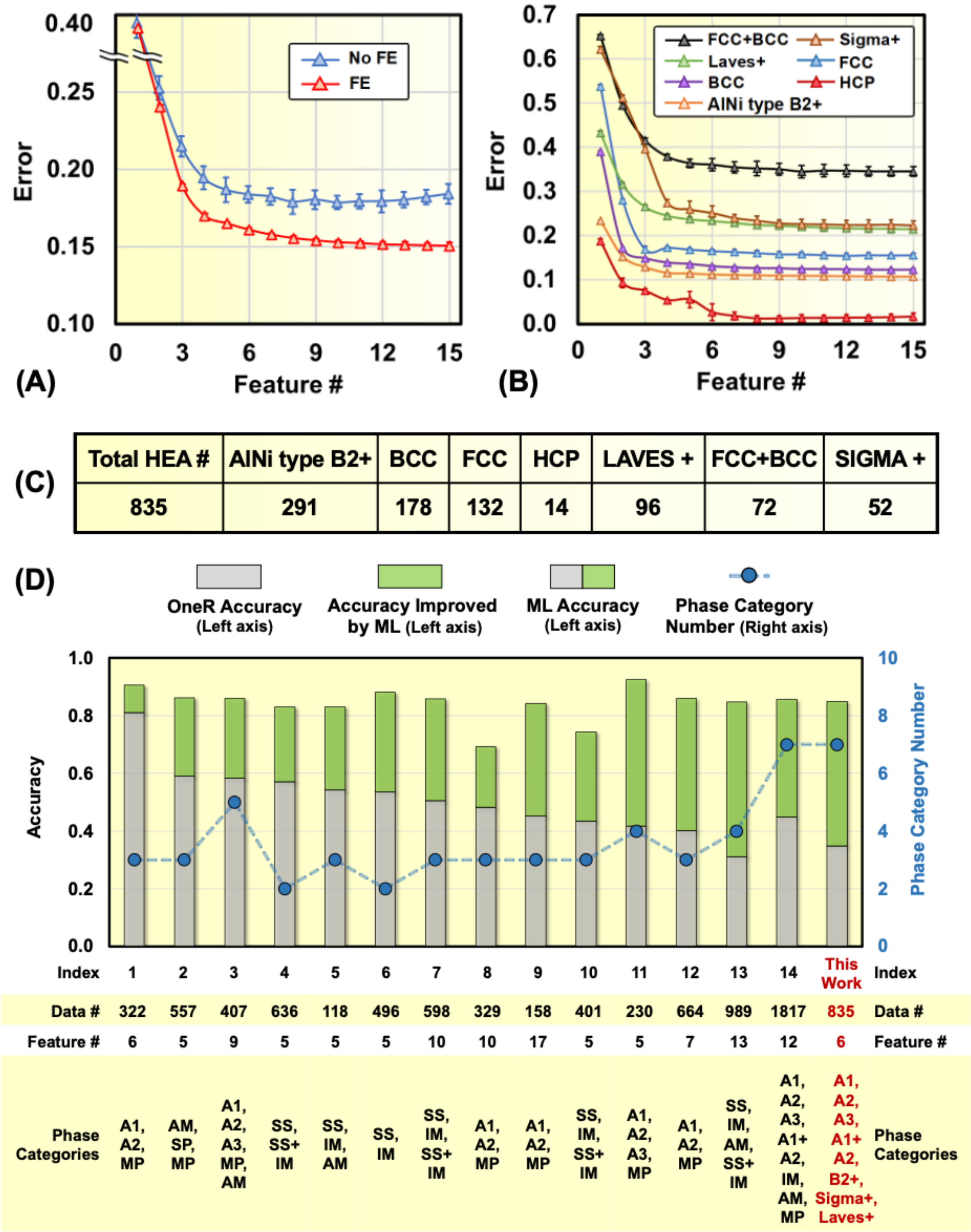
As described in the methodology section, the multi-phase prediction model in the first layer (Figure 3.7) has seven categories: FCC, BCC, HCP, FCC+BCC, AlNi type B2+, Laves+, and Sigma+. The Random Forest classification algorithm is used to perform sequential learning (SL). Thirty rounds of SL were conducted. Figure 3.9A shows the overall classification errors and the error bars (standard deviation) plotted against the number of top-ranked raw features (labeled as "No FE") and engineered features (labeled as "FE"), respectively. We only keep the first six engineered features to train the ML prediction model because adding more features only increases the risk of over-fitting disproportionately to the diminishing gains in accuracy. A list of these features is presented in Table 3.1. The FE classification error with six features is 0.161, 10 % smaller than the error without FE. Figure 3.9B shows the classification errors of the individual phase category plotted against the number of top-ranked engineered features. HCP, AlNi type B2+, FCC, and BCC predictions have lower errors while FCC+BCC, Sigma+, and Laves+ predictions are relatively less accurate. Therefore, the IM formation needs to be verified by the models discussed in the next section.

We have collected ~ 40 ML models on HEA phase prediction studied by other researchers, 15[50–52,54,130,139,145–150] of which that best represent advances in terms of model accuracy and complexity are summarized and compared with our work in Figure 3.9D. The comparison is based on criteria: OneR accuracy, ML accuracy, the level of detail on phase categories, the number of phase categories, and the number of features.

OneR accuracy is a baseline accuracy to show the data complexity of a classification model. In general, the models with fewer categories and more biased data distribution are less complex and have higher accuracy. The accuracy of OneR is computed in Eqn. 3.5.

$$\text{OneR accuracy} = \frac{\#\text{ of data in the largest category}}{\#\text{ of data in the entire database}} \qquad 3.5$$

OneR predicts all HEAs to be the phase with the most data. For example, in a database with 70 phases A and 30 phase B HEAs, the most accurate random guess is predicting all HEAs to be in phase A with an OneR accuracy of 0.7. A high OneR accuracy indicates that the database lacks complexity, and the model can easily obtain high accuracy. As a result, instead of using the ML accuracy, using the accuracy improvement from OneR accuracy to ML accuracy is a better evaluation metric of the model effectiveness. As shown in Figure 3.9D, our model has the second-lowest OneR baseline accuracy (gray bar height), among the highest accuracy improvement by ML (green height), and an overall accuracy (green and gray bars total height) of 0.84. At the same time, our model can classify more phase categories (seven, shown by the blue dashed line.) with a much higher level of category detail while using fewer features (six) than the other models. Moreover, we also address the important phase categories AlNi type B2+, Laves+, and Sigma+, which have been rarely attempted by ML. Overall, our FE-assisted ML model shows high capability in classifying HEA phases compared to the current models.



Total HEA #	AlNi type B2+	BCC	FCC	HCP	LAVES +	FCC+BCC	SIGMA +
835	291	178	132	14	96	72	52

Index	1	2	3	4	5	6	7	8	9	10	11	12	13	14	This Work	Index
Data #	322	557	407	636	118	496	598	329	158	401	230	664	989	1817	835	Data #
Feature #	6	5	9	5	5	5	10	10	17	5	5	7	13	12	6	Feature #
Phase Categories	A1, A2, MP	AM, SP, MP	A1, A2, A3, MP, AM	SS, SS+ IM	SS, IM, AM	SS, IM	SS, IM, SS+ IM	A1, A2, MP	A1, A2, MP	SS, IM, SS+ IM	A1, A2, A3, MP	A1, A2, MP	SS, IM, AM, SS+ IM	A1, A2, A3, A1+ A2, IM, AM, MP	A1, A2, A3, A1+ A2, B2+, Sigma+, Laves+	Phase Categories

Figure 3.9: (A) Overall classification error of multi-phase prediction model (first layer) versus the number of top-ranked features is plotted with error bars (standard deviation). Results with and without FE are shown. (B) Classification errors for individual phase categories versus the number of engineered features. (C) The number of HEA data in the database and each phase category. (D) Comparison of fifteen ML HEA phase prediction models. Models with index number 1-14 are from Li et al[52], Xiong et al.[147], Di et al.[130], Krishna et al.[148], Islam et al.[50], Xiong et al.[147], Risal et al.[149], Roy et al.[150], Agarwal et al.[54], Huang et al.[51], Xiong et al.[147], Jaiswal et al.[145], Lee et al.[146],

and Lee et al.[139], respectively. Accuracy values are labeled on the left vertical axis. Gray bar height represents the OneR accuracy. Green bar height represents the accuracy improvement from OneR accuracy to ML accuracy. Gray and green bars' total height represents the ML accuracy. The Blue dashed line represents the number of phase categories (right vertical axis). Below the plot is a list of references, numbers of data and features, and the phase categories for each model. A1, A2, A3, B2, SP, MP, AM, and IM are abbreviations for FCC, BCC, HCP, AlNi type B2, single phase, multi-phase, amorphous, and intermetallic phases.

Table 3.1: Engineered Features selected for phase prediction.

Prediction Model	**Features**
Multi-phase	η, $\mathrm{PFP_{A1}} - \mathrm{e}^{\mathrm{PFP_{A3}}}$, $\frac{\mathrm{E_2}}{\mathrm{E_0}} \cdot \Delta\mathrm{H_{mix}}$, $\Delta\chi^2 \cdot \sqrt{\mathrm{PFP_{Laves}}}$, $\mathrm{PFP_{Sigma}} \cdot \Phi$, $\mathrm{PFP_{A3}}/\mathrm{e}^{\delta}$
Laves+	$\mathrm{k_1^{cr}}/\ln(\mathrm{PFP_{Laves}})$, $\Delta\mathrm{H_{mix}} \cdot \sqrt{\Omega}$, $\mathrm{PFP_{Laves}} \cdot \mathrm{PFP_{A1}}$, $\Phi \cdot \sqrt{\mathrm{PFP_{Laves}}}$
Sigma +	$\Delta\chi^2 \cdot \ln(\mathrm{PFP_{Sigma}})$, $\Delta\chi \cdot \mathrm{VEC}^2$, $\mathrm{PFP_{A1}} \cdot \sqrt{\mathrm{PFP_{A3}}}$, ${\mathrm{PFP_{B2}}}^2/\ln(\mathrm{PSP})$
Heusler +	δ/Φ, $\mathrm{PFP_{Sigma}} \cdot {\Delta\mathrm{H_{mix}}}^2$, $\mathrm{PFP_{B2}}/{\mathrm{PFP_{A2}}}^2$, $\mathrm{PFP_{B2}} \cdot \mathrm{PFP_{A3}}$
Al-X-Y type B2+	$\eta + \Delta\chi$, $\Delta\mathrm{S_{mix}} \cdot \mathrm{VEC}^2$, $\mathrm{PSP} \cdot \mathrm{PFP_{A3}}$

Part 2: Laves+, Sigma+, Heusler+, and Al-X-Y Type B2+ Prediction Models

The four models in the second layer (Figure 3.7) use Support Vector Machine as the classification algorithms. For the Sigma+ and Laves+ prediction models, the appreciable imbalanced data distributions require special handling. For example, the Sigma+ prediction model database consists of 52 Sigma-containing HEAs ($\mathrm{HEA_{Sigma}}$) and 783 HEAs without the Sigma phase ($\mathrm{HEA_{No\text{-}Sigma}}$). The imbalance makes the ML model biased to the dominant category $\mathrm{HEA_{No\text{-}Sigma}}$ and adversely affects the predictions for $\mathrm{HEA_{Sigma}}$. Conventional methods of handling imbalanced databases include *under-sampling*, and over-sampling methods such as the *Random Over-sampling*, Adaptive Synthetic Sampling Approach for Imbalanced Learning (*ADASYN*)[151] and Synthetic Minority Over-sampling Technique (*SMOTE*) [152].

The Four methods of handling imbalanced databases are compared using errors from 5-fold cross validations. Random Over-sampling method randomly generates new samples by repeating the samples in the minor dataset. ADASYN[151] and SMOTE[152] are synthetic over-sampling methods, which create virtual samples based on samples in the minor database. Under-sampling method randomly draws samples from the major dataset, to form a balanced training database with the minor dataset.

As shown in Figure 3.10, the under-sampling method shows the lowest error compared to other over-sampling methods. More importantly, although under-sampling method may have info loss due to data removal in the majority class in each round of simulation, this information loss can be overcome by bootstrapping the database and training multiple ML models based on the bootstrapped sub-database. Random Over-sampling method that creates repeated data for minority class may cause overfitting. ADASYN and SMOTE would expand the minority class by creating virtual data that are not physically existed. These new data do not have any physical meaning. Based on these reasons, we choose under-sampling as our unbalanced database handling method.

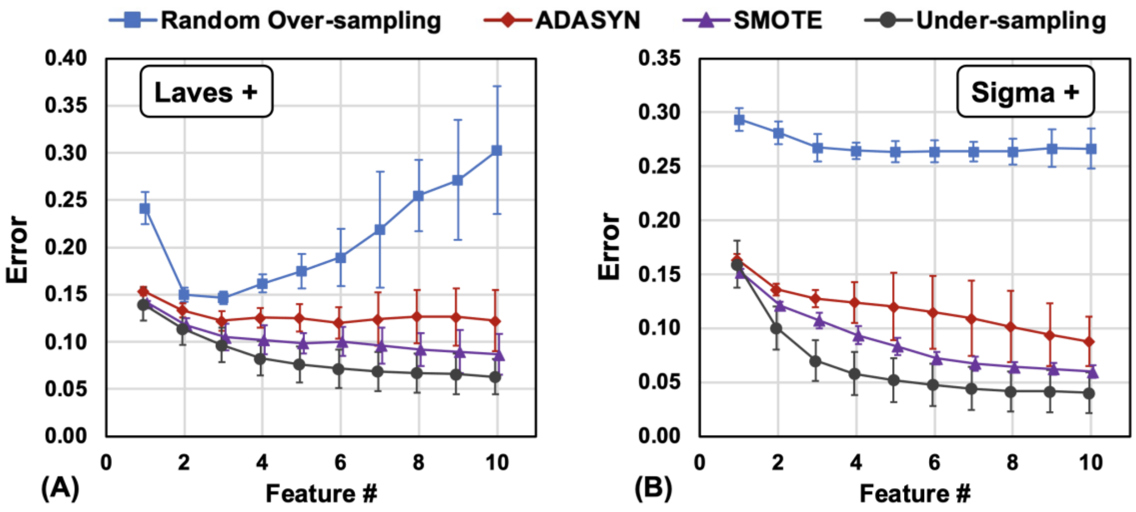


Figure 3.10: ML classification error decreases as the number of engineered features increases. The comparisons of the errors are presented for: (A) Laves+, and (B) Sigma+ models, among using Random Over-sampling, ADASYN, SMOTE, and Under-sampling methods. Error bars are presented in both plots.

The under-sampling method will randomly pick 52 samples from the $HEA_{No\text{-}Sigma}$ to constitute a ML database with the 52 HEA_{Sigma}. Thirty rounds of random samplings followed by sequential learning (SL) are conducted, and the average performance is presented.

Similarly, the Laves+ prediction model database consists of 96 Laves-containing HEAs (HEA_{Laves}) and 739 HEAs without the Laves phase ($HEA_{No\text{-}Laves}$). $HEA_{No\text{-}Laves}$ are under-sampled to 96 to constitute a ML database with the 96 HEA_{Laves} in each of the thirty random sampling rounds.

Figure 3.11A and B show how errors decrease with more features and compare the results with and without FE for Sigma+ and Laves+ predictions. In both models, only the first four engineered features will be kept for ML prediction, and we obtain low errors of 0.06 and 0.08 for Sigma+ and Laves+ predictions, respectively. FE suppresses the error by around 0.05 from No-FE results. The four features giving lowest error among the thirty rounds are presented in Table 3.1.

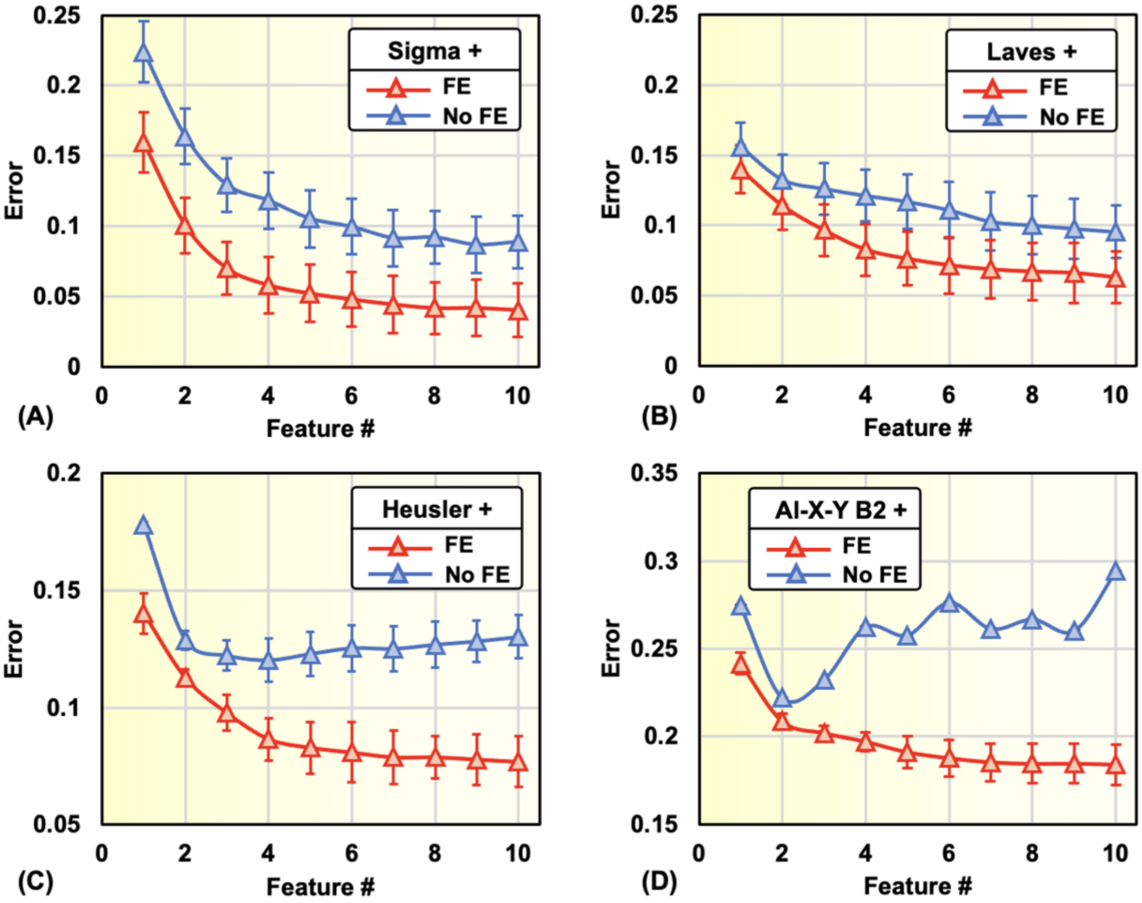


Figure 3.11: ML classification error decreases as the number of engineered features increases. The comparisons of the results between using and not using FE are presented for: (A) Sigma +; (B) Laves +; (C) Heusler +; and (D) Al-X-Y B2 + prediction models. Error bars (standard deviation) are presented in all plots. Small error bars may be invisible in figure (D).

A Heusler phase has a general composition X_2YZ, where X, Y, and Z symbolize specific groups of elements in the periodic table[153]. The database constitutes 77 HEAs containing the

Heusler phase (HEA_{L21}), and 109 HEAs without the Heusler phase ($HEA_{Non-L21}$). $HEA_{Non-L21}$ are selected based on the criteria: (1) they include appropriate X, Y, and Z elements for forming the Heusler phase; and (2) they are annealed to ascertain the non-emergence of the Heusler phase. Thirty rounds of SL are conducted. The average classification errors for using FE and not using FE are presented in Figure 3.11C. As more features are included, FE error becomes saturated, and No-FE error increases due to over-fitting. The top-ranked four engineered features (listed in Table 3.1) are kept for ML prediction with a classification error of 0.08. FE suppresses the error by 0.05 over No FE.

The refractory Al-X-Y type B2 phase comprises at least three components: X is Ti, Zr, or Hf; and Y is Cr, Mo, Nb, or V[15]. The database consists of 52 HEAs with Al-X-Y type B2 phase ($HEA_{AlXY-B2}$) and 35 without Al-X-Y type B2 phase ($HEA_{Non-AlXY-B2}$) but having Al, X, and Y elements. From thirty rounds of SL, the average classification errors for using FE and not using FE are presented in Figure 3.11D. As more features are included, the FE error continuously drops while the No-FE error increases rapidly due to over-fitting. The best three engineered features (listed in Table 3.1) are kept with a classification error of 0.2.

3.4 Phase formation interpretation

Although ML is a powerful classification tool, it is a black box and does not show the input and output relationships. Therefore, appropriate techniques are needed to evaluate the features' importance in determining phase formation.

Two sample T-test is the traditional statistical method to show the correlations between the features and phase categories. The permutation method shuffles each feature's values at a time and uses the error increase as the feature importance. The feature elimination method determines the feature importance by dropping a feature and computing the classification error increase or the prediction score decrease[139]. The single accuracy method uses only one feature for ML at a time and takes the prediction accuracy as the importance of the feature. The frequency selection method creates bootstrapping samples of the database, trains ML models on each of them, and uses the frequency of feature appearance in the bootstrapping models to represent the importance.

Additionally, specific ML algorithms, such as the LR, support vector machine (SVM), and neuron network (NN), can take the optimized feature weights as their importance scores.

The single accuracy method is utilized in this work. Although FE is found to reduce the prediction error, the feature variants generated are not amenable to direct physical meaning interpretation. Therefore, FE is not applied in this part. Moreover, PD features as phenomenological parameters are primarily effective in predicting SS phases in either the single-phase or mixed-phase state[38,55] without direct inferring physical mechanism of phase formation. On the other hand, Thermo and HR features can reflect the physics and are deemed to play an important role in the classification of specific IM. Therefore, we will identify the three most important IM formation determining Thermo and HR features from the feature importance values shown in Figure 3.12, and plot the HEA distribution probability density function based on the values of these features in Figure 3.13 to interpret their influence on specific IM formation.

From Figure 3.12A, the Heusler phase formation is mainly controlled by VEC, Φ, and $\frac{\mathrm{E}_2}{\mathrm{E}_0}$. $\mathrm{HEA}_{\mathrm{L21}}$ generally have lower VEC values than $\mathrm{HEA}_{\mathrm{Non\text{-}L21}}$ (Figure 3.13-H1). The low VEC implies that a BCC-prone environment[73,74,144] is favored for the Heusler phase formation, potentially due to the structural similarity between the Heusler and BCC lattices. Φ is a Thermo feature controlling IM/SS formation tendency. IM formation is favored when Φ is small[62]. Figure 3.13-H2 shows that $\mathrm{HEA}_{\mathrm{L21}}$ are generally low in Φ and energetically favored to form IM. Finally, $\mathrm{HEA}_{\mathrm{L21}}$ generally have larger $\frac{\mathrm{E}_2}{\mathrm{E}_0}$ values (Figure 3.13-H3), which represent larger atomic size difference[69]. This makes specific elements, such as Al, whose atomic size is different from the transition-metal elements, confined to certain sites on a crystal lattice, forming the ordered Heusler phase.

Al-X-Y type B2 formation is predominantly controlled by η, ΔS_{mix}, and Ω (Figure 3.12B). $\mathrm{HEA}_{\mathrm{AlXY\text{-}B2}}$ generally have more negative η values (Figure 3.13-B1), which indicates the IM formation tendency, consistent with DFT results[63]. $\mathrm{HEA}_{\mathrm{AlXY\text{-}B2}}$ also have a wide ΔS_{mix} distribution spectrum (Figure 3.13-B2) while $\mathrm{HEA}_{\mathrm{Non\text{-}AlXY\text{-}B2}}$ are clustered at the high ΔS_{mix} value region. Higher ΔS_{mix} prompts the disordering and suppresses the ordered $\mathrm{HEA}_{\mathrm{AlXY\text{-}B2}}$ formation. Ω is another Thermo feature showing the SS and IM formation tendencies[59]. $\mathrm{HEA}_{\mathrm{AlXY\text{-}B2}}$ generally have low Ω values (Figure 3.13-B3), which favors the ordered IM such as the B2 formation. More importantly, all three dominant features are thermodynamic, and HR features show limited influence. Electron environment-related HR features, VEC and $\Delta\chi$, are found to be correlated to FCC, BCC[39], and topological close-packed Sigma and Laves[72,144] but not B2 formation. Lattice distortion-related HR features, $\frac{E_2}{E_0}$ and δ, are relatively more important for predicting the IM with non-cubic structures (e.g., Laves) which can accommodate the severe atomic size mismatch. The B2 phase retains the BCC structure, where small lattice distortion should be expected for both disordered BCC and B2 phases. Despite the low effectiveness of the HR features, the key to ML predicting Al-X-Y type B2 is to distinguish it from the disordered BCC, where enthalpy and thermodynamic consideration are proven to be crucial in determining BCC/B2 ordering by a Monte Carlo and DFT combined study[44]. Our ML model draws a similar conclusion. In future, first-principles methods such as *ab-initio* simulations and DFT are promising to give an accurate, in-depth analysis of the order-disorder transition of such alloy systems.

For the Laves phase formation, $\frac{E_2}{E_0}$, η, δ, and ΔH_{mix} are the four most important features (Figure 3.12C). $\frac{E_2}{E_0}$ and δ both indicate the atomic size difference and the internal strain. As shown in Figure 3.13-L1 and L3, HEA_{Laves} have higher atomic size mismatches than the $HEA_{Non\text{-}Laves}$. The severe lattice distortion favors the ordered IM formation. From the thermodynamic aspects, the inset box plot in Figure 3.13-L2 shows that HEA_{Laves} all cluster at a region with low η absolute values while $HEA_{Non\text{-}Laves}$ has wide η distribution. Besides, HEA_{Laves} also show more negative ΔH_{mix} values than $HEA_{Non\text{-}Laves}$. The η and ΔH_{mix} distribution trends of HEA_{Laves} favors the IM formation.

Figure 3.12D shows that multiple features have weak impacts on Sigma formation. However, when these features are combined using FE, a low classification error of 0.05 is attained, illustrating the efficacy of the FE methodology used herein. The important roles of these features can be seen primarily in η, VEC, $\Delta\chi$, and ΔH_{mix} as examples. The inset of Figure 3.13-S1 shows that HEA_{Sigma} cluster at a region with low η absolute values, indicating a higher IM formation tendency. Similarly, HEA_{Sigma} shows more negative ΔH_{mix} values that favors IM formation. The influence of VEC, and $\Delta\chi$, both electron-related features, is shown in Figure 3.13-S2 and S3. It is previously reported that the formation of the topological close-packed Sigma phase formation is favored when $\Delta\chi > 0.133$[72] and $6.88 < VEC < 7.84$[144]. The current work obtains similar results based on a larger database. The first, second (i.e., median), and third quartiles of VEC distribution are 7.36, 7.61, and 7.86 (7.36 < VEC < 7.86 is the region for the middle 50 % of the distribution). This new Sigma-prone VEC region overlaps the FCC-prone VEC region[39]. A further review of the database also shows that ~80 % of HEA_{Sigma} contain FCC phase. Finally, the larger Δχ values of HEA_{Sigma} provide clear separation from the $HEA_{Non\text{-}Sigma}$. Therefore, one should consider decreasing the electronegativity discrepancy of the constituent elements to avoid Sigma formation during HEA design. The current work identifies the electron configuration as the most important HR factor in controlling Sigma formation.

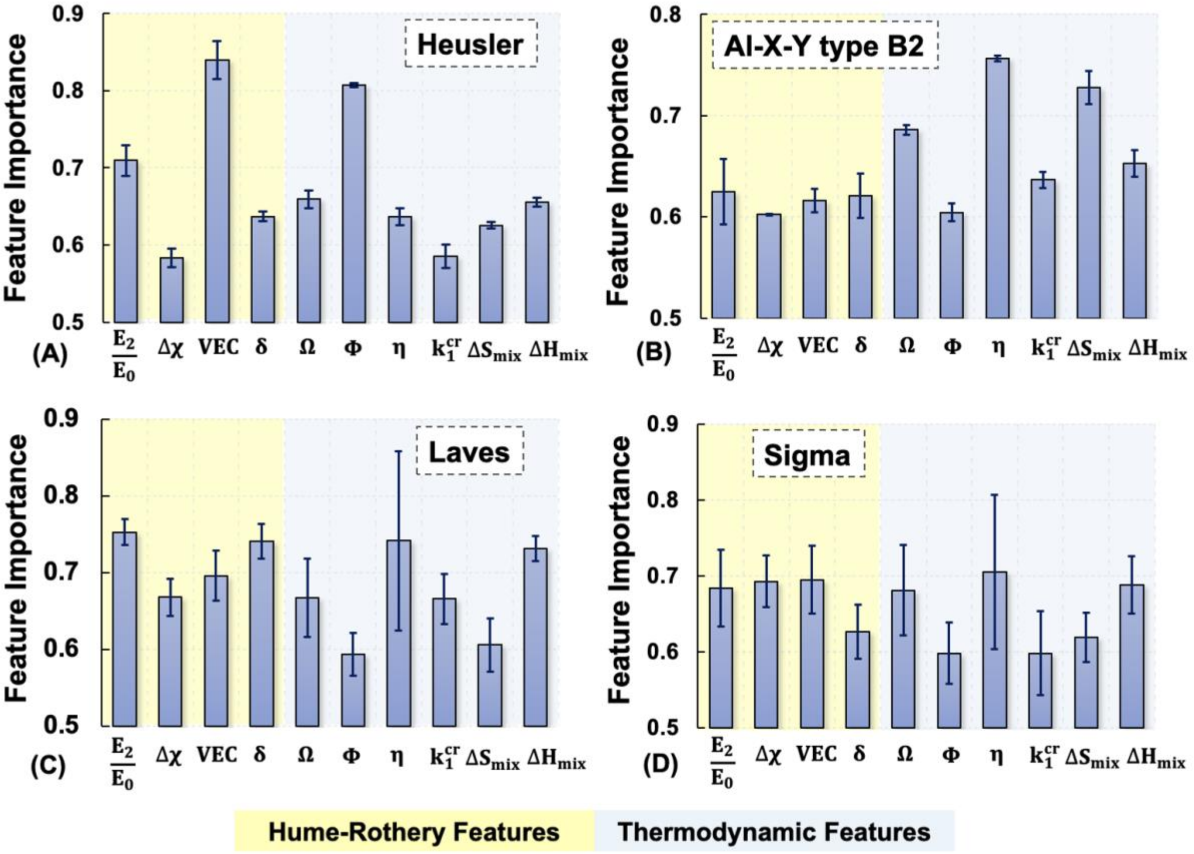


Figure 3.12: Feature importance in determining different phases' formation. Figures A-D are plotted for Heusler, Al-X-Y type B2, Laves, and Sigma phases. Yellow and blue backgrounds correspond to HR and Thermo features. Error bars (standard deviation) are shown.

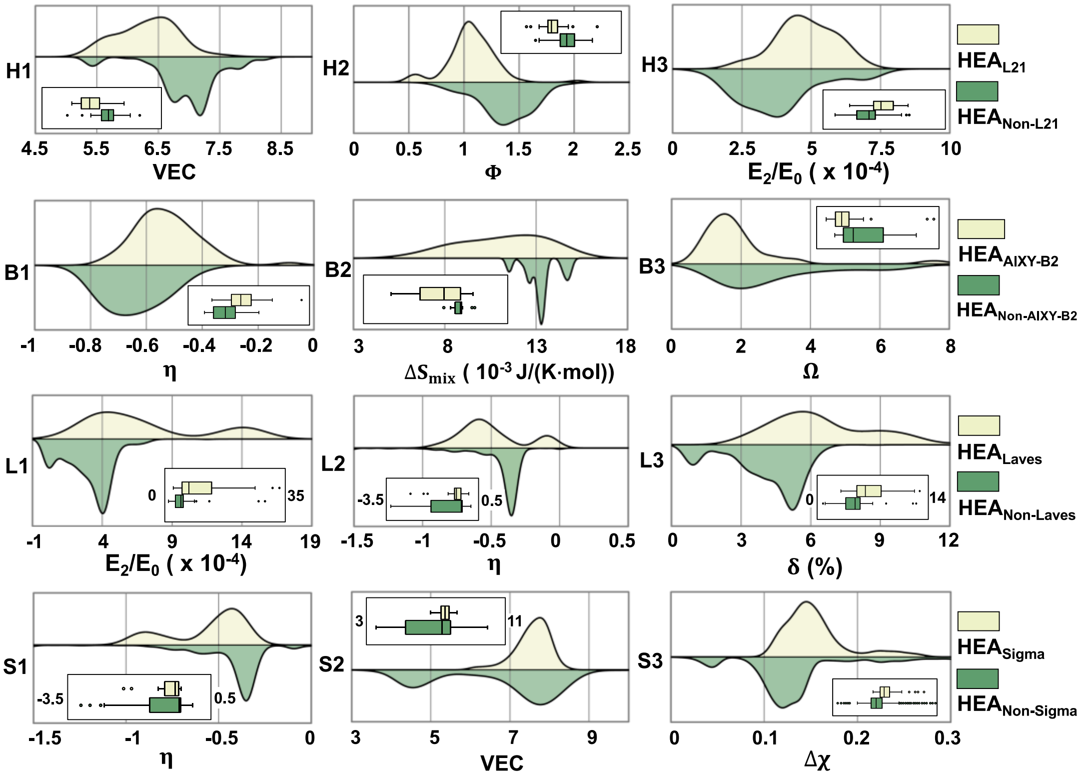


Figure 3.13: HEA distribution probability density functions based on the values of the three most important top-ranked features. The horizontal axis represents feature values. The vertical axis represents distribution probability density. Insets are box plots showing the relative positions of the two categories' distribution. The upper and lower bounds of box plots are labeled if different from the main plots. Figures H1-H3 show the HEA distribution based on VEC, Φ, and $\frac{E_2}{E_0}$ values in the Heusler+ prediction model. Figures B1-3 show the HEA distribution based on η, ΔS_{mix}, and Ω values in the Al-X-Y type B2+ prediction model. Figures L1-L3 show the HEA distribution based on $\frac{E_2}{E_0}$, η, and δ values in the Laves+ prediction model. Figures S1-3 show the HEA distribution based on η, VEC, and Δχ values in the Sigma+ prediction model.

3.5 Experimental validation for the models

3.5.1 Experiment methods

Alloys for validation were synthesized using arc melting (Section 0). Raw materials with a minimum purity of 99.97 wt.% were placed into a water-cooled copper crucible. Raw materials were melted five times under a high purity argon atmosphere. Each melt was conducted for a minimum of a minute. The sample was flipped over between melts to ensure homogeneity. All HEAs were characterized in the as-cast state, consistent with most data used in training the presented ML models. The ML models are set in the high-temperature ranges most suitable for as-cast alloys or alloys annealed at high temperatures, e.g., ~ 0.8 of the melting temperatures[55]. Finally, alloys were polished using grinding papers with grit sizes 180, 320, 600, and 1200. X-Ray Diffraction (XRD) measurements were conducted on a PANalytical Empyrean diffractometer with Cu Kα radiation and a scanning rate of ~0.15 degree/s.

3.5.2 Summary of validation results

Experimental validation is important to provide an unbiased evaluation of a ML model trained on available databases. As such, the palette of elements for the validation alloys should be an unbiased representation of the compositional space where the model is trained. Accordingly, of the 86 validation alloys, the multi-phase prediction model will have 60 alloys (Table 3.2A) with randomly chosen compositions based on the common element in the training database, located both inside and outside the feature space covered by the current database. The distributions of validation HEAs in each predicted phase category are proportional to the database phase distribution. 50 alloys are predicted correctly, giving a validation accuracy of 83 %. Since the Laves+, Sigma+, and multi-phase prediction models are trained on the same database, the same 60 HEAs also validate the Laves+ and Sigma+ models, with validation accuracies of 92 % and 95 %, respectively (Table 3.2A). To validate the Al-X-Y B2+ prediction model, another 14 new HEAs containing the Al-X-Y type B2 phase essential elements are randomly chosen that involves two or more refractory elements (Table 3.2B). For Al-X-Y type B2 formation, 12 out of the 14 HEAs are predicted correctly with an accuracy of 86 %. For a similar consideration, another 12 HEAs (Table

3.2C) containing the Heusler phase essential elements were synthesized to validate the Heusler+ prediction model. For the Heusler phase formation, 11 out of the 12 HEAs are predicted correctly, with an accuracy of 92 %. Overall, the validation accuracies essentially match the classification accuracies. All the XRD patterns can be found in the Figure 3.14.

Table 3.2: Validation HEAs compositions, phase prediction results, and experimental phase characterization results obtained from XRD are listed. The number subscripts in compositions are elemental atomic percentages. Detailed phase contents are listed in the experimental results column. True or False in Laves+, Sigma+, Al-X-Y B2+, and Heusler+ prediction columns represent forming or not forming the corresponding IM, respectively. Abbreviations AlNi B2+, A1, A2, Mix A1-A2, Al-X-Y B2+, and L21 represent the AlNi type B2 forming with other solid solution phases, disordered FCC_A1 phase, BCC_A2 phase, mixed A1-A2 phase (coexistence of multiple A1 or A2, or mixture of A1 and A2), Al-X-Y type B2+ forming with other phases, and Heusler phase. The incorrect predictions are underlined and bolded.

A. Multi-phase, Laves+, and Sigma+ prediction models validation HEAs

Composition	Multi-phase Prediction	Laves+ Prediction	Sigma+ Prediction	Experimental results
$Ag_{20}Al_{20}Cr_{20}Mn_{20}Ni_{20}$	AlNi B2+	False	False	B2+A1
$Ag_{5}Al_{38}Cr_{19}Mn_{19}Ni_{19}$	AlNi B2+	False	False	B2+A1
$Al_{5}Co_{20}Cr_{10}Fe_{40}Ni_{20}Ti_{5}$	**AlNi B2+**	**True**	False	A1
$Al_{10}Co_{20}Cu_{20}Fe_{20}Ni_{20}V_{10}$	AlNi B2+	False	False	B2+A1
$Al_{11}Co_{22}Cr_{11}Cu_{11}Ni_{33}V_{12}$	AlNi B2+	False	**True**	B2+A1
$Al_{15}Cr_{15}Mo_{15}Ni_{46}W_{9}$	AlNi B2+	False	False	B2+A1+A2
$Al_{15}Cr_{31}Fe_{31}Mn_{15}Ni_{8}$	AlNi B2+	False	False	B2
$Al_{16}Co_{20}Fe_{20}Mn_{18}Ni_{20}V_{6}$	AlNi B2+	False	False	B2
$Al_{16}Co_{21}Cr_{21}Fe_{21}Ni_{21}$	AlNi B2+	False	False	B2+A1
$Al_{16}Cr_{16}Fe_{16}Mn_{16}Ni_{31}V_{5}$	AlNi B2+	False	False	B2
$Al_{19}Cr_{19}Cu_{19}Fe_{19}Ni_{19}Si_{5}$	AlNi B2+	False	False	B2+A1+A2
$Al_{20}Co_{20}Cr_{20}Fe_{20}Mn_{20}$	AlNi B2+	False	False	B2
$Al_{21}Co_{11}Cr_{21}Cu_{5}Fe_{21}Mn_{21}$	AlNi B2+	False	False	B2
$Al_{22}Co_{26}Fe_{26}Ni_{26}$	**AlNi B2+**	False	False	A2
$Al_{23}Co_{23}Cu_{23}Fe_{23}V_{8}$	AlNi B2+	False	False	B2+A1
$Al_{23}Cu_{23}Fe_{23}Ni_{23}V_{8}$	AlNi B2+	False	False	B2+A1
$Al_{24}Co_{24}Cu_{23}Ni_{23}Ti_{6}$	AlNi B2+	False	False	B2+A1
$Al_{25}Co_{25}Cr_{25}Fe_{25}$	AlNi B2+	False	False	B2
$Al_{25}Cu_{25}Fe_{25}Ni_{25}$	AlNi B2+	False	False	B2+A1
$Al_{29}Co_{29}Cu_{13}Fe_{29}$	AlNi B2+	False	False	B2+A1
$Al_{33}Co_{17}Nb_{33}Ni_{17}$	**AlNi B2+**	**False**	False	B2+Laves
$Co_{7}Ta_{31}Ti_{31}V_{31}$	BCC	False	False	A2
$Cr_{6}Ti_{56}V_{19}Zr_{19}$	BCC	False	False	A2
$Cr_{25}Mo_{25}Ti_{25}V_{25}$	BCC	False	False	A2

$Cr_{33}Mo_{22}Nb_{12}V_{33}$	BCC	False	False	A2
$Hf_{25}Nb_{25}Ta_{25}Zr_{25}$	BCC	False	False	A2
$Hf_{30}Nb_{30}Ti_{30}V_{10}$	BCC	False	False	A2
$Hf_{30}Ta_{30}Ti_{30}V_{10}$	BCC	False	False	A2
$Mo_{29}Nb_{13}Ti_{29}V_{29}$	BCC	False	False	A2
$Nb_{22}Ta_{22}Ti_{22}V_{22}Zr_{12}$	BCC	False	False	A2
$Nb_{29}Ta_{29}Ti_{29}Zr_{13}$	BCC	False	False	A2
$Co_{15}Cr_{15}Fe_{15}Mn_{15}Ni_{32}V_{8}$	FCC	False	**True**	A1
$Co_{18}Cu_{18}Fe_{18}Mn_{18}Ni_{18}V_{10}$	**FCC**	False	False	A1+A1
$Co_{19}Cr_{29}Fe_{29}Ni_{19}Si_{4}$	FCC	False	False	A1
$Co_{21}Cr_{11}Fe_{42}Ni_{21}Ti_{5}$	FCC	**True**	False	A1
$Co_{22}Fe_{22}Mn_{12}Ni_{44}$	FCC	False	False	A1
$Co_{24}Cr_{24}Fe_{24}Ni_{24}Si_{4}$	FCC	False	False	A1
$Co_{24}Fe_{24}Ni_{47}V_{5}$	FCC	False	False	A1
$Co_{25}Cr_{8}Cu_{5}Fe_{25}Ni_{25}V_{12}$	FCC	False	False	A1
$Cr_{19}Cu_{19}Fe_{19}Mn_{18}Ni_{19}Ti_{6}$	**FCC**	**True**	False	A1+A2
$Al_{4}Cr_{32}Cu_{32}Fe_{11}Mn_{21}$	MIX A1-A2	False	False	A1+A2
$Al_{8}Cr_{56}Fe_{14}Mn_{22}$	**MIX A1-A2**	False	False	A2
$Al_{10}Co_{20}Cr_{10}Cu_{20}Mn_{20}Ni_{20}$	MIX A1-A2	False	False	A1+A2
$Al_{24}Co_{24}Cr_{23}Fe_{23}Ti_{6}$	**MIX A1-A2**	False	False	B2
$Co_{16}Cr_{16}Cu_{16}Fe_{16}Mn_{14}Ni_{16}Ti_{6}$	MIX A1-A2	False	False	A1+A1
$Co_{25}Cr_{25}Cu_{25}Fe_{25}$	**MIX A1-A2**	False	False	A1+A1+Unknown
$Cr_{25}Cu_{25}Fe_{25}Mn_{25}$	MIX A1-A2	False	False	A1+A2
$Cr_{40}Fe_{40}Mn_{10}Ni_{10}$	**MIX A1-A2**	False	False	A2
$Co_{20}Fe_{20}Mn_{20}Ni_{20}Ti_{10}V_{10}$	Laves+	True	False	Laves+A2
$Co_{20}Fe_{20}Mo_{20}Ni_{20}Ti_{20}$	Laves+	True	False	Laves+A1+A2
$Co_{21}Cr_{21}Cu_{21}Mn_{16}Ti_{21}$	Laves+	True	False	Laves+A1
$Co_{25}Cr_{25}Fe_{25}Nb_{13}Ti_{12}$	Laves+	True	False	Laves+A1+A2
$Cr_{20}Nb_{20}Ni_{20}Ti_{20}Zr_{20}$	Laves+	True	False	Laves+A2
$Cr_{40}Fe_{20}Ni_{20}Ti_{20}$	Laves+	True	False	Laves+A1+A2
$Cu_{17}Fe_{17}Mn_{17}Ni_{17}Ti_{32}$	Laves+	True	False	Laves+A1+A2
$Co_{15}Cr_{15}Cu_{8}Fe_{15}Ni_{31}Ti_{8}V_{8}$	**Sigma+**	**True**	**True**	A1
$Co_{18}Cr_{18}Fe_{18}Mo_{18}Ni_{18}V_{10}$	Sigma+	False	True	Sigma+A1
$Co_{20}Cr_{20}Fe_{20}Mo_{20}V_{20}$	Sigma+	False	True	Sigma+A2
$Co_{26}Cr_{26}Fe_{26}Mo_{22}$	Sigma+	False	True	Sigma+A2
$Cu_{20}Fe_{20}Mn_{20}Ni_{20}V_{20}$	Sigma+	False	True	Sigma+A1

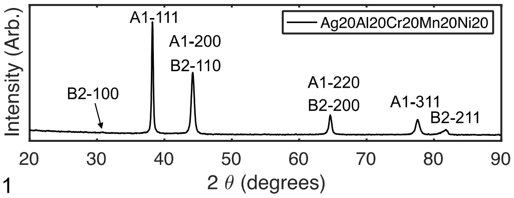
B. Al-X-Y type B2+ prediction model validation HEAs

Composition	Al-X-Y B2 + prediction	Experimental results		Composition	Al-X-Y B2 + prediction	Experimental results
$Al_{10}Hf_{20}Nb_{22}Ti_{33}V_{15}$	True	B2		$Al_{30}Nb_{20}Ta_{15}Ti_{20}V_{10}Zr_{5}$	True	B2
$Al_{15}Hf_{25}Nb_{32}Ti_{28}$	True	B2		$Al_{30}Nb_{20}Ta_{20}Ti_{20}Zr_{10}$	True	B2+Unknown
$Al_{20}Hf_{24}Nb_{29}Ti_{27}$	True	B2		$Al_{4}Hf_{6}Nb_{42}Ti_{18}V_{24}W_{6}$	False	A2
$Al_{23}Hf_{23}Nb_{23}Ti_{23}V_{8}$	True	B2		$Al_{8}Cr_{15}Mo_{15}Nb_{15}Ti_{15}V_{32}$	False	A2
$Al_{23}Hf_{23}Ta_{23}Ti_{23}V_{8}$	True	B2		$Al_{10}Hf_{18}Nb_{18}Ta_{18}Ti_{18}Zr_{18}$	False	A2

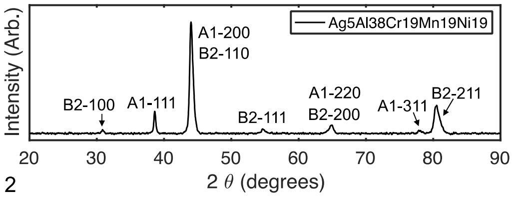
$Al_{26}Mo_{21}Nb_{11}Ti_{21}V_{21}$	**True**	A2		$Al_{32}Nb_{17}Ta_{17}Ti_{17}V_{17}$	**False**	B2
$Al_{30}Mo_{20}Nb_{20}Ti_{30}$	True	B2		$Al_{6}Nb_{21}Ta_{21}Ti_{21}V_{21}Zr_{10}$	False	A2

C. Heusler+ prediction model validation HEAs

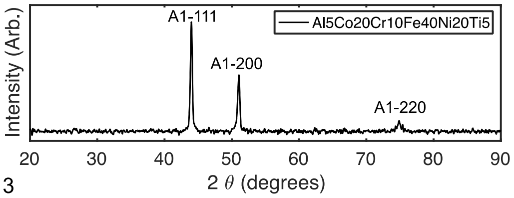
Composition	Heusler+ prediction	Experimental results		Composition	Heusler+ prediction	Experimental results
$Al_{10}Co_{25}Fe_{25}Mn_{25}Ti_{15}$	**True**	A2+Unknown		$Al_{25}Cr_{10}Fe_{20}Mn_{10}Ni_{20}Ti_{15}$	True	$L2_1$
$Al_{10}Cr_{5}Fe_{45}Mn_{12}Ni_{20}Ti_{8}$	True	$L2_1$+A1		$Al_{10}Co_{20}Mn_{20}Ni_{30}Ti_{10}$	False	A1+A2
$Al_{12}Co_{28}Fe_{19}Ni_{29}Ti_{12}$	True	$L2_1$+A1		$Al_{10}Co_{30}Fe_{20}Ni_{32}Ti_{8}$	False	A1
$Al_{14}Cr_{4}Fe_{17}Mn_{4}Mo_{1}Ni_{44}Ti_{16}$	True	$L2_1$+A1		$Al_{15}Co_{30}Fe_{30}Ni_{10}Ti_{15}$	False	B2
$Al_{15}Cr_{10}Fe_{30}Ni_{30}Ti_{15}$	True	$L2_1$+A2		$Al_{20}Fe_{10}Mn_{30}Ni_{40}$	False	B2+A1
$Al_{15}Fe_{40}Mn_{20}Ni_{10}Ti_{10}$	True	$L2_1$		$Al_{7}Co_{30}Fe_{30}Mn_{25}Ti_{8}$	False	B2+A1



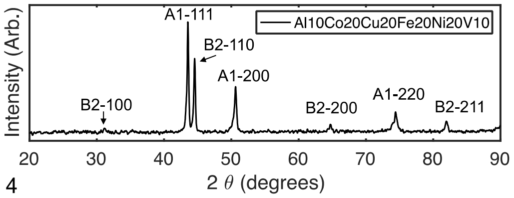
1



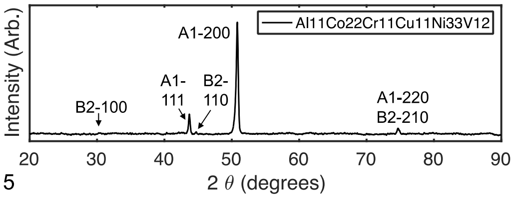
2



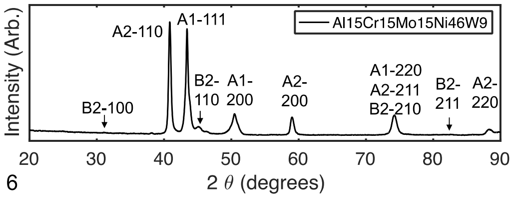
3



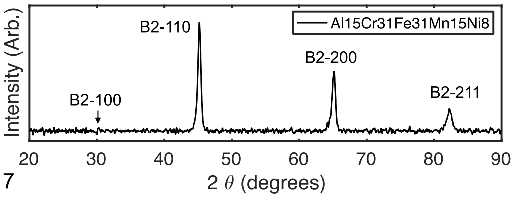
4



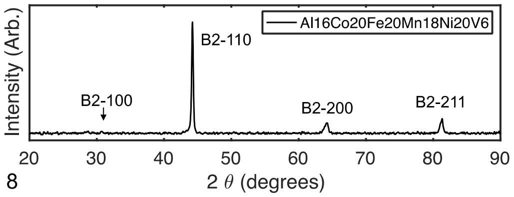
5



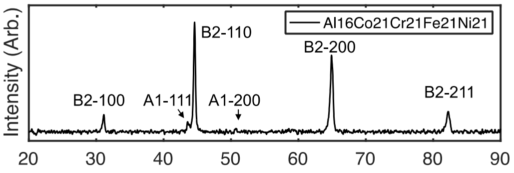
6



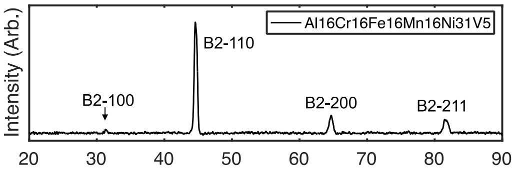
7



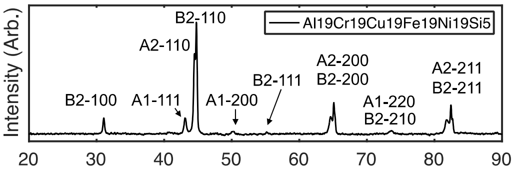
8



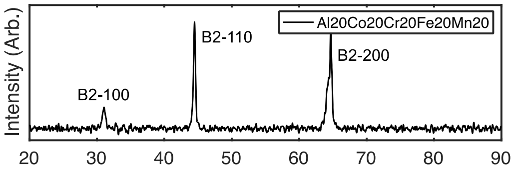
9



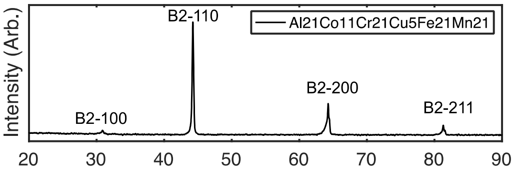
10



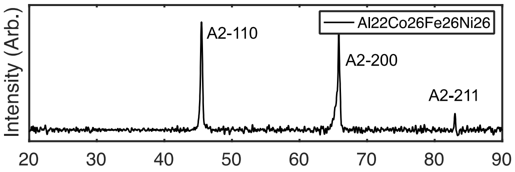
11



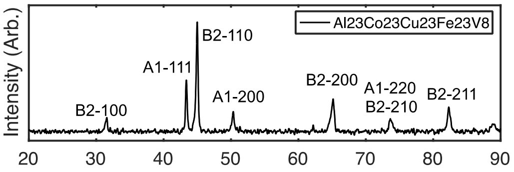
12



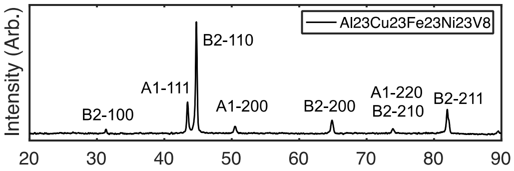
13



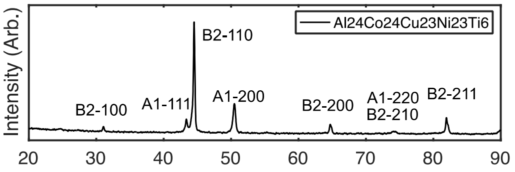
14



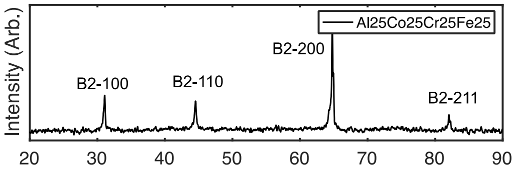
15



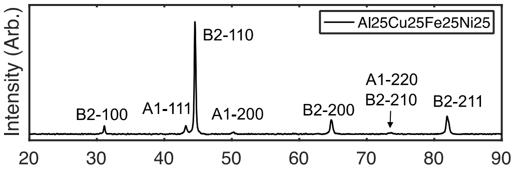
16



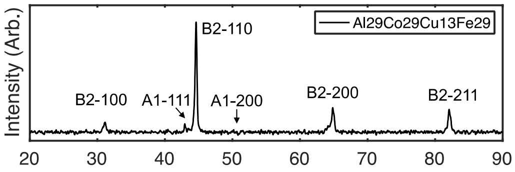
17



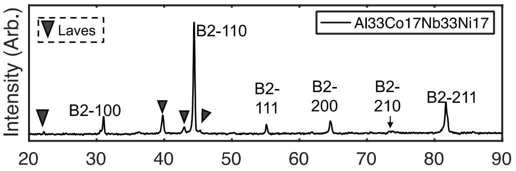
18



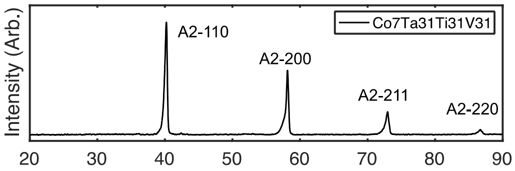
19



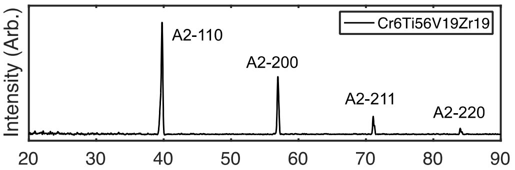
20



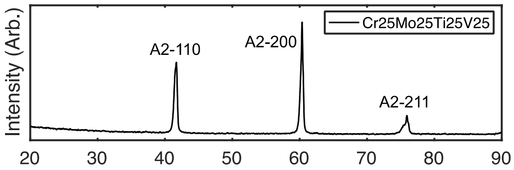
21



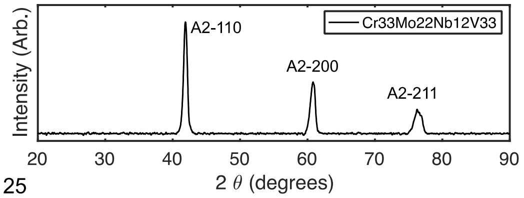
22



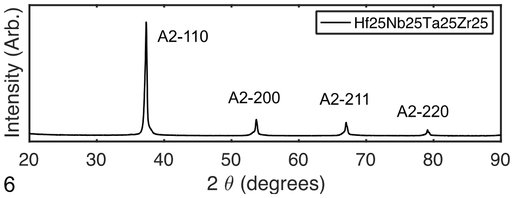
23



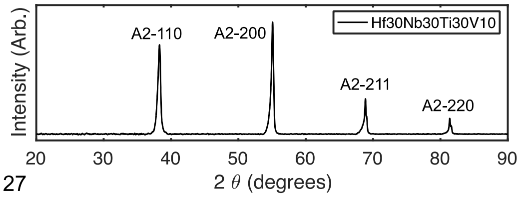
24



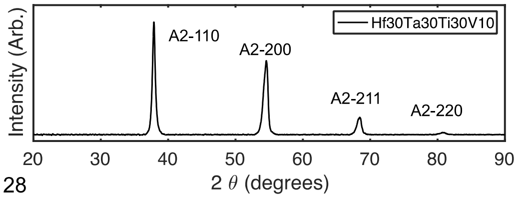
25



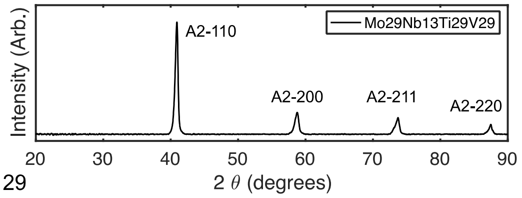
6



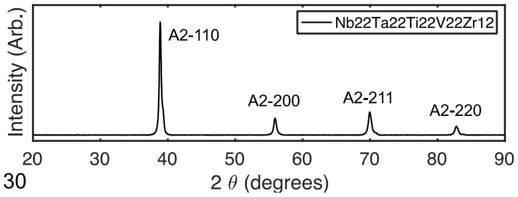
27



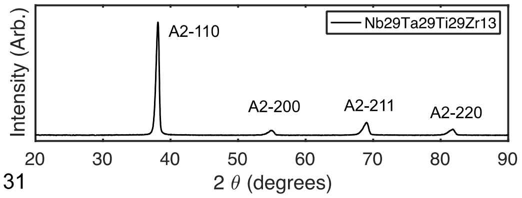
28



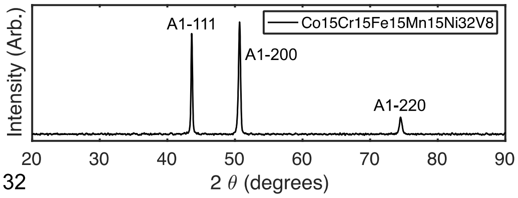
29



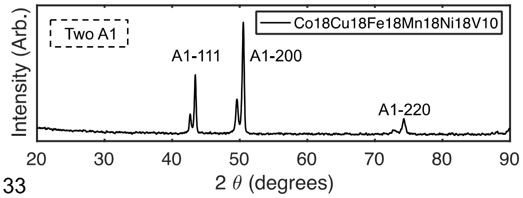
30



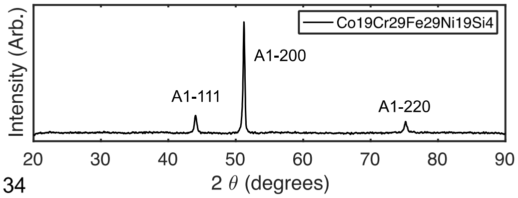
31



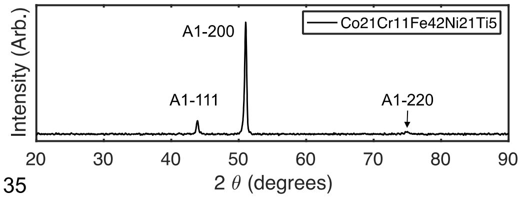
32



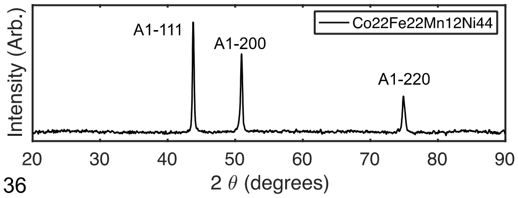
33



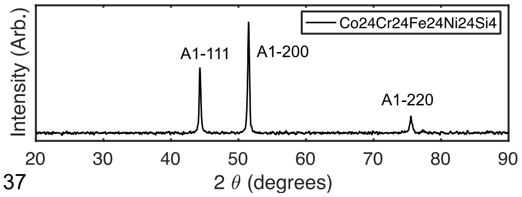
34



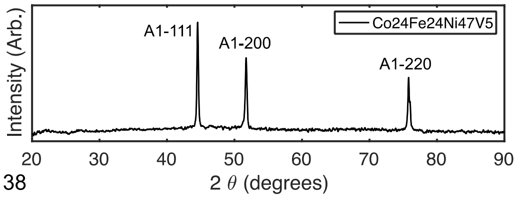
35



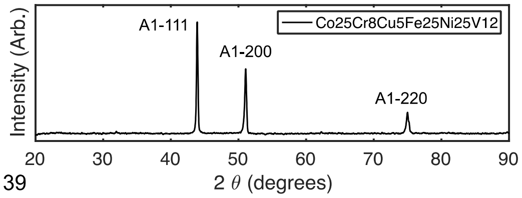
36



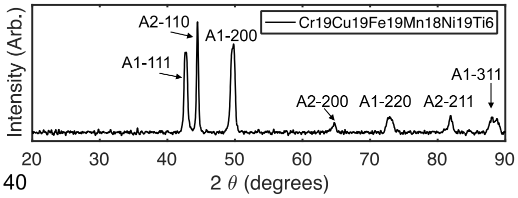
37



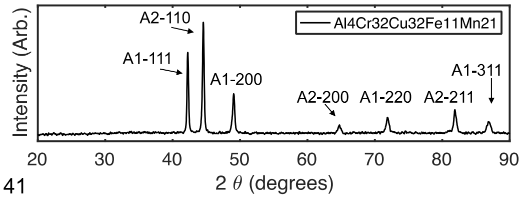
38



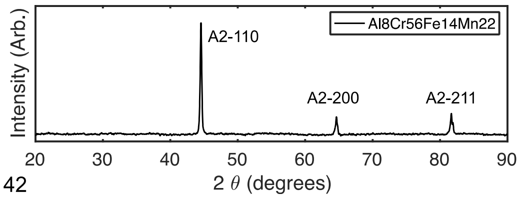
39



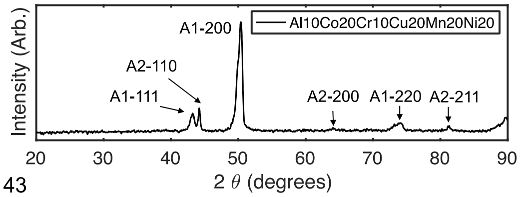
40



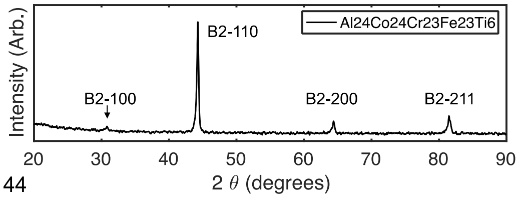
41



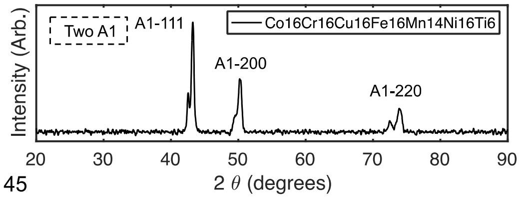
42



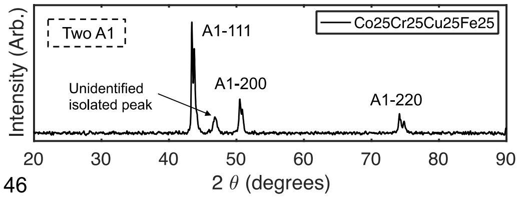
43



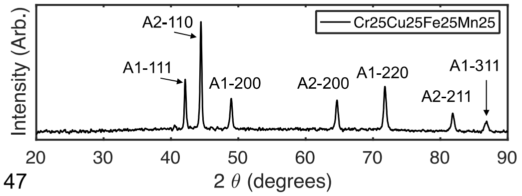
44



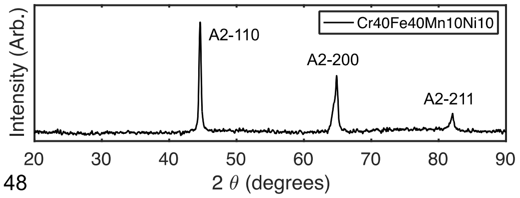
45



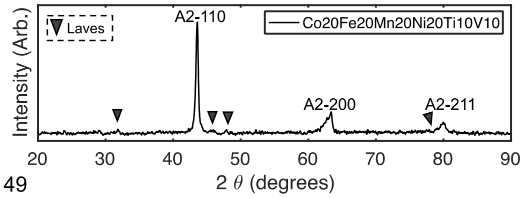
46



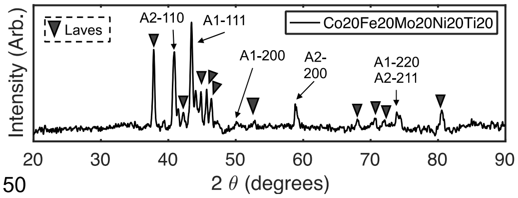
47



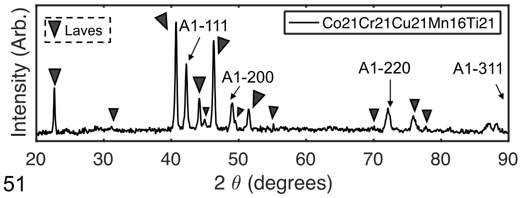
48



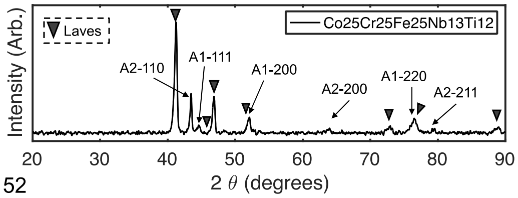
49



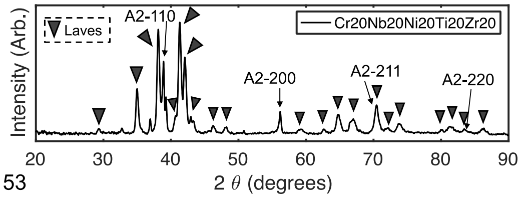
50



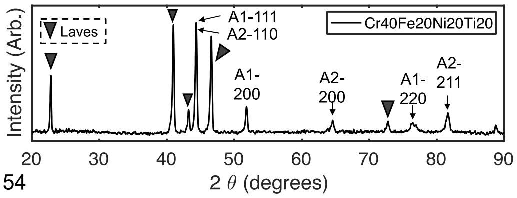
51



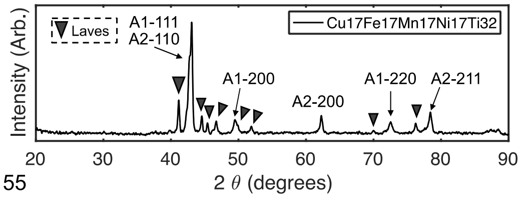
52



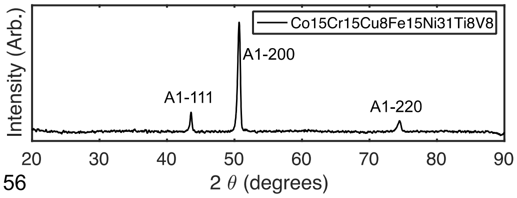
53



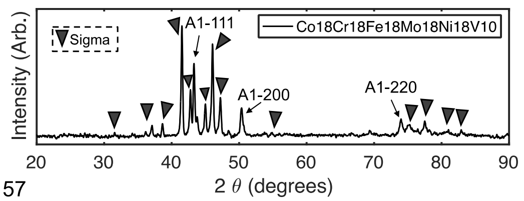
54



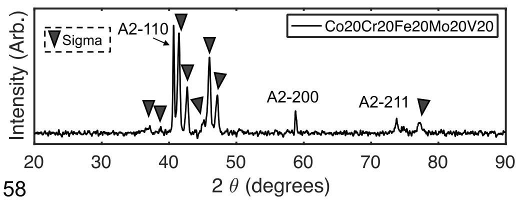
55



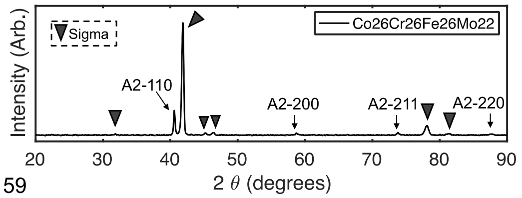
56



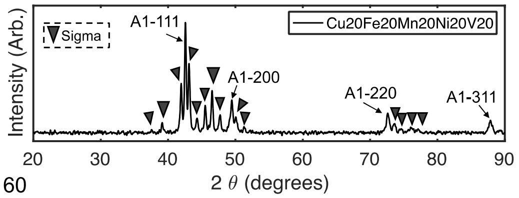
57



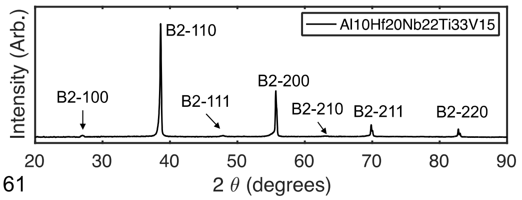
58



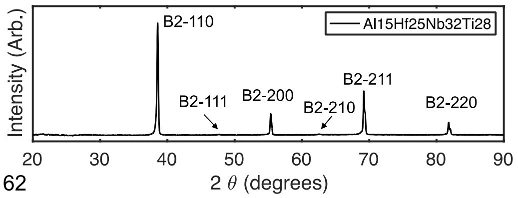
59



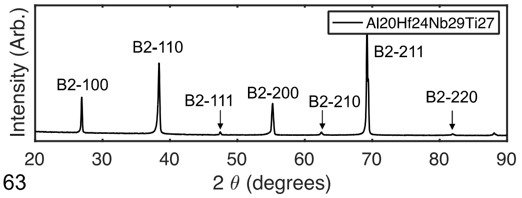
60



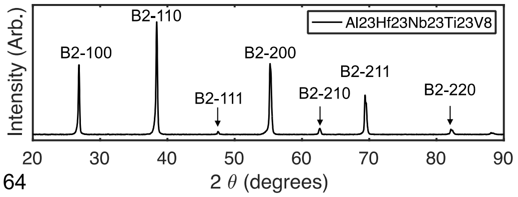
61



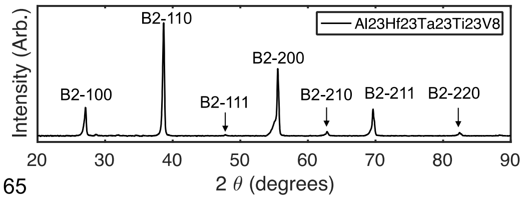
62



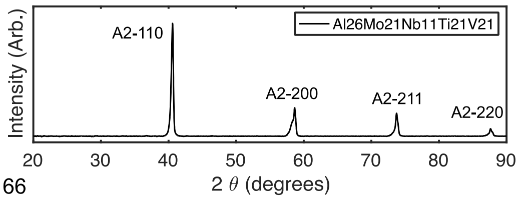
63



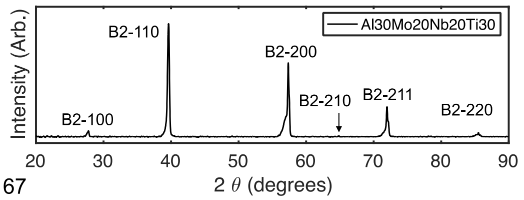
64



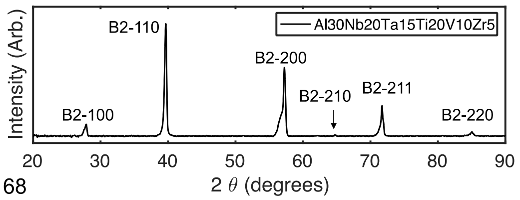
65



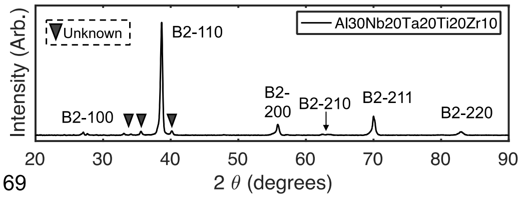
66



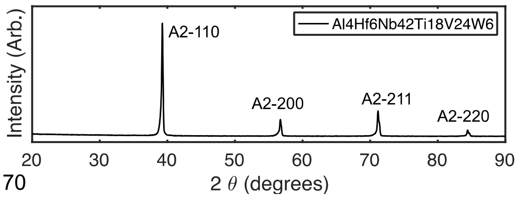
67



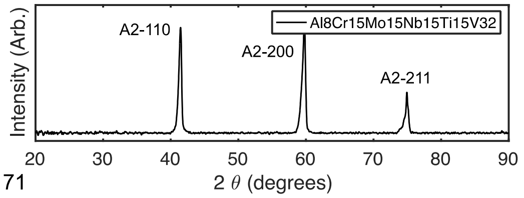
68



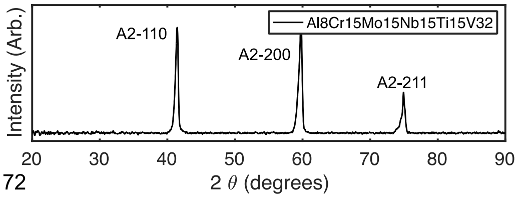
69



70



71



72

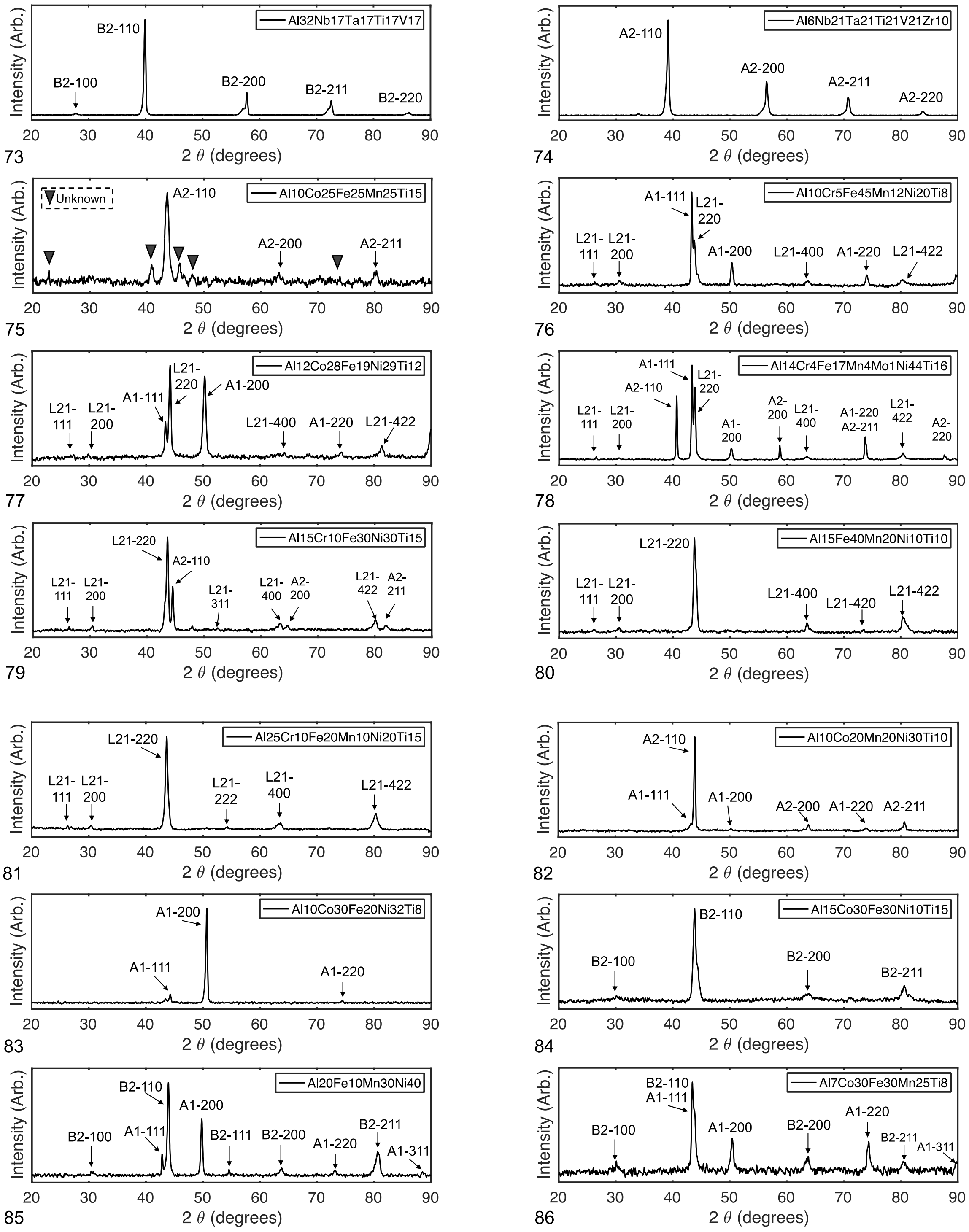


Figure 3.14: XRD patterns for newly synthesized validation HEAs.

Chapter 4 Review on High Entropy Alloys Properties Prediction Methods

This chapter will cover prediction models of fundamental properties like density, cost, and melting point as well as mechanical properties like hardness, strength, and fracture strain.

4.1 Density, cost, and melting Temperature

Alloy density is estimated by:

$$\frac{1}{\rho_{\mathrm{alloy}}} = \sum\nolimits_{\mathrm{all\ element\ i}} \frac{\mathrm{wt.\%_i}}{\rho_{\mathrm{i}}} \qquad 4.1$$

where ρ_{alloy} and ρ_i are the densities of the alloy and the i-th element, $\mathrm{wt.\%}_i$ is the weight percentage of the i-th element.

Alloy cost is calculated from:

$$\mathrm{Cost_{alloy}} = \sum\nolimits_{\mathrm{all\ element\ i}} \mathrm{wt.\%_i} \times \mathrm{Cost_i} \qquad 4.2$$

where $\mathrm{Cost_{alloy}}$ and $\mathrm{Cost_i}$ are the cost of the alloy and the i-th element, $\mathrm{wt.\%}_i$ is the weight percentage of the i-th element.

Melting temperature T_m of the alloy can be estimated using Eqn. 3.1 in Section 3.2.2.

4.2 Hardness

An empirical relationship between hardness and tensile strength for bulk metallic glass[154] may provide an quick and rough estimate of the HEA hardness. It is found that:

$$H_V \approx 3\ \sigma_{\mathrm{UTS}} \qquad 4.3$$

where H_V and σ_{UTS} are the Vickers hardness and ultimate tensile strength (UTS) of the metallic glass. This relationship works empirically for bulk metallic glass when σ_{UTS} is between 1 Gpa and 5 Gpa. However, one should obtain the σ_{UTS} before estimating hardness.

Recently, there are emerging efforts for using machine learning (ML) models to design HEAs with improved mechanical properties, principally hardness and strength. Chang et al. [48] used the artificial neuron network (ANN) ML algorithm to predict the hardness of HEAs and find new compositions with optimized hardness. Ninety-one HEAs containing Al, Co, Cr, Cu, Fe, Ni, Mn, and Mo with hardness data reported were contained in the database. The solid density, hardness, and atomic mass of each element, weighted by the atomic percentage of that element, were the features input in ANN. The ML model showed great agreement between the predicted and the experimental hardness results, with a value of 0.94 for the Pearson correlation coefficient. For designing new alloys with high hardness, a simulated annealing algorithm was adopted to change the composition systematically for finding global maximum hardness. HEAs designed from this model showed improvements in hardness. A general trend that the hardness increases for the same alloy system when the phase transforms from FCC to FCC+BCC to BCC+B2 is found. The similar method can be applied to other element systems.

Wen et al. [49] developed a robust ML method of making HEAs with high hardness. A radial basis function kernel (svr.r) ML model was used. ML features were the HEA compositions together with the empirical parameters $\mathrm{e/a}$, modulus mismatch, and the sixth square of the work function defined by the author. The dataset was composed of 155 AlCoCrCuFeNi HEA systems and their hardness values. As can be seen in Figure 4.1a, the method gave reliable hardness prediction results. After that, an iterating process was used to find the HEA with the highest hardness. In each iteration, ML could sample the HEA composition space and find three new HEAs with the highest predicted hardness. The hardness values of the new HEAs were then measured experimentally and added into the training dataset for the next iteration of ML. After seven iterations, 21 new HEAs with high hardness were obtained. Their hardness values compared to the original 155 systems were plotted in Figure 4.1(b). The HEA Al47Co20Cr18Cu5Fe5Ni5 with the highest hardness was obtained in the fourth iteration. Its hardness was 883 HV, which was 14 % higher than the highest hardness value of 775 HV in the original training data.

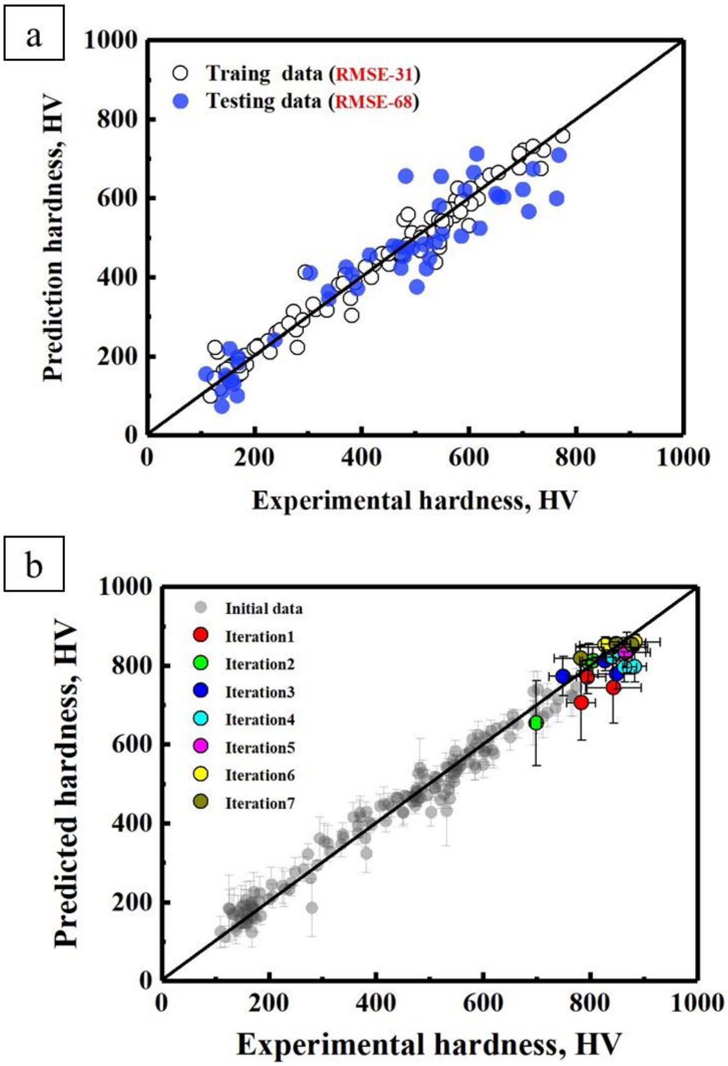


Figure 4.1: Plots from Wen et al [49]. (a) Comparison between the predicted and the experimental hardness values of the HEAs in both training and test dataset. (b) The predicted hardness values versus the measured values for the alloys of the original 155 training data and those synthesized in successive seven rounds of iterations.

4.3 Strength and ductility

The strength and ductility of alloys can be influenced by many factors. Intrinsically, stronger atomic bonding corresponds to the higher strength of a single lattice. Furthermore, factors such as the phase, grain boundary segregation, and grain size can all influence the strength and ductility of HEAs.

Many mechanical properties prediction models are based on the assumption of single-phase formation, i.e., the HEA will be in either a single FCC or BCC phase. In a recent important study, Hu et al. [155] reveals the correlation between the D-parameter and ductility. D-parameter is defined as $\frac{\gamma_{\mathrm{surf}}}{\gamma_{\mathrm{usf}}}$, where γ_{surf} is the surface energy and γ_{usf} is the unstable stacking fault (USF) energy. Alloys with high D parameter values tend to be intrinsically ductile. The calculation of γ_{surf} and γ_{usf} are accomplished by DFT, which is tedious and time-consuming. Alternatively, they developed a ML regression model and achieved high consistency between the ML predicted values and DFT calculation results. Finally, they showed that the D parameter strongly correlates to the compression fracture strain of HEAs from the experimental results.

As shown in Figure 4.2A-C, values of γ_{usf}, γ_{surf}, and D parameters calculated from DFT are plotted against valence electron concentration (VEC) of multiple binary refractory element systems. It is interesting to note that all three curves have parabolic shapes with extreme values around VEC = 6. The D parameter curves indicate that either low or high VEC values are needed to form intrinsically ductile BCC HEAs. Figure 4.2D-E shows the great match of γ_{usf} and γ_{surf} values between ML prediction and DFT calculation. Thus, by extrapolating from the DFT calculated values of 106 binary, ternary, and quaternary BCC alloys, ML was able to calculate D parameters for higher-order HEA systems. Figure 4.2F shows the γ_{usf} values against D parameters. USF energy indicates the material's ability to resist deformation, while D parameter indicates the ductility. Figure 4.2F reproduces the typical strength-ductility reciprocal relationship in alloys. Finally, Figure 4.2G shows the alloy fracture strain experimental results against the D parameter values. Generally, larger D parameter values correspond to greater fracture strains. Overall, this work provides valuable guidance in searching for ductile HEAs. The D parameter can be used in the future work to give exact value prediction on HEA mechanical properties.

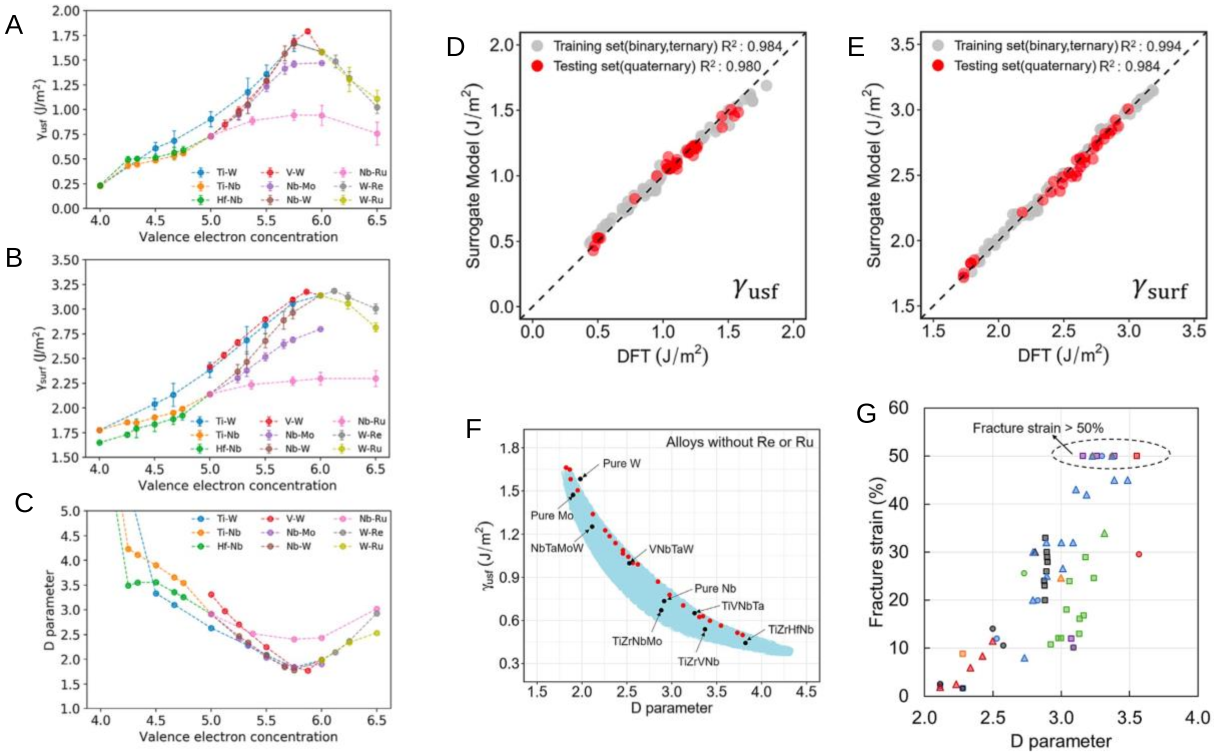


Figure 4.2: (A-C) Values of γ_{usf}, γ_{surf}, and D parameters against valence electron concentration of multiple binary refractory element systems. (D-E) Comparison of γ_{usf} and γ_{surf} values from ML prediction and DFT calculation. (F) γ_{usf} values against D parameters, indicating the strength against the ductility of alloys. (G) Experiment results of fracture strain against the D parameter values of alloys. Figure from Hu et al. [155]

ML has been widely applied to directly predict the strength and ductility of conventional alloys. For example, Dey et al. [156] used the ML ANN algorithm, inputting alloy compositions and processing parameters, to obtain the Pareto solutions for aluminum alloys with optimized strength and ductility. Similarly, Mohanty et al. [157] applied the ANN algorithm with the input of compositions and processing parameters for cold rolled interstitial free (IF) steel sheets, and successfully predicted the yield strength, tensile strength, and % elongation. It is expected that many well-developed machine learning models will be developed in the coming years to predict the strength and ductility of HEAs.

In addition to the intrinsic factors that affect ductility and strength, we will discuss other factors as well. Firstly, properties prediction models normally assume a single SS formation, whereas most HEAs form composite phases. The precipitation of Laves and Sigma can easily embrittle the alloy[158]. On the other hand, grain size and morphology are also important. Alloys with small grains and fine precipitation of IM can maintain ductility and be significantly strengthened[159]. Furthermore, grain boundary segregation, a phenomenon in which elements such as oxygen accumulate at grain boundaries, can embrittle the alloy[160]. It is important to consider all these aspects when optimizing the mechanical properties of alloys.

Chapter 5 High-throughput High Entropy Alloys Design Model

We are now able to solve the problem of finding suitable HEAs from a vast compositional space with the phase and property prediction models. The structure of the high-throughput HEA design model is in Figure 5.1.

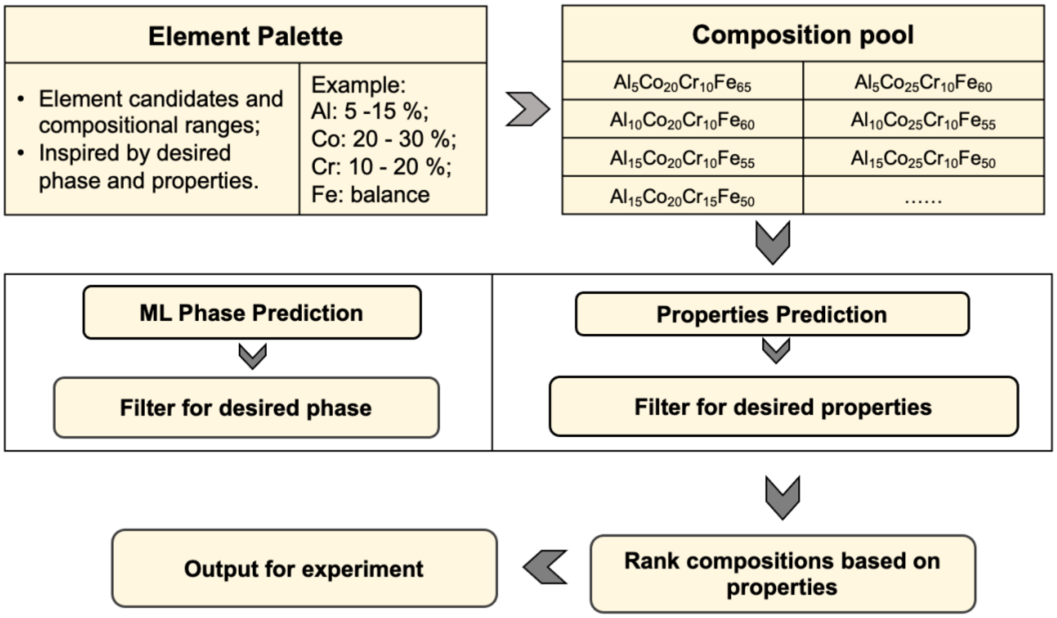


Figure 5.1: Structure of the high-throughput HEA design model.

According to the alloy design requirements, the metallurgist should first construct the element palette and input the desired phase and properties (density, melting temperature, cost, strain, strength, etc.) into the program. The program generates all possible compositions based on a grid search. ML models for predicting phases and properties can then be applied to screen out some compositions based on whether they meet the phase and property requirements. The selected compositions will then be scored according to their properties, and the compositions with the highest scores will be sent for experimental validation.

Here is an example of how to set up the input to the program of the high-temperature functional material design. Firstly, Al-containing refractory HEAs (Al-RHEA)[15] can be the ideal candidate because they generally have high melting temperatures, and the Al-X-Y B2 phase formed in Al-RHEA can help maintain mechanical properties at elevated temperatures. The element palette will include elements such as Al, Ti, Cr, Zr, Nb, Hf, V, etc. The desired microstructure should include the Al-X-Y B2 phase and exclude Laves and Sigma phases as they can embrittle the alloys. The desired properties can include melting temperature > 2000 ℃ and compression strain > 30 % to guarantee certain compression ductility.

The main advantage of this method is its efficiency, i.e., its ability to search a vast space of compositional possibilities within a short computation time. However, certain limitations still exist. As discussed in Section 2.5, ML phase prediction doesn't provide information on phase volume fractions, transformation temperatures, and element distribution among phases. This information is crucial in optimizing the mechanical and corrosion resistance properties and should be obtained from CALPHAD in more detailed alloy optimization. In addition, many properties prediction models are developed based on assumptions that HEA forms single FCC or BCC phases, and their predicted values may be inaccurate if HEA forms a composite phase. In spite of this, the predictions of these properties are still helpful because they provide a comparison among the compositions.

Chapter 6 High Entropy Alloys Synthesis and Characterization

6.1 Overview

This chapter will describe the various experimental methods in alloy synthesis and characterization. After that, the experiment results of a series of HEAs designed for the marine environment will be discussed.

6.2 Experiment methods

6.2.1 Alloy synthesis

Alloys are synthesized from pure elements with high purities (normally > 99.9 wt. %). Arc melting is our primary tool to melt pure elements into alloys. A schematic diagram of the arc melting is provided in Figure 6.1A. The raw materials (elements) should be placed properly before closing the chamber. Generally, light materials should be covered by heavy materials in the crucible. Liquid light and heavy materials will move up and down in the liquid state, improving the mixing of the elements.

At first, the arc melting chamber is vacuumed to 300 mTorr. Then, the chamber will be backfilled with high purity argon gas and vacuumed again to 500 mTorr. This backfilling operation will be repeated at least three times to decrease the oxygen percentage in the residual gas inside the chamber. The chamber will eventually be vacuumed to 30 mTorr and then refilled with argon to 15 mBar in order to sustain the electric arc. Oxygen-gathering material such as Zr or Ti should be melted before the alloy to eliminate residual oxygen. The alloy will be melted at least three times. Each melt lasts for at least 30 seconds, and the alloy should be flipped between two melts. It is recommended to cool down the system after melting for several minutes before opening the chamber.

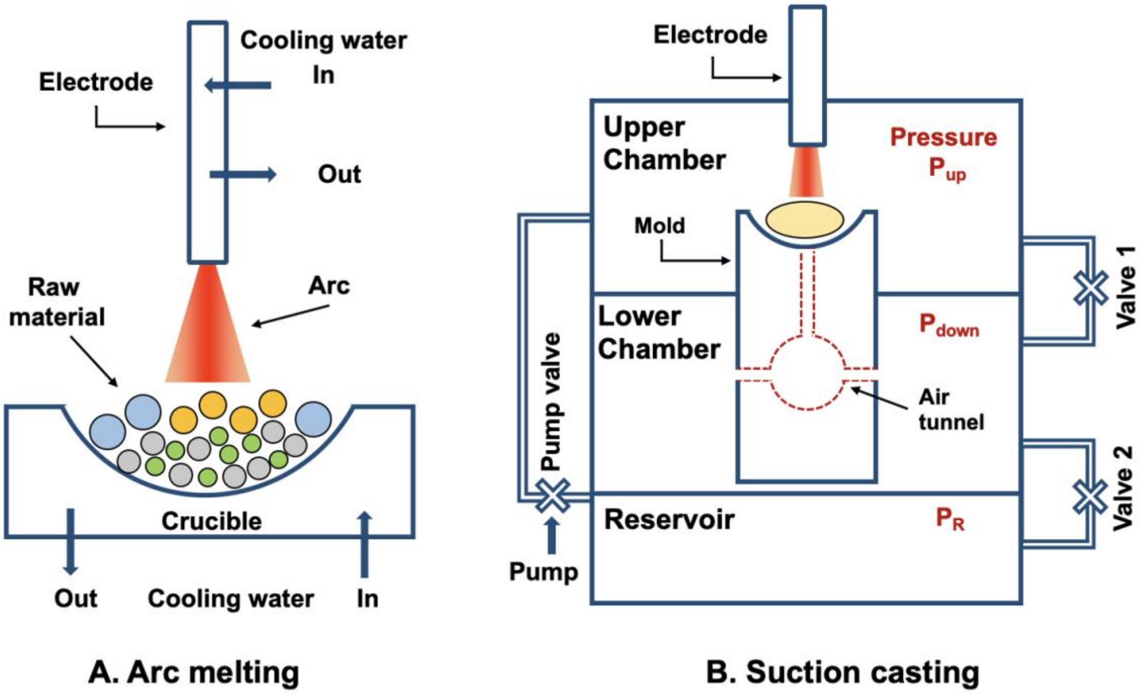


Figure 6.1: Schematic diagrams of: A - Arc melting, and B – Suction casting systems. The colored circles in figure A represent raw materials in the crucible, while their sizes indicate the densities of the materials. Materials of low density are covered by materials of high density. The dashed line in figure B represents the air tunnel inside the mold. Alloy will be sucked into the tunnel and solidified into the desired shape.

Suction casting can provide fast cooling and cast alloys into the desired shapes. As shown in Figure 6.1B, the suction casting instrument includes three chambers. The upper and lower chambers are connected with a highly heat-conductive copper mold. Inside the mold is an air tunnel. Liquid alloy will then be sucked into the tunnel and solidified into the desired shape. Valves 1 and 2 control the connections between the three chambers.

Valve 1, 2, and the pump valve are initially opened while the three chambers are vacuumed and backfilled three times following the same procedure in the arc melting. After closing the pump valve and valve 2, the reservoir is isolated. The upper and lower chambers are still connected and will be backfilled with argon gas to the atmospheric pressure (1 atm) to sustain the arc during melting. At this stage, $P_{up} = P_{down}$ = 1 atm and P_R = 30 mTorr. Next, the alloy is melted, with valve 1 being closed and valve 2 being opened. The P_{down} drops immediately when the lower chamber is isolated from the upper chamber but connected to the reservoir. The pressure difference between

P_{up} and P_{down} will suck the alloy into the air tunnel. The solidification finishes in a brief time due to the high heat conductivity of the copper mold.

6.2.2 Encapsulation, annealing and quenching

Annealing is conducted in a furnace with samples encapsulated into quartz tubes.

Figure 6.2 is the schematic diagram for the process of sample encapsulation. The alloy and a piece of Ta foil are placed inside the tube. Ta foil can gather oxygen during annealing. The high melting temperature of Ta (3020℃) guarantees it will not contaminate the alloy under a regular annealing temperature below 1100 ℃. The flame softens the tube and a narrow neck forms under the pulling force. The tube is then vacuumed and backfilled with argon gas three times, following the same procedure in the arc melting. Once the desired vacuum level is reached, the user can seal the tube either with a high vacuum inside or after backfilling a certain amount of argon. Finally, by heating the narrow neck again and pulling the tube, the upper and lower portions disconnect, and the lower portion is completely sealed.

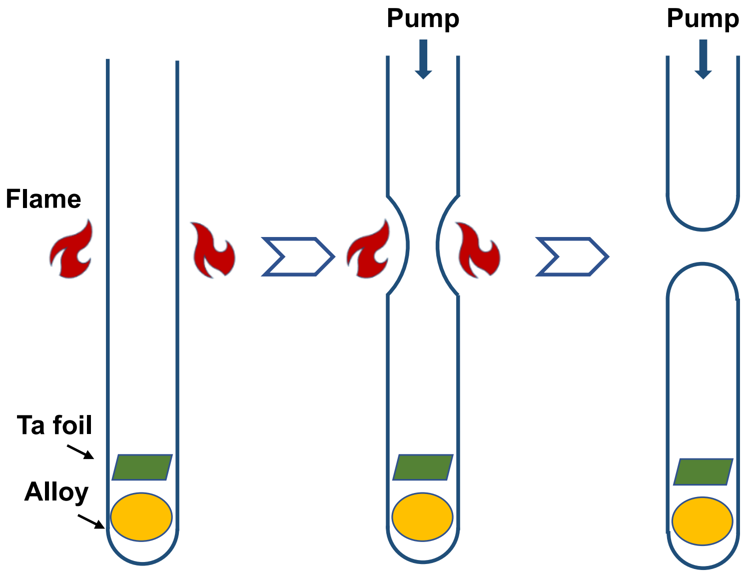


Figure 6.2: Schematic diagrams of the process of sample encapsulation.

After the annealing, water quenching is normally required to retain the homogenized phase at the annealing temperature due to its high cooling rate. The quartz tube should be quickly removed from the furnace and dipped into a bucket of water. It is important to break the tube inside the water in order to quench the alloy directly by the water.

6.3 HEA designed for marine environment

We designed a series of HEAs for the marine environment. The alloy requirements are low density, low cost, high strength, ductility, and good corrosion resistance. Low-cost elements Fe, Mn, and Al can be used as much as possible. Corrosion-resistant elements such as Cr, Mo, and Ti should be added. Low density can be achieved by including Al and Ti. Ductility and strength can be achieved by forming composite FCC+B2/Heusler. FCC phase guarantees ductility, while the B2 and Heusler precipitation can strengthen the alloys. FCC stabilizing elements are Co, Ni, and Mn. The addition of Al and Ti can induce B2 or Heusler phase. The element palette was constructed as Al, Co, Cr, Fe, Mn, Mo, Ni, and Ti (Table 6.1). Desired phases are FCC, B2, and Heusler, rejecting Laves and Sigma phases as they will embrittle the alloys. At the same time, we will control the density below 7.5 g/cc and cost below 5.5 U.S. dollars/kg.

Table 6.1: The initial element palette and grid search step size in marine environment HEA design.

Element	Al	Co	Cr	Fe	Mn	Mo	Ni	Ti
Scanning range (at. %)	3 - 15	0 - 20	5 - 15	20 - 50	1 - 10	1 - 6	10 - 40	0 - 60 (balance)
Scanning Step size (at. %)	3	5	5	5	2	2	5	balance

Using the method introduced in Chapter 5, four HEAs with varying Mn contents were designed[21]. Their compositions with the labels are listed in Table 6.2. Figure 6.3A and C are the XRD patterns and SEM images of the four HEAs. Their phases are FCC+B2, where fine B2 IM precipitates (dark particles in Figure 6.3C) are distributed evenly in the FCC matrix (light background in Figure 6.3C). This morphology can potentially strengthen the alloys and maintain the ductility. Figure 6.3B shows the corrosion resistance of the HEAs compared to the two benchmarks, HEA-Cantor and 316L. In particular, HEA-0.25Mn shows better corrosion resistance than other HEAs and the benchmarks, based on the evidence of lower current density in the passive potential domain and higher pitting potential.

Table 6.2: Compositions of HEAs designed for marine environment. Cantor alloy and 316L stainless steel are the benchmarks for corrosion resistance study.

Composition	Label
$Al_{0.3}Cr_{0.5}Fe_2Mn_0Mo_{0.15}Ni_{1.5}Ti_{0.3}$	HEA-0Mn
$Al_{0.3}Cr_{0.5}Fe_2Mn_{0.25}Mo_{0.15}Ni_{1.5}Ti_{0.3}$	HEA-0.25Mn
$Al_{0.3}Cr_{0.5}Fe_2Mn_{0.5}Mo_{0.15}Ni_{1.5}Ti_{0.3}$	HEA-0.5Mn
$Al_{0.3}Cr_{0.5}Fe_2Mn_1Mo_{0.15}Ni_{1.5}Ti_{0.3}$	HEA-1Mn
CoCrFeMnNi	HEA-Cantor
316L stainless steel	316L

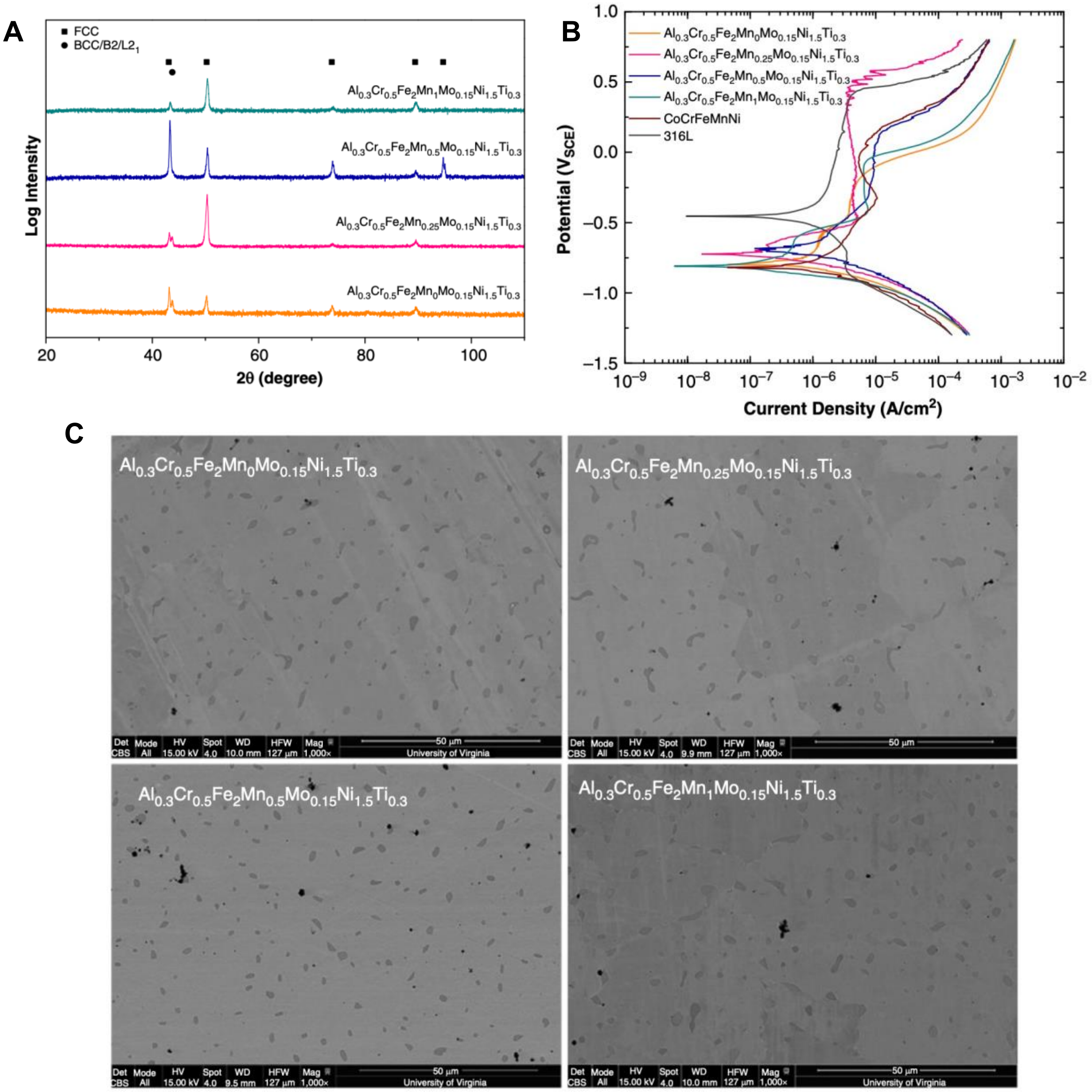


Figure 6.3: (A) X-ray diffraction (XRD) patterns of the HEAs designed for marine environment. (B) E-log(i) curves obtained during upward polarization of CCAs in 0.01 M NaCl after a 600 s application of a −1.3 V_{SCE} potential compared to CoCrFeMnNi and 316L. (C) Scanning electron microscope (SEM) images of the HEAs. Figures from Inman et al. [21]

Chapter 7 Summary and outlook

This thesis introduces high entropy alloys (HEA) phase and properties prediction methods, high-throughput HEA design model, and design results. We reviewed three types of phase prediction methods using single parameters, thermodynamic and first-principles calculations, and machine learning (ML). Based on the limitations of the current phase prediction methods, we developed ML models using phase diagram-based phenomenological parameters and feature engineering techniques. As a result, we achieved high accuracies (80 % ~ 90 %) in predicting the common solid solution (SS) and intermetallic (IM) phase and also interpreted the physical factors controlling each phase's formation. Based on the phase and properties prediction methods, we built a high-throughput HEA design model which can automatically screen all compositions and find the ones with desired phase and properties. A series of HEAs were finally designed for the marine environment. The experiment results validate the effectiveness of this model.

In the future, we expect to see the following developments in HEA research. Firstly, integrated phase and properties models with the synergetic use of ML, CALPHAD, and first-principles calculations will be applied to alloy design. As a result, they can rapidly locate the optimal compositions, and perform detailed and comprehensive analyses or optimization of phase and properties. Secondly, more HEA properties can be predicted accurately, especially using ML. ML was first applied to HEA phase prediction in 2016 and has matured recently as a result of the growing database on HEA phase. There will be a similar growth trend for ML HEA property prediction models. Lastly, the development of high-throughput experiments will expand the database and study the HEAs in compositional space that have seldom been explored. Therefore, more HEAs with exceptional properties will be discovered, pushing the limits of current materials.

Reference

1. Cantor, B., Chang, I. T. H., Knight, P. & Vincent, A. J. B. Microstructural development in equiatomic multicomponent alloys. *Mater. Sci. Eng. A* **375**–**377**, 213–218 (2004).
2. Yeh, J. W. *et al.* Nanostructured high-entropy alloys with multiple principal elements: Novel alloy design concepts and outcomes. *Adv. Eng. Mater.* **6**, 299-303+274 (2004).
3. Miracle, D. B. & Senkov, O. N. A critical review of high entropy alloys and related concepts. *Acta Mater.* **122**, 448–511 (2017).
4. Yeh, J. W. Recent progress in high-entropy alloys. *Ann. Chim. Sci. des Mater.* **31**, 633–648 (2006).
5. Graf, T., Felser, C. & Parkin, S. S. P. Simple rules for the understanding of Heusler compounds. *Prog. Solid State Chem.* **39**, 1–50 (2011).
6. Wolff-Goodrich, S. *et al.* Combinatorial exploration of B2/L21 precipitation strengthened AlCrFeNiTi compositionally complex alloys. *J. Alloys Compd.* **853**, 156111 (2021).
7. Inman, S. *et al.* Effect of Mn Content on the Passivation and Corrosion of Al0.3Cr0.5Fe2MnxMo0.15Ni1.5Ti0.3 FCC Compositionally Complex Alloys. *Corrosion* (2021) doi:10.5006/3906.
8. Feng, R. *et al.* High-throughput design of high-performance lightweight high-entropy alloys. *Nat. Commun. 2021 121* **12**, 1–10 (2021).
9. Shi, P. *et al.* Enhanced strength–ductility synergy in ultrafine-grained eutectic high-entropy alloys by inheriting microstructural lamellae. *Nat. Commun.* **10**, 1–8 (2019).
10. Feng, R. *et al.* Enhancing fatigue life by ductile-transformable multicomponent B2 precipitates in a high-entropy alloy. *Nat. Commun. 2021 121* **12**, 1–10 (2021).
11. Polvani, R. S., Tzeng, W. S. & Strutt, P. R. High temperature creep in a semi-coherent NiAl-Ni2AlTi alloy. *Metall. Trans. A* **7**, 33–40 (1976).
12. Strutt, P. R., Polvani, R. S. & Ingram, J. C. Creep behavior of the heusler type structure alloy Ni2AlTi. *Metall. Trans. A* **7**, 23–31 (1976).
13. Li, C. *et al.* Microstructures and mechanical properties of body-centered-cubic (Al,Ti)0.7(Ni,Co,Fe,Cr)5 high entropy alloys with coherent B2/L21 nanoprecipitation.

Mater. Sci. Eng. A **737**, 286–296 (2018).

14. Qi, Y. *et al.* L21-strengthened face-centered cubic high-entropy alloy with high strength and ductility. *Mater. Sci. Eng. A* **797**, 140056 (2020).
15. Miracle, D. B., Tsai, M. H., Senkov, O. N., Soni, V. & Banerjee, R. Refractory high entropy superalloys (RSAs). *Scr. Mater.* **187**, 445–452 (2020).
16. Qin, G. *et al.* CoCrFeMnNi high-entropy alloys reinforced with Laves phase by adding Nb and Ti elements. *J. Mater. Res.* **34**, 1011–1020 (2019).
17. Soni, V. K., Sanyal, S., Rao, K. R. & Sinha, S. K. A review on phase prediction in high entropy alloys. *Proc. Inst. Mech. Eng. Part C J. Mech. Eng. Sci.* **235**, 6268–6286 (2021).
18. Stepanov, N. D. *et al.* Effect of v content on microstructure and mechanical properties of the CoCrFeMnNiVx high entropy alloys. *J. Alloys Compd.* **628**, 170–185 (2015).
19. Hsu, C. Y., Wang, W. R., Tang, W. Y., Chen, S. K. & Yeh, J. W. Microstructure and mechanical properties of new AlCoxCrFeMo 0.5Ni high-entropy alloys. *Adv. Eng. Mater.* **12**, 44–49 (2010).
20. Gludovatz, B. *et al.* Exceptional damage-tolerance of a medium-entropy alloy CrCoNi at cryogenic temperatures. *Nat. Commun. 2016 71* **7**, 1–8 (2016).
21. Inman, S. B. *et al.* Effect of Mn Content on the Passivation and Corrosion of Al0.3Cr0.5Fe2MnxMo0.15Ni1.5Ti0.3 Compositionally Complex Face-Centered Cubic Alloys. *Corrosion* **78**, 32–48 (2022).
22. Senkov, O. N., Miracle, D. B., Chaput, K. J. & Couzinie, J. P. Development and exploration of refractory high entropy alloys - A review. *J. Mater. Res.* **33**, 3092–3128 (2018).
23. Lu, P. *et al.* Computational design and initial corrosion assessment of a series of non-equimolar high entropy alloys. *Scr. Mater.* **172**, 12–16 (2019).
24. Egami, T., Guo, W., Rack, P. D. & Nagase, T. Irradiation resistance of multicomponent alloys. *Metall. Mater. Trans. A Phys. Metall. Mater. Sci.* **45**, 180–183 (2014).
25. Chou, H. P., Chang, Y. S., Chen, S. K. & Yeh, J. W. Microstructure, thermophysical and electrical properties in AlxCoCrFeNi ($0 \leq x \leq 2$) high-entropy alloys. *Mater. Sci. Eng. B* **163**, 184–189 (2009).
26. Zhang, K. & Fu, Z. Effects of annealing treatment on properties of CoCrFeNiTiAlx multi-component alloys. *Intermetallics* **28**, 34–39 (2012).
27. Kao, Y. F. *et al.* Electrical, magnetic, and Hall properties of AlxCoCrFeNi high-entropy

alloys. *J. Alloys Compd.* **509**, 1607–1614 (2011).

28. Perrin, A., Sorescu, M., Burton, M. T., Laughlin, D. E. & McHenry, M. The Role of Compositional Tuning of the Distributed Exchange on Magnetocaloric Properties of High-Entropy Alloys. *Jom* **69**, 2125–2129 (2017).
29. Kurniawan, M., Perrin, A., Xu, P., Keylin, V. & McHenry, M. Curie Temperature Engineering in High Entropy Alloys for Magnetocaloric Applications. *IEEE Magn. Lett.* **7**, (2016).
30. Koželj, P. *et al.* Discovery of a superconducting high-entropy alloy. *Phys. Rev. Lett.* **113**, 107001 (2014).
31. Kao, Y. F. *et al.* Hydrogen storage properties of multi-principal-component CoFeMnTixVyZrz alloys. *Int. J. Hydrogen Energy* **35**, 9046–9059 (2010).
32. Kunce, I., Polanski, M. & Bystrzycki, J. Structure and hydrogen storage properties of a high entropy ZrTiVCrFeNi alloy synthesized using Laser Engineered Net Shaping (LENS). *Int. J. Hydrogen Energy* **38**, 12180–12189 (2013).
33. Kunce, I., Polanski, M. & Bystrzycki, J. Microstructure and hydrogen storage properties of a TiZrNbMoV high entropy alloy synthesized using Laser Engineered Net Shaping (LENS). *Int. J. Hydrogen Energy* **39**, 9904–9910 (2014).
34. Yao, Y. *et al.* Carbothermal shock synthesis of high-entropy-alloy nanoparticles. *Science (80-.).* **359**, 1489–1494 (2018).
35. Green, M. L., Takeuchi, I. & Hattrick-Simpers, J. R. Applications of high throughput (combinatorial) methodologies to electronic, magnetic, optical, and energy-related materials. *J. Appl. Phys.* **113**, 231101 (2013).
36. Liu, R. *et al.* Entropy as a Gene-Like Performance Indicator Promoting Thermoelectric Materials. *Adv. Mater.* **29**, 1–7 (2017).
37. Poon, S. J. & He, J. Multi-Principal-Element Approach to High-Performance Thermoelectric Materials. *Ref. Modul. Mater. Sci. Mater. Eng.* (2019) doi:10.1016/B978-0-12-803581-8.11719-9.
38. Poon, S. J., Qi, J. & Cheung, A. M. Harnessing the Complex Compositional Space of High-Entropy Alloys. in *High-Entropy Materials: Theory, Experiments, and Applications* (eds. Jamieson Brechtl & Peter K. Liaw) 63–113 (Springer, Cham, 2021). doi:10.1007/978-3-030-77641-1_3.

39. Gao, M. C. *et al.* Thermodynamics of concentrated solid solution alloys. *Curr. Opin. Solid State Mater. Sci.* **21**, 238–251 (2017).
40. Toda-Caraballo, I. & Rivera-Díaz-Del-Castillo, P. E. J. A criterion for the formation of high entropy alloys based on lattice distortion. *Intermetallics* **71**, 76–87 (2016).
41. Ikeda, Y., Grabowski, B. & Körmann, F. Ab initio phase stabilities and mechanical properties of multicomponent alloys: A comprehensive review for high entropy alloys and compositionally complex alloys. *Mater. Charact.* **147**, 464–511 (2019).
42. Senkov, O. N., Miller, J. D., Miracle, D. B. & Woodward, C. Accelerated exploration of multi-principal element alloys for structural applications. *Calphad Comput. Coupling Phase Diagrams Thermochem.* **50**, 32–48 (2015).
43. Senkov, O. N. *et al.* CALPHAD-aided development of quaternary multi-principal element refractory alloys based on NbTiZr. *J. Alloys Compd.* **783**, 729–742 (2019).
44. Santodonato, L. J., Liaw, P. K., Unocic, R. R., Bei, H. & Morris, J. R. Predictive multiphase evolution in Al-containing high-entropy alloys. *Nat. Commun.* **9**, 1–10 (2018).
45. Tancret, F., Toda-Caraballo, I., Menou, E. & Rivera Díaz-Del-Castillo, P. E. J. Designing high entropy alloys employing thermodynamics and Gaussian process statistical analysis. *Mater. Des.* **115**, 486–497 (2017).
46. Kube, S. A. *et al.* Phase selection motifs in High Entropy Alloys revealed through combinatorial methods: Large atomic size difference favors BCC over FCC. *Acta Mater.* **166**, 677–686 (2019).
47. Ruiz-Yi, B. *et al.* The Different Roles of Entropy and Solubility in High Entropy Alloy Stability. *ACS Comb. Sci.* **18**, 596–603 (2016).
48. Chang, Y.-J. J., Jui, C.-Y. Y., Lee, W.-J. J. & Yeh, A.-C. C. Prediction of the Composition and Hardness of High-Entropy Alloys by Machine Learning. *Jom* **71**, 3433–3442 (2019).
49. Wen, C. *et al.* Machine learning assisted design of high entropy alloys with desired property. *Acta Mater.* **170**, 109–117 (2019).
50. Islam, N., Huang, W. & Zhuang, H. L. Machine learning for phase selection in multi-principal element alloys. *Comput. Mater. Sci.* **150**, 230–235 (2018).
51. Huang, W., Martin, P. & Zhuang, H. L. Machine-learning phase prediction of high-entropy alloys. *Acta Mater.* **169**, 225–236 (2019).
52. Li, Y. & Guo, W. Machine-learning model for predicting phase formations of high-entropy

alloys. *Phys. Rev. Mater.* **3**, 95005 (2019).

53. Zhou, Z. *et al.* Machine learning guided appraisal and exploration of phase design for high entropy alloys. *npj Comput. Mater.* **5**, 1–9 (2019).
54. Agarwal, A. & Prasada Rao, A. K. Artificial Intelligence Predicts Body-Centered-Cubic and Face-Centered-Cubic Phases in High-Entropy Alloys. *Jom* **71**, 3424–3432 (2019).
55. Qi, J., Cheung, A. M. & Poon, S. J. High Entropy Alloys Mined From Binary Phase Diagrams. *Sci. Rep.* **9**, 1–10 (2019).
56. Yeh, J. W. *et al.* Nanostructured high-entropy alloys with multiple principal elements: Novel alloy design concepts and outcomes. *Adv. Eng. Mater.* **6**, 299–303 (2004).
57. Zhang, Y., Zhou, Y. J., Lin, J. P., Chen, G. L. & Liaw, P. K. Solid-solution phase formation rules for multi-component alloys. *Adv. Eng. Mater.* **10**, 534–538 (2008).
58. Takeuchi, A. & Inoue, A. Classification of Bulk Metallic Glasses by Atomic Size Difference, Heat of Mixing and Period of Constituent Elements and Its Application to Characterization of the Main Alloying Element. *Mater. Trans.* **46**, 2817–2829 (2005).
59. Zhang, Y. *et al.* Guidelines in predicting phase formation of high-entropy alloys. *MRS Commun.* **4**, 57–62 (2014).
60. Ye, Y. F., Wang, Q., Lu, J., Liu, C. T. & Yang, Y. Design of high entropy alloys: A single-parameter thermodynamic rule. *Scr. Mater.* **104**, 53–55 (2015).
61. Ye, Y. F., Wang, Q., Lu, J., Liu, C. T. & Yang, Y. The generalized thermodynamic rule for phase selection in multicomponent alloys. *Intermetallics* **59**, 75–80 (2015).
62. King, D. J. M., Middleburgh, S. C., McGregor, A. G. & Cortie, M. B. Predicting the formation and stability of single phase high-entropy alloys. *Acta Mater.* **104**, 172–179 (2016).
63. Troparevsky, M. C., Morris, J. R., Kent, P. R. C., Lupini, A. R. & Stocks, G. M. Criteria for predicting the formation of single-phase high-entropy alloys. *Phys. Rev. X* **5**, 1–6 (2015).
64. Senkov, O. N. & Miracle, D. B. A new thermodynamic parameter to predict formation of solid solution or intermetallic phases in high entropy alloys. *J. Alloys Compd.* **658**, 603–607 (2016).
65. Pei, Z., Yin, J., Hawk, J. A., Alman, D. E. & Gao, M. C. Machine-learning informed prediction of high-entropy solid solution formation: Beyond the Hume-Rothery rules. *npj Comput. Mater.* **6**, 1–8 (2020).

66. Kanzaki, H. Point defects in face-centred cubic lattice-I distortion around defects. *J. Phys. Chem. Solids* **2**, 24–36 (1957).

67. Rudman, P. S., Stringer, J. & Jaffee, R. I. *Phase Stability in Metals and Alloys: Papers and Proceedings*. (McGraw-Hill, 1967).

68. Ye, Y. F., Liu, C. T. & Yang, Y. A geometric model for intrinsic residual strain and phase stability in high entropy alloys. *Acta Mater.* **94**, 152–161 (2015).

69. Wang, Z., Qiu, W., Yang, Y. & Liu, C. T. Atomic-size and lattice-distortion effects in newly developed high-entropy alloys with multiple principal elements. *Intermetallics* **64**, 63–69 (2015).

70. Toda-Caraballo, I., Wróbel, J. S., Dudarev, S. L., Nguyen-Manh, D. & Rivera-Díaz-Del-Castillo, P. E. J. Interatomic spacing distribution in multicomponent alloys. *Acta Mater.* **97**, 156–169 (2015).

71. Fang, S., Xiao, X., Xia, L., Li, W. & Dong, Y. Relationship between the widths of supercooled liquid regions and bond parameters of Mg-based bulk metallic glasses. *J. Non. Cryst. Solids* **321**, 120–125 (2003).

72. Dong, Y., Lu, Y., Jiang, L., Wang, T. & Li, T. Effects of electro-negativity on the stability of topologically close-packed phase in high entropy alloys. *Intermetallics* **52**, 105–109 (2014).

73. Guo, S., Ng, C., Lu, J. & Liu, C. T. Effect of valence electron concentration on stability of fcc or bcc phase in high entropy alloys. *J. Appl. Phys.* **109**, (2011).

74. Yang, S., Lu, J., Xing, F., Zhang, L. & Zhong, Y. Revisit the VEC rule in high entropy alloys (HEAs) with high-throughput CALPHAD approach and its applications for material design-A case study with Al-Co-Cr-Fe-Ni system. *Acta Mater.* **192**, 11–19 (2020).

75. Tsai, M. H. *et al.* Criterion for sigma phase formation in Cr- and V-Containing high-entropy alloys. *Mater. Res. Lett.* **1**, 207–212 (2013).

76. Lu, Y. *et al.* A criterion for topological close-packed phase formation in high entropy alloys. *Entropy* **17**, 2355–2366 (2015).

77. Poletti, M. G. & Battezzati, L. Electronic and thermodynamic criteria for the occurrence of high entropy alloys in metallic systems. *Acta Mater.* **75**, 297–306 (2014).

78. Chen, H. L., Mao, H. & Chen, Q. Database development and Calphad calculations for high entropy alloys: Challenges, strategies, and tips. *Mater. Chem. Phys.* **210**, 279–290 (2018).

79. Mallik, A. K. Computer calculations of phase diagrams. *Bull. Mater. Sci.* **8**, 107–121 (1986).

80. Andersson, J. O., Helander, T., Höglund, L., Shi, P. & Sundman, B. Thermo-Calc & DICTRA, computational tools for materials science. *Calphad Comput. Coupling Phase Diagrams Thermochem.* **26**, 273–312 (2002).

81. Gorsse, S. & Senkov, O. N. About the reliability of CALPHAD predictions in multicomponent systems. *Entropy* **20**, 1–9 (2018).

82. Chou, K.-C. & Austin Chang, Y. A Study of Ternary Geometrical Models. *Berichte der Bunsengesellschaft für Phys. Chemie* **93**, 735–741 (1989).

83. Wu, M., Wang, S., Huang, H., Shu, D. & Sun, B. CALPHAD aided eutectic high-entropy alloy design. *Mater. Lett.* **262**, 127175 (2020).

84. Abu-Odeh, A. *et al.* Efficient exploration of the High Entropy Alloy composition-phase space. *Acta Mater.* **152**, 41–57 (2018).

85. Ge, S. *et al.* Effects of Al addition on the microstructures and properties of MoNbTaTiV refractory high entropy alloy. *Mater. Sci. Eng. A* **784**, 139275 (2020).

86. Feng, R. *et al.* Phase stability and transformation in a light-weight high-entropy alloy. *Acta Mater.* **146**, 280–293 (2018).

87. Gao, M. & Alman, D. Searching for Next Single-Phase High-Entropy Alloy Compositions. *Entropy* **15**, 4504–4519 (2013).

88. Gao, M. C., Zhang, B., Guo, S. M., Qiao, J. W. & Hawk, J. A. High-Entropy Alloys in Hexagonal Close-Packed Structure. *Metall. Mater. Trans. A Phys. Metall. Mater. Sci.* **47**, 3322–3332 (2016).

89. Gao, M. C., Zhang, B., Yang, S. & Guo, S. M. Senary Refractory High-Entropy Alloy HfNbTaTiVZr. *Metall. Mater. Trans. A Phys. Metall. Mater. Sci.* **47**, 3333–3345 (2016).

90. Senkov, O. N., Miller, J. D., Miracle, D. B. & Woodward, C. Accelerated exploration of multi-principal element alloys with solid solution phases. *Nat. Commun.* **6**, 1–10 (2015).

91. Gorsse, S. & Tancret, F. Current and emerging practices of CALPHAD toward the development of high entropy alloys and complex concentrated alloys. *J. Mater. Res.* **33**, 2899–2923 (2018).

92. Guruvidyathri, K., Kumar, K. C. H., Yeh, J. W. & Murty, B. S. Topologically Close-packed Phase Formation in High Entropy Alloys: A Review of Calphad and Experimental Results. doi:10.1007/s11837-017-2566-5.

93. Yin, B. & Curtin, W. A. First-principles-based prediction of yield strength in the RhIrPdPtNiCu high-entropy alloy. *npj Comput. Mater.* **5**, 1–7 (2019).

94. Yu, P. *et al.* The influence of dilute aluminum and molybdenum on stacking fault and twin formation in FeNiCoCr-based high entropy alloys based on density functional theory. *Sci. RepoRtS* | **9**, 10940 (2019).

95. Yin, B., Maresca, F. & Curtin, W. A. Vanadium is an optimal element for strengthening in both fcc and bcc high-entropy alloys. *Acta Mater.* **188**, 486–491 (2020).

96. King, D. J. M. *et al.* High temperature, low neutron cross-section high-entropy alloys in the Nb-Ti-V-Zr system. *Acta Mater.* **166**, 435–446 (2019).

97. Singh, P., Smirnov, A. V., Alam, A. & Johnson, D. D. First-principles prediction of incipient order in arbitrary high-entropy alloys: exemplified in Ti0.25CrFeNiAlx. *Acta Mater.* **189**, 248–254 (2020).

98. Jafary-Zadeh, M., Khoo, K. H., Laskowski, R., Branicio, P. S. & Shapeev, A. V. Applying a machine learning interatomic potential to unravel the effects of local lattice distortion on the elastic properties of multi-principal element alloys. *J. Alloys Compd.* **803**, 1054–1062 (2019).

99. Kostiuchenko, T., Körmann, F., Neugebauer, J. & Shapeev, A. Impact of lattice relaxations on phase transitions in a high-entropy alloy studied by machine-learning potentials. *npj Comput. Mater.* **5**, 1–7 (2019).

100. Beyramali Kivy, M. & Asle Zaeem, M. Generalized stacking fault energies, ductilities, and twinnabilities of CoCrFeNi-based face-centered cubic high entropy alloys. *Scr. Mater.* **139**, 83–86 (2017).

101. Rao, S. I. *et al.* Modeling solution hardening in BCC refractory complex concentrated alloys: NbTiZr, Nb1.5TiZr0.5 and Nb0.5TiZr1.5. *Acta Mater.* **168**, 222–236 (2019).

102. Zaddach, A. J., Niu, C., Koch, C. C. & Irving, D. L. Mechanical properties and stacking fault energies of NiFeCrCoMn high-entropy alloy. *JOM* **65**, 1780–1789 (2013).

103. Leyson, G. P. M., Hector, L. G. & Curtin, W. A. Solute strengthening from first principles and application to aluminum alloys. *Acta Mater.* **60**, 3873–3884 (2012).

104. Tian, F. *et al.* Structural stability of NiCoFeCrAlx high-entropy alloy from ab initio theory. *Phys. Rev. B - Condens. Matter Mater. Phys.* **88**, 1–5 (2013).

105. Ma, D., Grabowski, B., Körmann, F., Neugebauer, J. & Raabe, D. Ab initio

thermodynamics of the CoCrFeMnNi high entropy alloy: Importance of entropy contributions beyond the configurational one. *Acta Mater.* **100**, 90–97 (2015).

106. Li, Z., Körmann, F., Grabowski, B., Neugebauer, J. & Raabe, D. Ab initio assisted design of quinary dual-phase high-entropy alloys with transformation-induced plasticity. *Acta Mater.* **136**, 262–270 (2017).
107. Wang, W. R. *et al.* Effects of Al addition on the microstructure and mechanical property of Al xCoCrFeNi high-entropy alloys. *Intermetallics* **26**, 44–51 (2012).
108. Wang, Y. *et al.* Computation of entropies and phase equilibria in refractory V-Nb-Mo-Ta-W high-entropy alloys. *Acta Mater.* **143**, 88–101 (2018).
109. Rogal, L. *et al.* Computationally-driven engineering of sublattice ordering in a hexagonal AlHfScTiZr high entropy alloy. *Sci. Rep.* **7**, 1–14 (2017).
110. Wu, B. *et al.* Microstructures and thermodynamic properties of high-entropy alloys CoCrCuFeNi. *Intermetallics* **93**, 40–46 (2018).
111. Körmann, F., Ruban, A. V. & Sluiter, M. H. F. Long-ranged interactions in bcc NbMoTaW high-entropy alloys. *Mater. Res. Lett.* **5**, 35–40 (2017).
112. Nguyen, M. C. *et al.* Cluster-Expansion Model for Complex Quinary Alloys: Application to Alnico Permanent Magnets. *Phys. Rev. Appl.* **8**, 1–8 (2017).
113. Niu, C., Windl, W. & Ghazisaeidi, M. Multi-Cell Monte Carlo Relaxation method for predicting phase stability of alloys. *Scr. Mater.* **132**, 9–12 (2017).
114. Tamm, A., Aabloo, A., Klintenberg, M., Stocks, M. & Caro, A. Atomic-scale properties of Ni-based FCC ternary, and quaternary alloys. *Acta Mater.* **99**, 307–312 (2015).
115. Widom, M., Huhn, W. P., Maiti, S. & Steurer, W. Hybrid Monte Carlo/molecular dynamics simulation of a refractory metal high entropy alloy. *Metall. Mater. Trans. A Phys. Metall. Mater. Sci.* **45**, 196–200 (2014).
116. Huhn, W. P. & Widom, M. Prediction of A2 to B2 phase transition in the high-entropy alloy Mo-Nb-Ta-W. *Jom* **65**, 1772–1779 (2013).
117. Singh, P. *et al.* Design of high-strength refractory complex solid-solution alloys. *npj Comput. Mater.* **4**, (2018).
118. Singh, P., Smirnov, A. V. & Johnson, D. D. Ta-Nb-Mo-W refractory high-entropy alloys: Anomalous ordering behavior and its intriguing electronic origin. *Phys. Rev. Mater.* **2**, 1–6 (2018).

119. Ogura, M., Fukushima, T., Zeller, R. & Dederichs, P. H. Structure of the high-entropy alloy AlxCrFeCoNi: fcc versus bcc. *J. Alloys Compd.* **715**, 454–459 (2017).

120. Feng, R., Liaw, P. K., Gao, M. C. & Widom, M. First-principles prediction of high-entropy-alloy stability. *npj Comput. Mater.* **3**, 1–6 (2017).

121. Fernández-Caballero, A., Wróbel, J. S., Mummery, P. M. & Nguyen-Manh, D. Short-Range Order in High Entropy Alloys: Theoretical Formulation and Application to Mo-Nb-Ta-V-W System. *J. Phase Equilibria Diffus.* **38**, 391–403 (2017).

122. Feng, W., Qi, Y. & Wang, S. Effects of short-range order on the magnetic and mechanical properties of FeCoNi(AlSi)x high entropy alloys. *Metals (Basel).* **7**, (2017).

123. Lederer, Y., Toher, C., Vecchio, K. S. & Curtarolo, S. The search for high entropy alloys: A high-throughput ab-initio approach. *Acta Mater.* **159**, 364–383 (2018).

124. Curtarolo, S. *et al.* AFLOW: An automatic framework for high-throughput materials discovery. *Comput. Mater. Sci.* **58**, 218–226 (2012).

125. Sher, A., Schilfgaarde, M. Van, Chen, A. & Chen, W. Quasichemical approximation in binary alloys. *Phys. Rev. B* **36**, 4297 (1987).

126. Schmidt, J., Marques, M. R. G., Botti, S. & Marques, M. A. L. Recent advances and applications of machine learning in solid-state materials science. *npj Computational Materials* vol. 5 (2019).

127. Zhang, J. *et al.* Robust data-driven approach for predicting the configurational energy of high entropy alloys. *Mater. Des.* **185**, 108247 (2020).

128. Kaufmann, K. *et al.* Discovery of high-entropy ceramics via machine learning. *npj Comput. Mater.* **6**, 42 (2020).

129. Zhang, Y. *et al.* Phase prediction in high entropy alloys with a rational selection of materials descriptors and machine learning models. *Acta Mater.* **185**, 528–539 (2020).

130. Dai, D. *et al.* Using machine learning and feature engineering to characterize limited material datasets of high-entropy alloys. *Comput. Mater. Sci.* **175**, (2020).

131. Wu, Q. *et al.* Uncovering the eutectics design by machine learning in the Al–Co–Cr–Fe–Ni high entropy system. *Acta Mater.* **182**, 278–286 (2020).

132. Grabowski, B. *et al.* Ab initio vibrational free energies including anharmonicity for multicomponent alloys. **5**, 1–6 (2019).

133. Dai, F. Z., Wen, B., Sun, Y., Xiang, H. & Zhou, Y. Theoretical prediction on thermal and

mechanical properties of high entropy (Zr0.2Hf0.2Ti0.2Nb0.2Ta0.2)C by deep learning potential. *J. Mater. Sci. Technol.* **43**, 168–174 (2020).

134. Rickman, J. M. *et al.* Materials informatics for the screening of multi-principal elements and high-entropy alloys. **10**, 1–10 (2019).

135. Butler, K. T., Davies, D. W., Cartwright, H., Isayev, O. & Walsh, A. Machine learning for molecular and materials science. *Nature* (2018) doi:10.1038/s41586-018-0337-2.

136. Lookman, T., Balachandran, P. V., Xue, D. & Yuan, R. Active learning in materials science with emphasis on adaptive sampling using uncertainties for targeted design. *npj Comput. Mater.* **5**, (2019).

137. Podryabinkin, E. V., Tikhonov, E. V., Shapeev, A. V. & Oganov, A. R. Accelerating crystal structure prediction by machine-learning interatomic potentials with active learning. *Phys. Rev. B* **99**, 064114 (2019).

138. Benesty, J., Chen, J., Huang, Y. & Cohen, I. Pearson Correlation Coefficient. in *Noise Reduction in Speech Processing. Springer Topics in Signal Processing* vol. 2 1–4 (Springer, Berlin, Heidelberg, 2009).

139. Lee, K., Ayyasamy, M. V., Delsa, P., Hartnett, T. Q. & Balachandran, P. V. Phase classification of multi-principal element alloys via interpretable machine learning. *npj Comput. Mater.* **8**, 1–12 (2022).

140. Derimow, N. & Abbaschian, R. Liquid phase separation in high-entropy alloys-a review. *Entropy* **20**, (2018).

141. Nagase, T. Microstructure of Co-Cr-Fe-Mn-Ni-Cu and Co-Cr-Fe-Mn-Ni-Ag high entropy alloys with liquid phase separation. *Mater. Sci. Forum* **941 MSF**, 1238–1241 (2018).

142. Tsai, M. H., Tsai, R. C., Chang, T. & Huang, W. F. Intermetallic Phases in High-Entropy Alloys: Statistical Analysis of their Prevalence and Structural Inheritance. *Met. 2019, Vol. 9, Page 247* **9**, 247 (2019).

143. Xie, S. R. *et al.* Machine learning of superconducting critical temperature from Eliashberg theory. *npj Comput. Mater. 2022 81* **8**, 1–8 (2022).

144. Tsai, M.-H. H. *et al.* Criterion for sigma phase formation in Cr- and V-Containing high-entropy alloys. *Mater. Res. Lett.* **1**, 207–212 (2013).

145. Jaiswal, U. K., Vamsi Krishna, Y., Rahul, M. R. & Phanikumar, G. Machine learning-enabled identification of new medium to high entropy alloys with solid solution phases.

Comput. Mater. Sci. **197**, 110623 (2021).

146. Lee, S. S. Y., Byeon, S., Kim, H. S., Jin, H. & Lee, S. S. Y. Deep learning-based phase prediction of high-entropy alloys: Optimization, generation, and explanation. *Mater. Des.* **197**, 109260 (2021).
147. Xiong, J., Shi, S. Q. & Zhang, T. Y. Machine learning of phases and mechanical properties in complex concentrated alloys. *J. Mater. Sci. Technol.* **87**, 133–142 (2021).
148. Krishna, Y. V., Jaiswal, U. K. & R, R. M. Machine learning approach to predict new multiphase high entropy alloys. *Scr. Mater.* **197**, 113804 (2021).
149. Risal, S., Zhu, W., Guillen, P. & Sun, L. Improving phase prediction accuracy for high entropy alloys with Machine learning. *Comput. Mater. Sci.* **192**, 110389 (2021).
150. Roy, A., Babuska, T., Krick, B. & Balasubramanian, G. Machine learned feature identification for predicting phase and Young's modulus of low-, medium- and high-entropy alloys. *Scr. Mater.* **185**, 152–158 (2020).
151. He, H., Bai, Y., Garcia, E. A. & Li, S. ADASYN: Adaptive synthetic sampling approach for imbalanced learning. *Proc. Int. Jt. Conf. Neural Networks* 1322–1328 (2008) doi:10.1109/IJCNN.2008.4633969.
152. Chawla, N. V., Bowyer, K. W., Hall, L. O. & Kegelmeyer, W. P. SMOTE: Synthetic Minority Over-sampling Technique. *J. Artif. Intell. Res.* **16**, 321–357 (2002).
153. Felser, C., Fecher, G. H. & Balke, B. Spintronics: A Challenge for Materials Science and Solid-State Chemistry. *Angew. Chemie Int. Ed.* **46**, 668–699 (2007).
154. Zhang, P., Li, S. X. & Zhang, Z. F. General relationship between strength and hardness. *Mater. Sci. Eng. A* **529**, 62–73 (2011).
155. Hu, Y. J., Sundar, A., Ogata, S. & Qi, L. Screening of generalized stacking fault energies, surface energies and intrinsic ductile potency of refractory multicomponent alloys. *Acta Mater.* **210**, 116800 (2021).
156. Dey, S., Sultana, N., Kaiser, M. S., Dey, P. & Datta, S. Computational intelligence based design of age-hardenable aluminium alloys for different temperature regimes. *Mater. Des.* **92**, 522–534 (2016).
157. Mohanty, I., Bhattacharjee, D. & Datta, S. Designing cold rolled IF steel sheets with optimized tensile properties using ANN and GA. *Comput. Mater. Sci.* **50**, 2331–2337 (2011).
158. Stepanov, N. D., Yurchenko, N. Y., Skibin, D. V., Tikhonovsky, M. A. & Salishchev, G. A.

Structure and mechanical properties of the AlCr<inf>x</inf>NbTiV (x = 0, 0.5, 1, 1.5) high entropy alloys. *J. Alloys Compd.* **652**, 266–280 (2015).

159. Soni, V., Senkov, O. N., Gwalani, B., Miracle, D. B. & Banerjee, R. Microstructural Design for Improving Ductility of An Initially Brittle Refractory High Entropy Alloy. *Sci. REpoRTS* | **8**, 8816 (2018).

160. Zou, Y., Maiti, S., Steurer, W. & Spolenak, R. Size-dependent plasticity in an Nb25Mo25Ta25W25 refractory high-entropy alloy. *Acta Mater.* **65**, 85–97 (2014).