PHASE PREDICTION OF HIGH ENTROPY ALLOYS
VIA MACHINE LEARNING METHOD

JIRAPRACHA THAMPIRIYANON

A Thesis Submitted in Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy in Materials Engineering

Suranaree University of Technology

Academic Year 2024

การทำนายเฟสของโลหะผสมเอนโทรปีสูงโดยวิธีการเรียนรู้ของเครื่อง

นายจิรประชา ธรรมพิริยานนท์

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรดุษฎีบัณฑิต
สาขาวิชาวิศวกรรมวัสดุ
มหาวิทยาลัยเทคโนโลยีสุรนารี
ปีการศึกษา 2567

PHASE PREDICTION OF HIGH ENTROPY ALLOYS
VIA MACHINE LEARNING METHOD

Suranaree University of Technology has approved this thesis submitted in partial fulfillment of the requirements for the Degree of Doctor of Philosophy.

Thesis Examining Committee

....

(Assoc. Prof. Dr. Vitoon Uthaisangsuk)

Chairperson

....

(Asst. Prof. Dr. Sakhob Khumkoa)

Member (Thesis Advisor)

....

(Prof Dr. Thanatchai Kulworawanichpong)

Member

....

(Asst. Prof. Dr. Tapany Patcharawit)

Member

....

(Asst. Prof. Dr. Sarum Boonmee)

Member

....
(Assoc. Prof. Dr. Yupaporn Ruksakulpiwat)
Vice Rector for Academic Affairs and
Quality Assurance

....
(Assoc. Prof. Dr. Pornsiri Jongkol)
Dean of Institute of Engineering

จิรประชา ธรรมพิริยานนท์ : การทำนายเฟสของโลหะผสมเอนโทรปีสูงโดยวิธีการเรียนรู้ของ
เครื่อง (PHASE PREDICTION OF HIGH ENTROPY ALLOYS VIA MACHINE LEARNING
METHOD) อาจารย์ที่ปรึกษา: ผู้ช่วยศาสตราจารย์ ดร. สงบ คำค้อ, 157 หน้า

คำสำคัญ: โลหะผสมเอนโทรปีสูง/การเรียนรู้ของเครื่อง/การทำนายเฟส

โลหะผสมเอนโทรปีสูง (High entropy alloys) เป็นวัสดุชนิดใหม่ที่มีสมบัติเชิงกลที่โดดเด่น
มีความเสถียรทางความร้อนและมีความต้านทานการกัดกร่อนที่ยอดเยี่ยม ทำให้เหมาะสมสำหรับการ
นำไปใช้งานในอุตสาหกรรมที่หลากหลาย อย่างไรก็ตาม โลหะผสมเอนโทรปีสูงมีองค์ประกอบที่
ซับซ้อน ดังนั้นการทำนายโครงสร้างเฟสที่เกิดขึ้นจากการปรับแต่งสมบัติเพื่อให้เหมาะสมกับการใช้
งานเฉพาะด้านจึงเป็นเรื่องที่ท้าทาย งานวิจัยนี้นำเสนอวิธีการเรียนรู้ของเครื่อง (Machine learning)
เพื่อสร้างแบบจำลองการทำนายโครงสร้างเฟสของโลหะผสมเอนโทรปีสูงโดยใช้ชุดข้อมูลที่ได้จากการ
ทดลองร่วมกับคุณลักษณะ (Feature) ที่คำนวณจากองค์ประกอบทางเคมีของวัสดุในการฝึก
แบบจำลอง งานวิจัยนี้ใช้เทคนิคการเข้ารหัสเวกเตอร์บูลีน (Boolean vector encoding) เพื่อแทน
ค่าการจำแนกเฟสแบบหลายเฟส ทำให้แบบจำลองสามารถจับความสัมพันธ์ที่ซับซ้อนของเฟสได้
อย่างมีประสิทธิภาพ โดยได้นำอัลกอรีทึม 4 แบบ ได้แก่ Support vector machine (SVM)
k-nearest neighbors (KNN) Random forest (RF) และ Neural network (NN) มาใช้สร้าง
แบบจำลองสำหรับทำนายเฟสและทำการประเมินความแม่นยำในการทำนายจากการทดสอบกับชุด
ข้อมูลที่ไม่เคยนำมาใช้ในการฝึกมาก่อน ผลการศึกษาพบว่า NN และ KNN เป็นแบบจำลองที่มี
ประสิทธิภาพสูงสุดในการทำนายโครงสร้างเฟสโดยมีความแม่นยำที่ 84.85% นอกจากนี้การ
วิเคราะห์ความสำคัญของคุณลักษณะชี้ให้เห็นว่า ความเข้มข้นของอิเล็กตรอนชั้นนอกสุดและอุณหภูมิ
หลอมเหลวเป็นปัจจัยที่มีอิทธิพลต่อการเกิดเฟส งานวิจัยนี้แสดงให้เห็นถึงศักยภาพของนำการเรียนรู้
ของเครื่องมาใช้ในด้านวัสดุศาสตร์เชิงคำนวณและเป็นแนวทางในการนำปัญญาประดิษฐ์มาใช้ใน
ออกแบบและพัฒนาโลหะผสมเอนโทรปีสูง

สาขาวิชา วิศวกรรมโลหการ   ลายมือชื่อนักศึกษา................................

ปีการศึกษา 2567   ลายมือชื่ออาจารย์ที่ปรึกษา................................

JIRAPRACHA THAMPIRIYANON : PHASE PREDICTION OF HIGH ENTROPY ALLOYS VIA MACHINE LEARNING METHOD   THESIS ADVISOR: ASSOC.PROF, SAKHOB KHUMKOA, Ph.D., 157 PP.

Keyword: High entropy alloy/Machine learning/Phase prediction

High entropy alloys (HEAs) represent a novel category of materials known for their outstanding mechanical strength, thermal stability, and corrosion resistance, making them highly suitable for diverse industrial applications. However, their complex and variable compositions pose significant challenges in accurately predicting phase structures, which is essential for tailoring their properties to specific applications. This study presents a machine learning (ML) approach for phase structure prediction in HEAs, utilizing experimentally validated datasets with features derived from chemical compositions. A Boolean vector encoding technique was employed to effectively represent multi-phase classifications, enhancing the model's ability to capture complex phase relationships. Four ML algorithms including support vector machine (SVM), k-nearest neighbors (KNN), random forest (RF), and neural network (NN) were implemented and evaluated through rigorous cross-validation (CV) and testing on unseen data. The findings indicate that NN and KNN outperform other models, achieving a maximum test accuracy of 84.85%. Feature importance analysis highlights valence electron concentration (VEC) and melting temperature (Tm) as the most influential factors in phase formation. This research demonstrates the growing potential of ML-driven approaches in computational materials science and suggest promising pathways for integrating artificial intelligence to enhance HEA design and development.

School of Metallurgical Engineering   Student's Signature..........................

Academic Year 2024   Advisor's Signature..........................

ACKNOWLEDGEMENT

TABLE OF CONTENTS

Page

CHAPTER

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued)

Page

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES (Continued)

LIST OF ABBREVIATIONS

AdaBoost	Adaptive boosting
Adam	Adaptive moment estimation
AM	Amorphous
AUC	Area under the curve
BCC	Body-centered cubic
BMG	Bulk metallic glass
CALPHAD	Calculation of phase diagrams
CV	Cross-validation
DFT	Density functional theory
DT	Decision Tree
FCC	Face-centered cubic
FN	False negative
FP	False positive
GA	Genetic algorithm
GB	Gradient boosting
GLMNET	General linear model via elastic net
HCP	Hexagonal closed-pack
HEA	High entropy alloy
IM	Imtermetallic
KNN	k-nearest neighbors
KPCA	Kernel principal component analysis
LASSO	Least absolute shrinkage and selection operator
LDA	Linear discriminant analysis
LEA	Low entropy alloy

LIST OF ABBREVIATIONS (Continued)

LR	Logistic regression
MEA	Medium entropy alloy
ML	Machine learning
MP	Multi-phase
MPEA	Multi-principal element alloy
NB	Naïve Bayes
NN	Neural network
PR	Precision-recall
ReLU	Rectified linear unit
RF	Random forest
RFE	Recursive feature elimination
ROC	Receiver operating characteristic
SGD	Stochastic gradient
SS	Solid solution
SVM	Support vector machine
TN	True negative
TP	True positive
TTT	Time-Temperature-Transformation
UTS	Ultimate tensile strength
VEC	Valence electron concentration
XGBoost	Extreme gradient boosting

CHAPTER I

INTRODUCTION

1.1   Background

High entropy alloys (HEAs) are an emerging class of materials that differ from conventional alloys by incorporating multiple principal elements in nearly equiatomic proportions. This unique composition strategy results in remarkable mechanical, thermal and corrosion-resistant properties, making HEAs highly promising for applications in aerospace, energy, and biomedical fields. However, the complex and variable compositions of HEAs pose significant challenges in accurately predicting their phase structures, which is crucial for tailoring their properties for specific applications. Traditional phase prediction methods, such as empirical phase formation rules, the Calculation of Phase Diagrams (CALPHAD), and Density functional Theory (DFT), have been employed to determine phase stability. While these methods provide valuable insights, they often require extensive computational resources and experiment validation. In the recent years, machine learning (ML) has emerged as a powerful tool in material science, offering efficient and accurate predictive capabilities by leveraging vast datasets and complex algorithms. The integration of ML with HEA phase prediction has the potential to accelerate materials discovery by reducing reliance on trial-and-error experimental approaches.

1.2   Problem statement

Despite the promising advancement in ML-driven phase prediction of HEAs, several challenges remain unaddressed:

1. Most existing ML models classify HEA phases into broad categories, such as, single-phase solid solutions (SS) such as body-centered cubic (BCC) ,and face- centered cubic (FCC) or intermetallic compounds (IM), without considering complex multi-phase combinations.

2. The dataset imbalance among different phase compositions affects the model's ability to generalize, particularly for underrepresented phase combinations like BCC+FCC+IM.

Addressing these challenges is essential to develop a more robust, generalizable and computationally efficient ML-based phase prediction model for HEAs.

1.3   Research objectives

This study aims to develop a machine learning-based framework for predicting complex phase combinations in HEAs with enhanced accuracy and generalizability. The specific objectives of this research are:

1. To construct a comprehensive dataset of HEA compositions and corresponding experimentally validated phase structures,

2. To implement and evaluate various ML algorithms, including support vector machine (SVM), k-nearest neighbors (KNN), random forest (RF), and neural networks (NN), for phase prediction.

3. To introduce a Boolean vector encoding technique to accurately classify multi-phase compositions in HEAs

4. To analyze the feature importance of difference thermodynamic and atomic descriptors in governing phase formation.

5. To compare predictive performance of ML models, highlighting the advantages and limitations of each approach.

1.4   Significance of the study

This research contributes to the field of computational materials science by demonstrating the effectiveness of ML-based phase prediction models for HEAs. The significance of this study includes:

1. Accelerated HEA design: Reduces reliance on time-consuming and costly experimental phase identification methods.

2. Enhanced predictive accuracy: Introduces an improved ML framework capable of handling complex phase combinations

3. Generalizability: Provides insights into model performance across diverse HEAs compositions.

4. Industrial applications: Supports the development of HEAs with tailored properties.

1.5   Thesis structure

The thesis is structured as follows:

Chapter 1 Introduction: Provides an overview of HEAs, the challenges in phase prediction, research objectives, and significance of the study.

Chapter 2 Literature Review: Discusses prior research on ML applications in materials science, HEA phase prediction, and gaps in existing approaches.

Chapter 3 Research Methodology: Details the dataset construction, ML model selection, feature engineering and evaluation metrics.

Chapter 4 Results and Discussion: Presents the findings of data analysis, ML-based phase prediction, compares model performance, and interprets key insights.

Chapter 5 Conclusion and Future work: Summarizes the study's contributions, discusses limitations, and outlines direction for future research.

This chapter establishes the foundation for the research by highlighting the importance of ML in HEA phase prediction and outlining the key research question address in this study. The subsequent chapters build upon this framework to provide a comprehensive analysis of ML-driven phase classification in HEAs.

CHAPTER II

LITERATURE REVIEWS

The purpose of this chapter is to provide an overview of existing research related to HEAs, phase formation mechanisms, and the application of ML technique in materials science. This review identifies gaps in current knowledge and highlights the necessity of employing ML for predicting complex phase formation in HEAs.

2.1   High entropy alloys (HEAs)

HEAs are a novel class of metallic materials that contain multiple principal elements in nearly equiatomic ratios. Unlike conventional alloys, which are based on a single dominant element, HEAs exhibit unique microstructures and exceptional mechanical, thermal, and chemical properties. The concept of HEAs was first introduced in 2004 (Yeh et al., 2004), emphasizing the role of configurational entropy in stabilizing multi-element SS.

2.1.1   Composition-based definition

Unlike conventional alloys that are predominantly based on one or two principal elements with minor additions of alloying elements, HEAs do not have a single dominant base element. Instead, each principal component contributes significantly to the overall composition, leading to a complex atomic arrangement. The key criteria defining HEAs based composition is the presence of at least five principal elements, each contributing between 5-35 atomic % to the overall composition.

2.1.2 Entropy-based definition

An alternative and more fundamental way to classify HEAs is through their configurational entropy (ΔS_{conf}), which is derived from statistical thermodynamics. The entropy-based classification considers the thermodynamic stability of the alloy based on its mixing entropy rather than simply the number of principal elements.

The configurational entropy of multicomponent system is given by the equation 2.1:

$$\Delta S_{conf} = -R \sum_{i=1}^{n} c_i \ln c_i \tag{2.1}$$

where:

R is the gas constant (8.314 J/(mol.K)),

c_i is the atomic fraction of the ith element,

n is the total number of principal elements in the system.

Based on the magnitude of ΔS_{conf}, metallic alloys can be categorized as follows

Low entropy alloys (LEAs): $\Delta S_{conf} < R$

Medium entropy alloys (MEAs): $R < \Delta S_{conf} < 1.5R$

High entropy alloys (HEAs): $\Delta S_{conf} > 1.5R$

2.1.3 Characteristics and properties of HEAs

HEAs exhibit unique characteristics and outstanding properties due to their complex composition and inherent phase stability. These properties stem from four core effects that distinguish HEAs from conventional alloys. The four core effects are:

1. High entropy effect.

The high entropy stabilizes SS phases (e.g., BCC, FCC) over IM, reducing phase segregation. This effect promotes thermodynamic stability and influences phase transformation behaviors.

2. Lattice distortion effect.

The significant atomic size differences among constituent elements create lattice distortions. These distortions affect dislocation motion, increasing strength and hardness while maintaining ductility. The atomic-level stress fields enhance strain hardening and mechanical properties. Figure 2.1 highlighted the superior mechanical performance of HEAs due to their unique combination of strength and ductility.
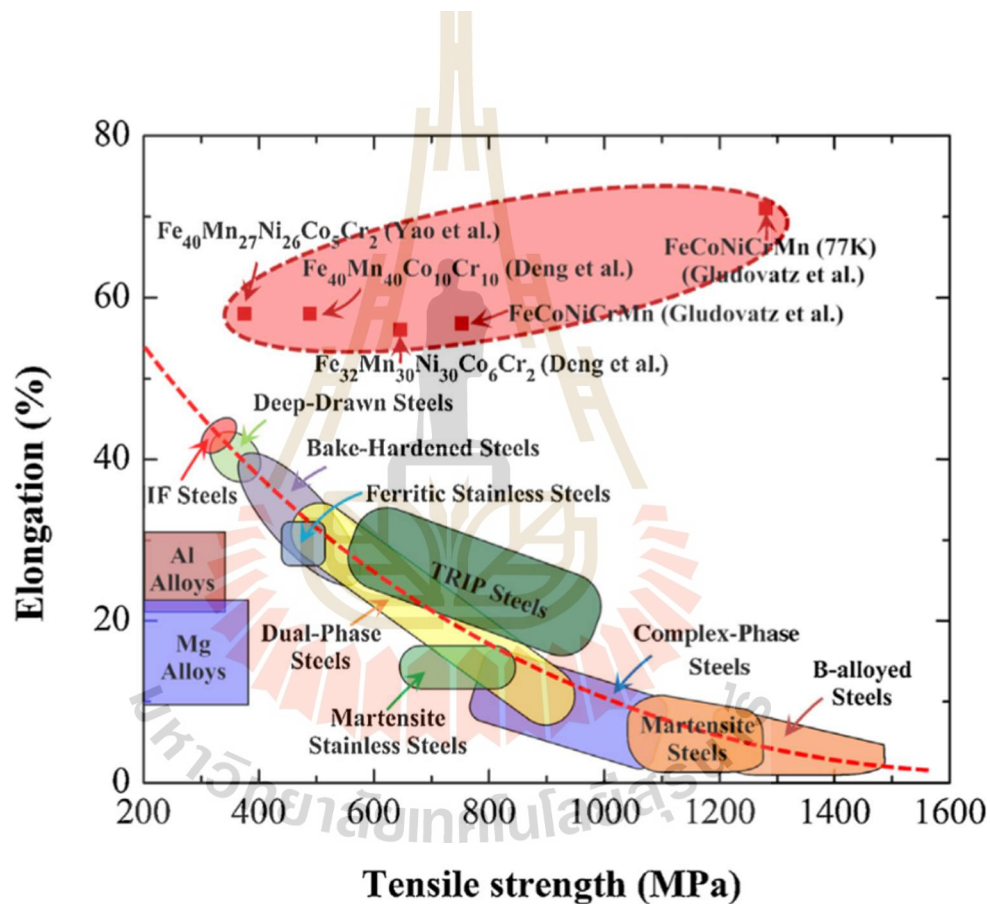
Figure 2.1 Strength versus ductility properties of HEAs in comparison with conventional alloys (Ye et al., 2016).

3. Sluggish diffusion effect.

HEAs exhibit slower atomic diffusion due to diverse atomic environments which suppresses grain growth at elevated temperatures. This effect contributes to enhanced thermal stability, oxidation resistance and creep resistance. The sluggish diffusion mechanism influences phase transformations and microstructural evolution. Figure 2.2 compared the ultimate tensile strength (UTS) of HEAs with superalloys at different high-temperature conditions.
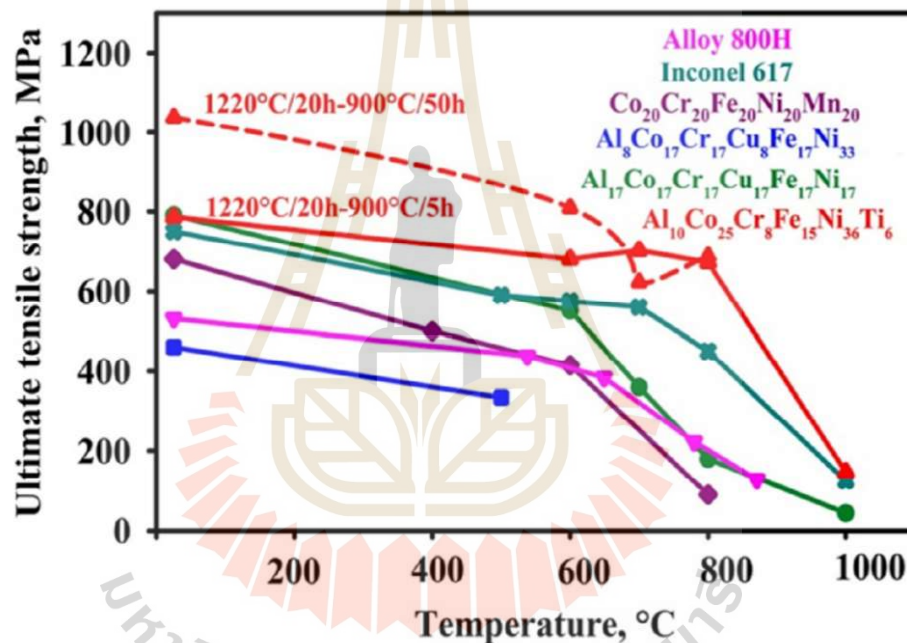
Figure 2.2 Ultimate tensile strength of HEAs at elevated temperatures compared to conventional superalloys (J. Chen et al., 2018).

    4.   Cocktail effect.

        The synergistic interaction among multiple elements results in properties that surpass those of individual components. This effect is responsible for superior mechanical, thermal, and corrosion resistance properties in HEAs. It enables the fine-tuning of alloy properties through compositional modification, providing flexibility for various application. Figure 2.3 compared the corrosion resistance of HEAs with other conventional alloys.
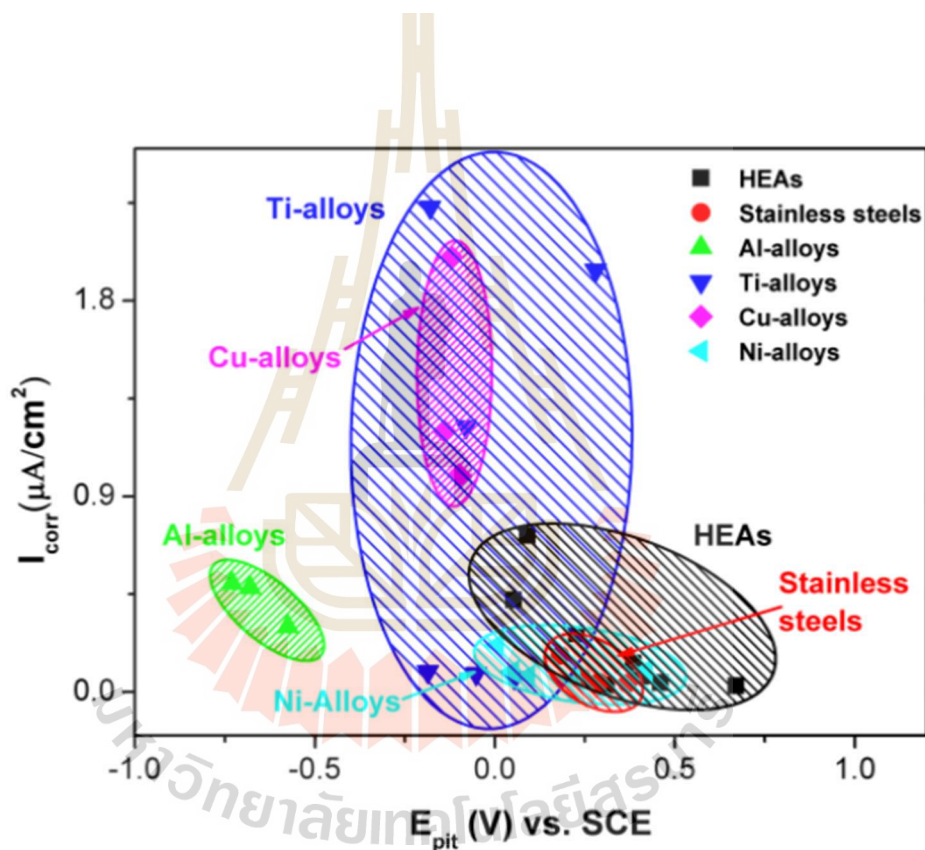
Figure 2.3 A comparison of current density and pitting potential between
HEAs and other alloys (Shi et al., 2017).

The combination of these four core effects makes HEAs promising materials for structural and functional applications. Their unique stability, strength, and durability make them suitable for aerospace, nuclear and biomedical industries. Further research on these effects will enable the development of next generation high-performance material.

2.1.4 Phase formation and microstructure in HEAs

The formation of phases and resulting microstructures in HEAs is highly complex due to their multi-principal element design and the vast compositional space. With a large number of possible element compositions, HEAs exhibit diverse microstructures, including single-phase solid solutions, multi-phase mixtures, and nanocrystalline structures. Among these, body-centered cubic (BCC) and face-centered cubic (FCC) phases are the most frequently observed, although other structures, such as hexagonal close-packed (HCP), intermetallic (IMs) phases have also reported.

The BCC phase is generally characterized by high strength, hardness and wear resistance, though it often suffers from limited ductility. In contrast, the FCC phase provide superior ductility, toughness, and corrosion resistance, making it more suitable for application requiring mechanical flexibility. The HCP phase, while less common, may form under specific composition and exhibits anisotropic mechanical behavior, which can be beneficial in lightweight or directional-strength materials. IM phases are ordered structures resulting from strong atomic bonding. These phases can enhance strength and thermal stability but may also reduce ductility, depending on their morphology and distribution.

In general, the configurational entropy in HEAs favors the formation of disordered solid solution (SS) phases with simple crystals structures rather than ordered IM phases. However, IM phase can still emerge, especially in composition with significant atomic size or electronegativity mismatches. The understanding of phase formation is essential for tailoring the properties of HEAs for specific applications.

.

2.2 Empirical and computational approaches for phase prediction in HEAs

Several approaches have been developed to predict phase formation in HEAs:

2.2.1 Empirical rules

Empirical models have been proposed to predict phase stability based on elemental properties:

Mixing enthalpy (ΔH_{mix}): Governs the thermodynamic interactions between constituent elements.

Mixing entropy (ΔS_{mix}): Influences the likelihood of SS formation.

Melting temperature (T_m): Influences phase stability and the formation of different structures by determining thermal stability.

Atomic size different (δ): Affects lattice distortions and phase stability.

Valence electron concentration (VEC): Differentiates between BCC and FCC structures

Electronegativity difference ($\Delta\chi$): Impacts phase separation and intermetallic formation.

Parameter for prediction solid solution (Ω): Assesses the likelihood of forming a stable SS, where higher values generally favor SS formation over IM.

Figure 2.4 presents a statistical analysis of experimental data on phase selection in HEAs based on δ and ΔH_{mix}. The phase selection between SS and AM phase can be effectively predicted using δ and ΔH_{mix}, SS phases are more likely to form when δ is small ($\delta < 0.666$) and ΔH_{mix} is either slightly positive of weakly negative ($-11.6 < \Delta H_{mix} < 3.2$ kJ/mol). In contrast, the formation of the AM phase is favored when is δ large ($\delta > 0.064$) and ΔH_{mix} is significantly negative ($\Delta H_{mix} < -12.2$ kJ/mol). These trends highlight the strong influence of atomic interactions and structural compatibility in determining phase formation in multi-component alloys.
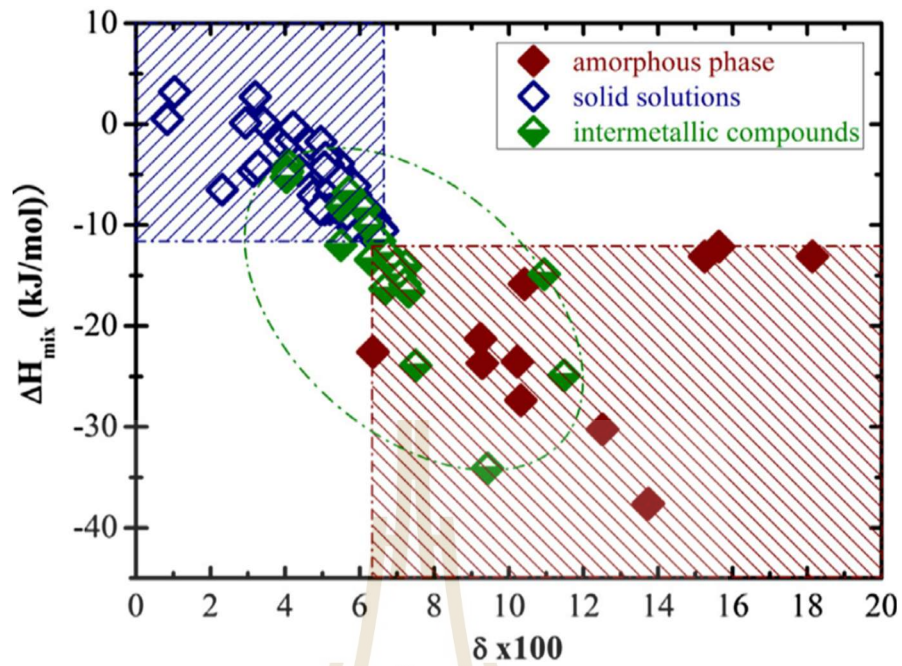
Figure 2.4 Distribution of AM, SS and IM phases effected by ΔH_{mix} and δ (Guo et al., 2013).

Figure 2.5 illustrates the relationship between δ and Ω for multi-component alloys. Alloys that form simple SS phases are clustered in the triangular region where lower δ and high Ω. Both SS and IM may form due to increase δ and reduce Ω value, making it a transition zone between SS and IM. The bulk metallic glasses (BMGs) exhibit larger δ and Ω value compared to HEAs. This analysis suggests that criteria for predicting SS formation in HEAs is $\Omega > 1.1$ and $\delta < 6.6\%$.
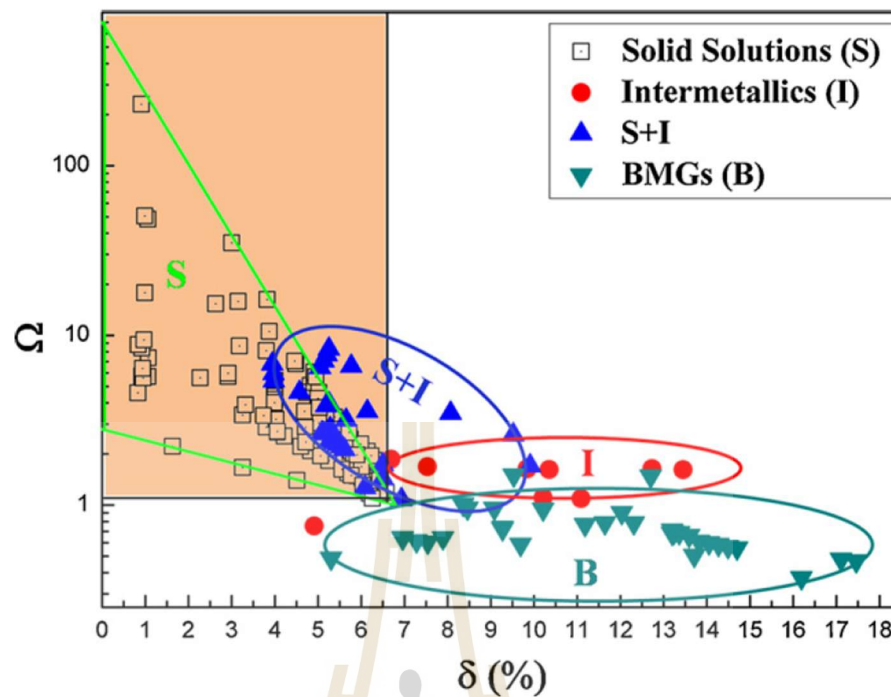
Figure 2.5 Distribution of SS, IM, SS+IM and BMG phases effected by Ω and δ (Yang & Zhang, 2012).

VEC is identified as a key factor governing the stability of FCC and BCC phase, with FCC structures being favored at higher VEC values and BCC phases stabilizing at lower VEC values. This findings provide valuable insights into phase stability in HEAs and offer a fundamental guideline for designing alloy (Guo et al., 2011).

2.2.2 Calculation of phase diagrams (CALPHAD)

The CALPHAD method is a widely adopted computational approach for predicting phase stability in multi-component systems. It employs thermodynamic models and experimentally derived databases to estimate equilibrium phase diagrams based on Gibbs free energy minimization. By integrating thermodynamic descriptions of individual elements and binary or ternary systems, CALPHAD enables the extrapolation of phase behavior in complex alloys, including HEAs (Y. Liu et al., 2021).

Despite its effectiveness, the CALPHAD method has inherent limitations (M. C. Gao et al., 2017). Its accuracy is dependent on the availability and reliability of experimental thermodynamic data, particularly for higher-order multi-element systems. Additionally, the approach relies on simplifying assumptions in phase modeling, which may lead to uncertainties when predicting metastable phases or new compositions beyond the established database.

2.2.3 Density functional theory (DFT)

Density Functional Theory (DFT) is a powerful computational approach used to investigate phase stability at the atomic scale. By solving the Schrödinger equation for many-body systems, DFT provides detailed insights into electronic structures, bonding interactions, and thermodynamic stability of HEAs (Singh et al., 2020). It enables the prediction of phase formation, mechanical properties, and electronic behaviors with high accuracy, making it an essential tool for materials design.

However, DFT has significant limitations. It is computationally intensive, requiring substantial processing time and resources, particularly for multi-component alloys with complex atomic configurations. This makes it impractical for large-scale phase stability predictions across the vast compositional space of HEAs. Additionally, approximations in exchange-correlation functionals can introduce uncertainties in predicted phase stability.

2.3 ML in materials science

ML has revolutionized materials science by enabling the rapid discovery, design, and optimization of novel materials. Unlike conventional trial-and-error approaches, ML utilizes data-driven models to predict material properties, phase structures, and microstructural evolution with high accuracy and efficiency. By leveraging large experimental and computational datasets, ML algorithms can identify complex patterns and correlations that traditional methods may overlook.

In recent years, ML provides advanced and efficient methods for predictive modeling in materials science, facilitating automated microstructure analysis in steels and specialized alloys (Mulewicz et al., 2019). It has been successfully applied to predict phase formation and mechanical properties in entropy alloys (Roy et al., 2020) and to expedite the interpretation of time–temperature–transformation (TTT) diagrams in stainless steels (X. Huang et al., 2020).

2.4   Previous studies on HEA phase prediction

Machine learning (ML) has been extensively utilized in recent years to classify and predict phase stability in HEAs. Various ML approaches, including NN, SVM, decision trees (DT), and RF, KNN have been applied to analyze phase formation and enhance predictive accuracy. Many studies have incorporated feature selection and engineering to improve model performance, demonstrating the effectiveness of ML in HEA phase classification. Table 2.1 presents a comprehensive overview of previous studies utilizing ML approaches for phase prediction in HEAs. It highlights various ML models, key methodologies and classification strategies used to distinguish between different phase structures, showcasing, advancements and challenges in predictive modeling for HEAs.

Islam (Islam et al., 2018) applied a NN model (shown in Figure 2.6) within a ML framework   to classify phase selection in multi-principal element alloys (MPEAs), including HEAs. Utilizing an experimental dataset, their NN model achieved a remarkably high training accuracy of over 99%, and a cross-validation (CV) accuracy exceeding 80%. This demonstrated the model's capability to generalize well and accurately predict phase types such as single-phase SS, AM, and IM. Notably, their findings highlighted the significance of VEC as the most influential feature in determining phase formation. Their study emphasizes the potential of NN-based ML models in exploring complex phase behavior and can be extended to various MPEA systems.
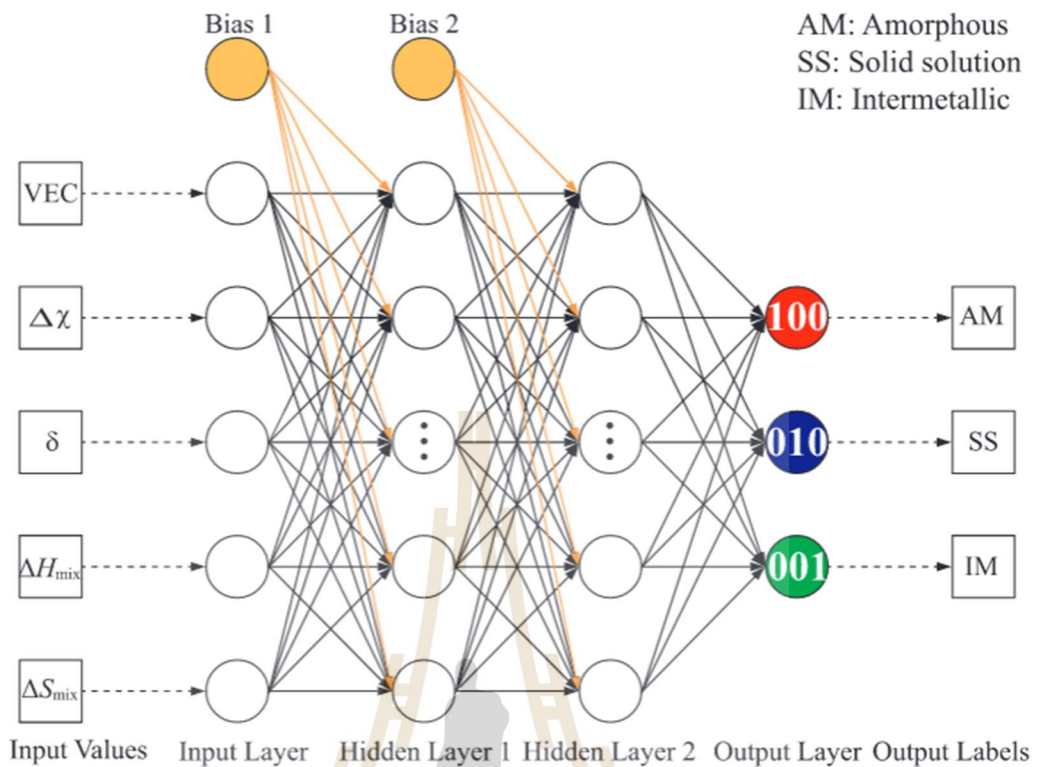
Figure 2.6 An NN model for AM, SS and IM phases prediction (Islam et al., 2018).

Similarly, Huang (W. Huang et al., 2019) explored phase selection rules using a large experimental dataset of 401 HEA compositions classified into SS, IM, and mixed SS+IM phases. They implemented three ML algorithms: KNN, SVM, and NN, and conducted a four-fold CV to prevent overfitting. Among the models, the NN (shown in Figure 2.7) achieved the highest multi-class classification accuracy of 74.3%. When limited to binary classification between pairs of phases, NN consistently outperformed SVM, achieving accuracies of 86.7% to 94.3%. This study demonstrated that ML approaches, particularly NN, can serve as efficient alternatives to computationally expensive methods like DFT for predicting phase selection in HEAs. The results confirmed that ML models could effectively learn correlations between elemental features and phase types, offering valuable insights for computational alloy design.
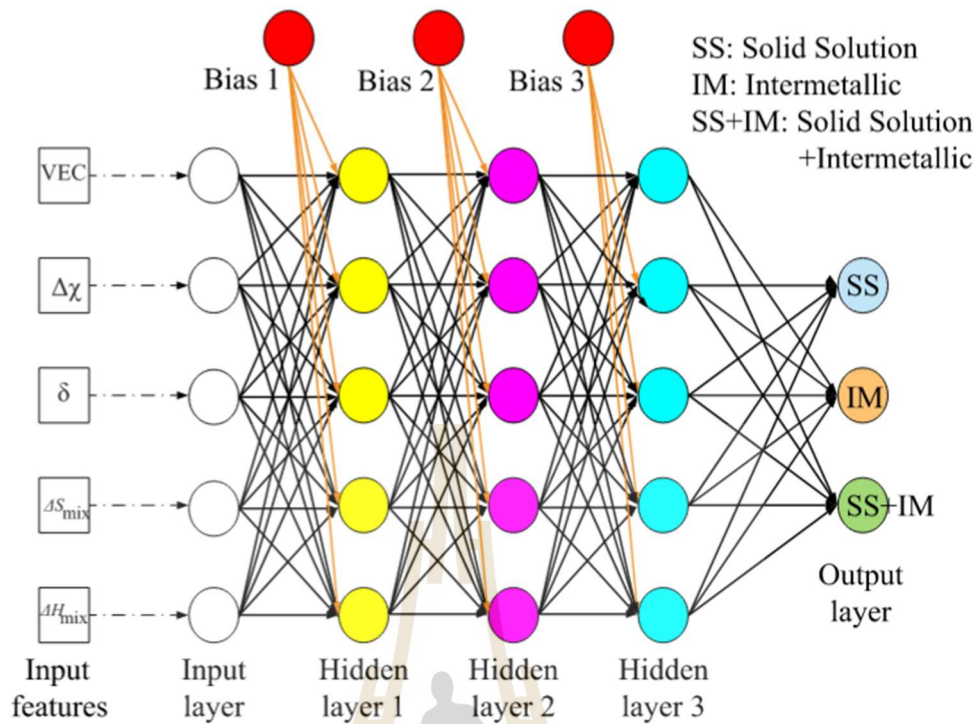
Figure 2.7 An NN model for SS, IM and SS+IM prediction (Huang et at, 2019).

Building on the use of NN and multiple algorithms in previous studies, Dai (Dai et al., 2020) proposed an approach centered on feature engineering combined with machine learning to improve phase prediction accuracy in HEAs. Recognizing the limitations of trial-and-error experimentation and thermodynamic models, the authors augmented the dimensionality of descriptors through non-linear combinations, effectively expanding the feature space from a small dataset. They evaluated six different ML models: SVM, adaptive boosting (AdaBoost), decision tree (DT), RF, gradient boosting (GB), and logistic regression (LR), and demonstrated that this dimensionality expansion significantly improved predictive performance compared to conventional approaches. Their methodology proved particularly useful in small-data scenarios, highlighting the versatility of feature engineering for improving classification accuracy. This study underscores the importance of data transformation and selection

techniques in enhancing the reliability of ML-based predictions, especially for single-phase SS HEAs as shown in Figure 2.7.
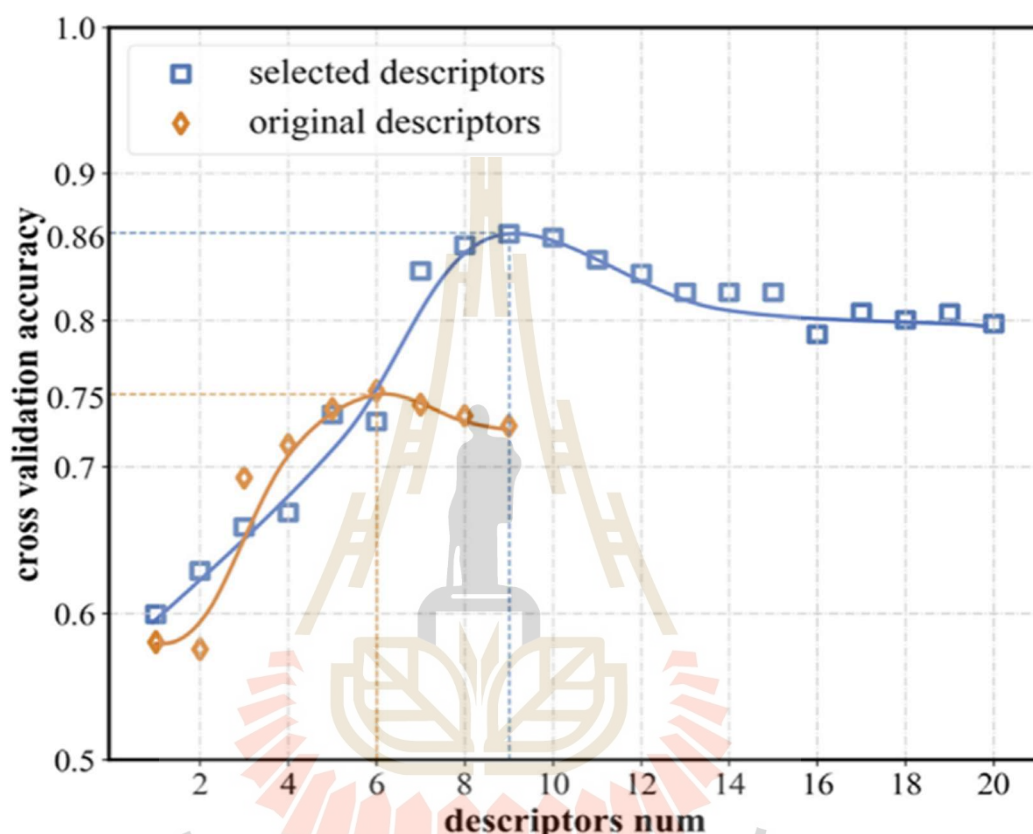
Figure 2.8 CV accuracy versus number of features (Dai et at.,2020).

Zhang (L. Zhang et al., 2020) further advanced the predictive modeling of HEA phases by addressing limitations in traditional ML methods when handling complex thermodynamic descriptors. They incorporated nine physical parameters including formation enthalpy based on the extended Miedema theory and mixing entropy and applied feature selection and transformation using Kernel principal component analysis (KPCA). By reducing and optimizing the feature space, they significantly

enhanced phase classification performance. The SVM model trained with four KPCA-transformed variables (4V-KPCA) achieved a remarkable test accuracy of 97.43% as shown in Figure 2.9. Detailed F1-scores for the model demonstrated high precision across four phase types: SS (0.9787), AM (0.9463), mixed SS+IM (0.9863), and IM (0.8103). Their study confirmed that elastic energy and δ are influential factors in phase formation and that combining physical theory with advanced ML techniques yields reliable predictive tools for HEA design.
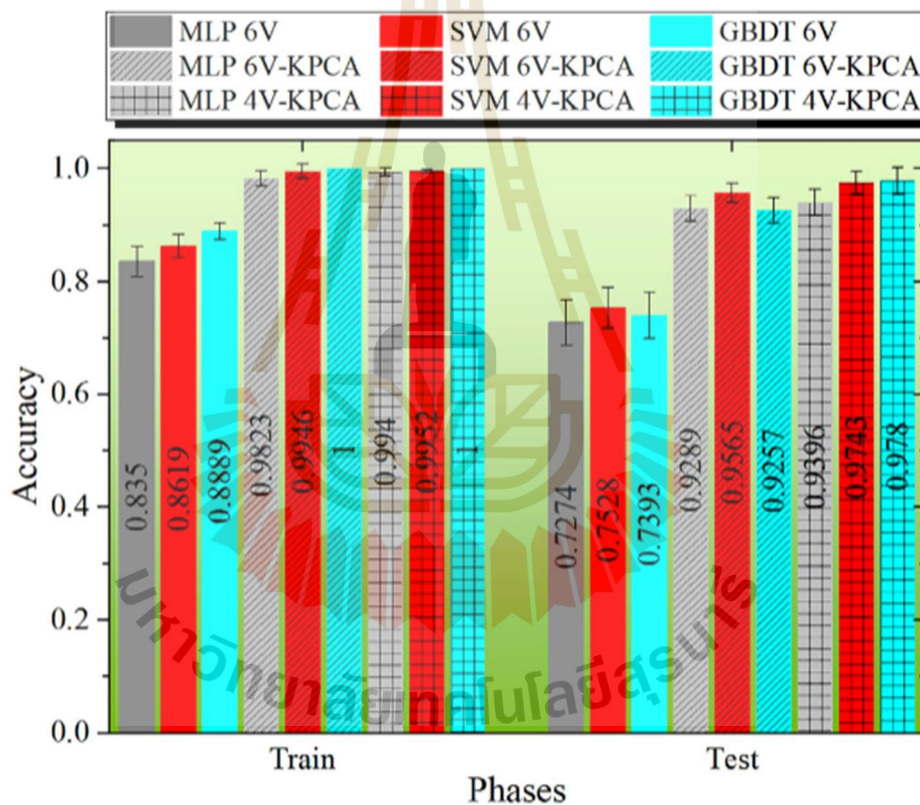
Figure 2.9 The prediction accuracies of difference preprocessing methods and difference models (L. Zhang et at., 2020).

Machaka (Machaka, 2021) addressed the fundamental question of why HEAs exhibit exceptional properties by focusing on phase classification using ML. The study compiled an extensive dataset derived from over 430 peer-reviewed experimental studies and included more than 40 metallurgy specific features. A robust framework (shown in Figure 2.10) was proposed, combining six feature selection methods, feature ensemble construction, and the evaluation of eight general purpose ML classifiers: DT, linear discriminant analysis (LDA), naïve Bayes (NB), generalized linear model via elastic net (GLMNET), RF, NN, KNN, and SVM. The models were trained to classify solid solution phases in HEAs. RF emerged as the top performer with an accuracy of 97.5%, followed by SVM (95.8%), KNN (94.5%), and NN (94.0%). The study revealed that traditional post-treatment heat-treatment features were less effective in phase prediction, while the most discriminating predictors were identified through feature selection. Additionally, the applicability of the developed models was validated across five alloy systems, showing their effectiveness in predicting phase stabilization and phase transitions, thereby bridging experimental and computational insights.
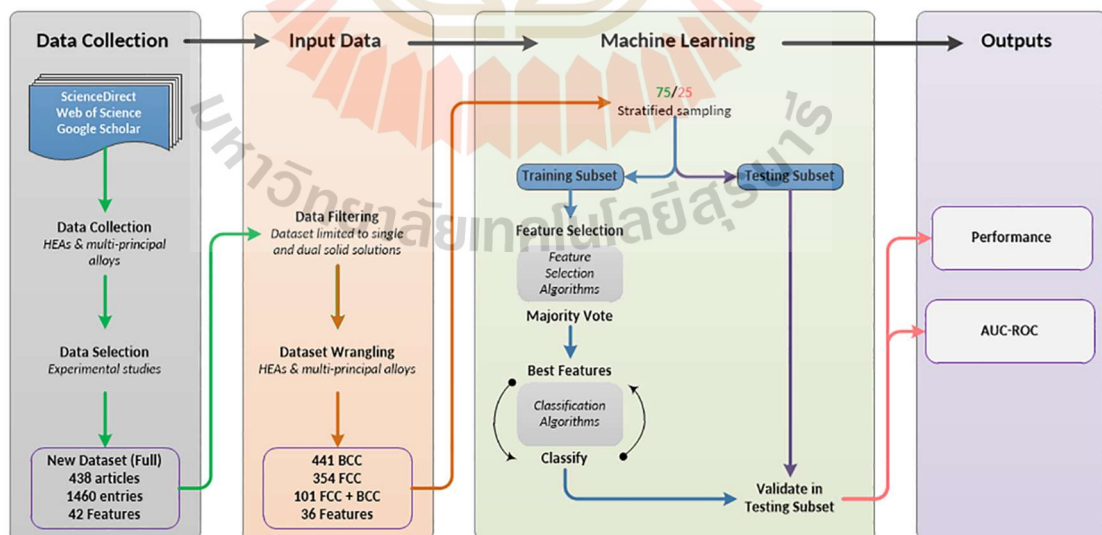
Figure 2.10 ML modeling framework for phase prediction (Machaka, 2021).

Nezhad (Ghouchan Nezhad Noor Nia et al., 2022) introduced a graph-based ML approach to enhance phase prediction in HEAs, addressing the limitations of traditional methods that often suffer from low accuracy and high costs. Their method employed a KNN algorithm within an HEA interaction network, where each compound is connected to others based on feature similarity. This network-based KNN model was used to predict five phase categories: FCC, BCC, HCP, MP, and AM. The key innovation in this work was utilizing network relationships to infer the phase of new compositions by evaluating the most similar neighboring compounds. The model achieved a prediction accuracy of 88.88%, outperforming many conventional ML techniques. This study demonstrated the effectiveness of graph-based models in capturing complex compositional relationships within HEAs, providing a scalable tool for accelerating phase discovery in advanced alloy design.

Syarif (Syarif et al., 2023) explored phase prediction in high-entropy alloys (HEAs) by focusing solely on the concentration of constituent elements, offering a novel approach that diverges from conventional ML methods that rely heavily on feature engineering and domain-specific descriptors. Their model applied pruned DT and linear correlation to develop simple, interpretable prediction rules. These rules were used with self-organizing maps and Euclidean space formulations to frame the identification of phase formation drivers as an optimization problem. Genetic algorithm (GA) optimization revealed that properties such as electron affinity, molar volume, and resistivity significantly influence phase formation. Notably, the study demonstrated that in the AlCoCrFeNiTiCu HEA family, FCC phase formation could be predicted with up to 87% accuracy using only the concentrations of Al and Cu. This approach not only simplifies input requirements but also offers insights into the elemental influences on phase behavior (shown in Figure 2.11), advancing data-driven HEA design.

Chemical Elements categorization into phase stabilizing and destablizing elements

	Stabilizing ⬤	Destabilizing ⬤
FCC	Cu	Al
BCC	Al, Cr, Fe	Cu, Ni
IM	Ti, Fe	Cu, Co

Figure 2.11 Stabilizing and destabilizing elements for FCC, BCC and IM phase formation (Syarif et al., 2023)

Gao (J. Gao et al., 2023) investigated phase prediction in HEAs using ensemble ML techniques on a dataset of 511 labeled compositions. The study applied feature selection via Extreme gradient boosting (XGBoost) to rank the importance of features derived from empirical design principles. Among the models tested, ensemble methods: Voting and Stacking, demonstrated the highest predictive accuracies, both exceeding 92%. The research also introduced a decision tree-based visualization to guide phase identification among FCC, BCC, and FCC+BCC categories. This combination of robust ensemble learning and intuitive model interpretation underscores the potential of ML in accelerating HEA phase design, while also providing criteria for selecting meaningful features for effective prediction.

He (He et al., 2024) explored the use of ML to predict phase structures in HEAs, focusing on SS and AM phases. The study applied five ML algorithms on a dataset of 399 HEAs categorized into BCC, FCC, BCC+FCC, and AM phases. Grid search combined with CV was used to optimize model performance, with VEC and ΔH_{mix} identified as

the most influential features (shown in Figure 2.12). Among the algorithms, the random forest (RF) model delivered the best performance, achieving a predictive accuracy of 87%. The model's predictions were experimentally validated using $CoCrFeNiAl_x$ (x = 0, 0.5, 1) alloys, showing phase transitions from FCC to BCC + FCC and then to BCC as Al content increased, results that aligned with the ML predictions. This highlights the RF model's effectiveness and the practical applicability of ML in guiding HEA phase design.
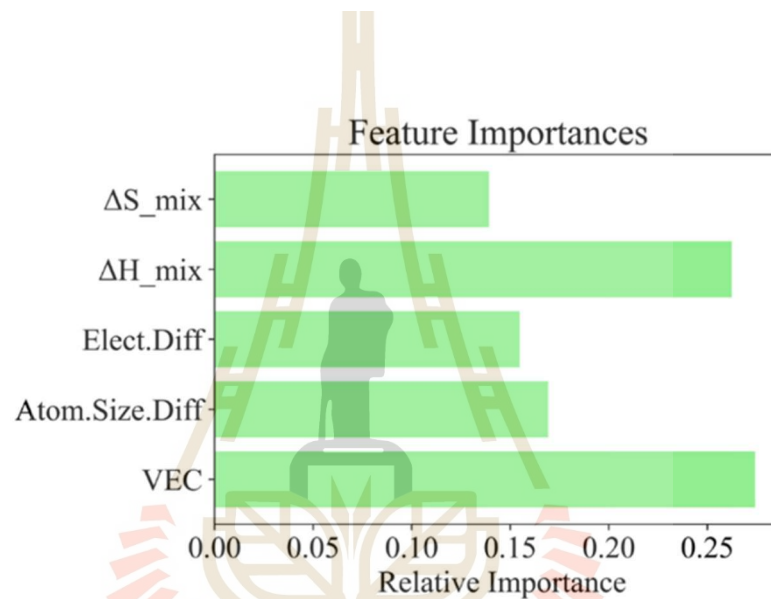
Figure 2.12 Feature importance ranking (He et al., 2024)

These studies demonstrate the potential of ML models in predicting phase stability in HEAs, highlighting the significance of feature engineering, ensemble learning techniques, and data-driven approaches. The selection of relevant input features, including thermodynamic and atomic parameters, plays a critical role in improving predictive accuracy. The continuous advancements in ML algorithms and their application to HEAs provide a foundation for further optimization in computational materials science.

2.5   Gaps in the literature

Most existing studies on ML-based phase prediction for HEAs primarily classify alloys into single-phase categories such as BCC, FCC, and IM phases. However, in reality, HEAs frequently exhibit complex multi-phase structures, including BCC+FCC, BCC+IM, FCC+IM, and even BCC+FCC+IM combinations. Current ML models often fail to accurately predict these multi-phase configurations, as they are typically trained on datasets that do not sufficiently represent multi-phase compositions. The lack of effective classification for multi-phase HEAs limits the applicability of these models in practical alloy design, where multi-phase structures play a crucial role in optimizing mechanical properties. Developing ML approaches that can robustly handle and classify multi-phase HEAs remains a critical research challenge.

This thesis aims to address this gap by developing an ML model that accurately predicts six distinct phase categories (BCC, FCC, BCC+FCC, BCC+IM, FCC+IM, BCC+FCC+IM), incorporating Boolean vector encoding to improve classification accuracy.
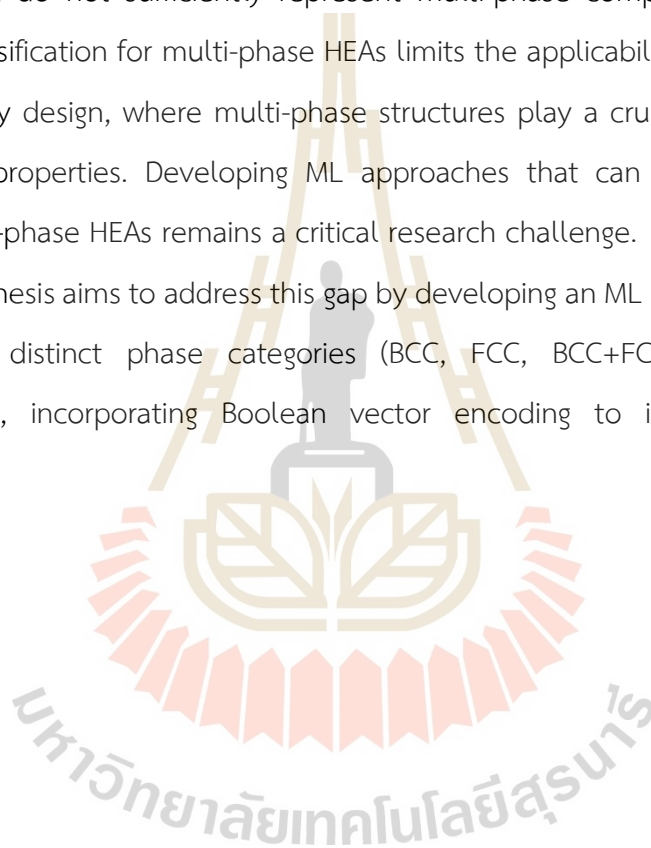
CHAPTER III

RESEARCH METHODOLOGY

This chapter describes the methodology used in this study to develop and evaluate ML models for predicting phase structures in HEAs. The research framework consists of dataset construction, data preprocessing, feature extraction, model selection, training and validation, and performance evaluation.

3.1 Research framework

Figure 3.1 outlines the workflow for this study, which consists of key stages:
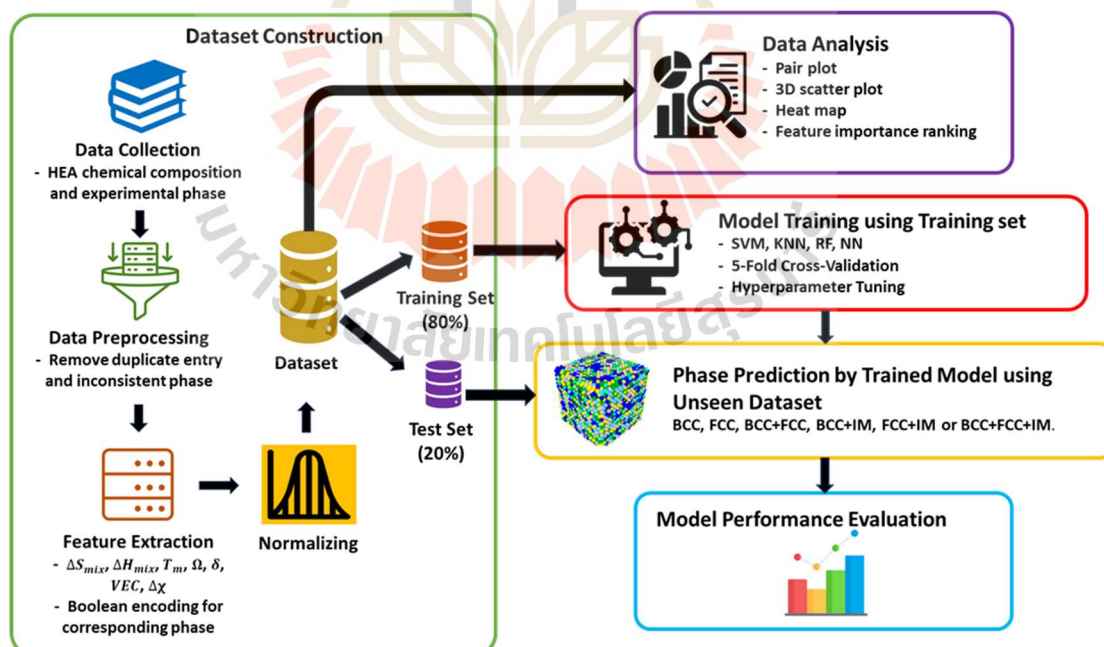
Figure 3.1 Comprehensive strategy for HEA phase prediction using ML.

Data collection: Gathering experimental HEA phase data from 34 literatures.

Feature extraction: Computing thermodynamic and atomic descriptors.

Data preprocessing: Remove duplicate sample and inconsistent phase.

Normalizing: Enhancing data quality by Standardizing the range of feature data.

Data analysis: Providing a detailed data analysis to understand feature correlations and their impact on phase classification.

Data splitting: Dividing data into two subsets at an 80:20 ratio for training and testing.

Model selection: Employing machine learning algorithms.

Model training and validation: Optimizing hyperparameters using CV.

Performance evaluation: Assessing predictive accuracy using multiple metrics.

3.2   Dataset construction

3.2.1   Data collection

The dataset was compiled from published experimental studies on 329 HEAs. The collected data included alloy compositions and experimentally verified phase structures. Only alloys synthesized via arc melting were considered to ensure consistency and reliability in the dataset. Vacuum arc melting process is one of the most commonly used synthesis methods for HEAs, especially in experimental phase studies. It provides a controlled environment with minimal contamination, and most reported phase structures from literature use vacuum arc melting in the as-cast condition. Other fabrication methods, such as powder metallurgy or mechanical alloying, could introduce variability in microstructure and phase stability due to different thermal histories, cooling rates, and impurity levels. By focusing on vacuum arc melting, the data remains uniform in process conditions, which helps the ML model learn patterns based on composition rather than processing-related variations. The detailed dataset sources can be found in appendix A1.

3.2.2 Feature extraction

The predictive models employed in this study utilize seven key features derived from the chemical compositions of HEAs, chosen based on their known correlation with phase stability. The dataset consists of numerical attributes extracted from the chemical composition of each alloy and its corresponding phase, calculated using the equations provide in Table 3.1. In these equations, R denotes the gas constant; c_i and c_j represent the atomic concentration of the ith and j^{th} components, respectively; ΔH_{ij} is the interaction parameter between i^{th} and j^{th} elements; T_{mi} corresponds to the melting temperature of the ith component; r_i signifies the atomic radius of the ith component, while $\bar{r}$ represents the average atomic radius; VEC_i refers to the valence electron concentration of the ith component; x_i denotes the Pauling electronegativity of the ith component, and $\bar{\chi}$ represents the average Pauling electronegativity. The atomic properties of each element for feature calculation can be found in appendix A2 and A3. The extracted features of each alloy can be found in appendix A4.
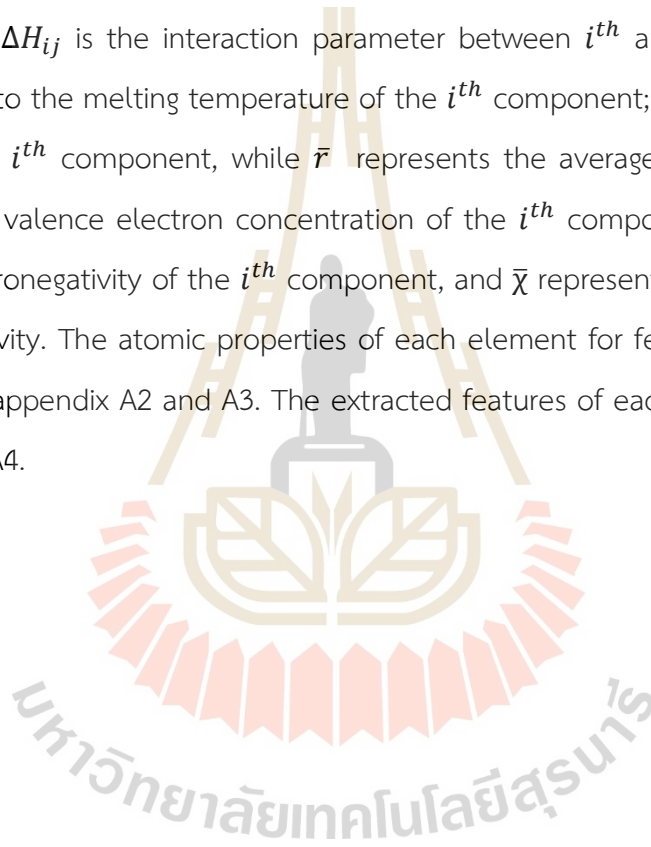
Table 3.1 Equations for feature extraction derived from chemical composition of HEAs.

Feature description	Unit	Equation	Reference
Mixing enthalpy	kJ/mol	$\Delta H_{mix} = \sum_{i=1, i \neq j}^{n} 4\Delta H_{ij}\, c_i c_j$	(L. Zhang et al., 2020)
Mixing entropy	J/(mol.K)	$\Delta S_{mix} = -R \sum_{i=1}^{n} c_i \ln c_i$	(Y. Zhang et al., 2008)
Melting temperature	K	$T_m = \sum_{i=1}^{n} c_i T_{mi}$	(Yang & Zhang, 2012)
Parameter for predicting solid solution formation	-	$\Omega = \dfrac{T_m \Delta S_{mix}}{\lvert \Delta H_{mix} \rvert}$	(Yang & Zhang, 2012)
Atomic size difference	%	$\delta = \sqrt{\sum_{i=1}^{n} c_i \left(1 - \dfrac{r_i}{\bar{r}}\right)}$	(Y. Zhang et al., 2008)
Valence electron concentration	-	$VEC = \sum_{i=1}^{n} c_i\, VEC_i$	(Guo et al., 2011)
Electronegativity difference	-	$\Delta \chi = \sqrt{\sum_{i=1}^{n} c_i \left(\chi_i - \bar{\chi}\right)}$	(Y. Zhang et al., 2020)

3.2.3   Boolean vector encoding

Since some HEAs exhibit multi-phase structures, a Boolean vector encoding scheme was implemented to handle multi-label classification. Each sample was assigned a binary vector indicating the presence (1) or absence (0) of three primary phase components: BCC, FCC and IM. For example:

[1,0,0] represents a single-phase BCC structure.

[1,1,0] represents a dual-phase BCC+FCC structure.

[1,1,1] represents a complex multi-phase BCC+FCC+IM structure.

3.2.4 Data normalization

Data normalization is a crucial preprocessing step in machine learning to ensure that all features contribute equally to the model. Since different features in the dataset have varying scales and magnitudes, normalization helps prevent any particular feature from dominating the learning process. This study employs Z-score normalization to standardize the dataset. Z-score normalization transforms each feature by subtracting the mean and dividing by the standard deviation, ensuring a mean of zero and a standard deviation of one. The transformation formula is as follows:

$$x_{new} = \frac{x_i - \bar{x}}{\sigma} \tag{3.1}$$

where:

x_i is the original feature value,

$\bar{x}$ is the mean of feature,

σ is the standard deviation of the feature.

This approach maintains the distribution characteristics of the original data while ensuring all features have comparable scales.

3.3 Machine learning models

3.3.1 Model selection

Machine learning (ML) algorithms, including SVM, KNN, RF, and NN, were utilized to predict HEA phases. These models were implemented using the scikit-learn library, chosen for their widespread use in similar research and proven effectiveness in classification tasks within materials science. Each algorithm has distinct advantages: SVM efficiently handles non-linear separability, KNN is a straightforward yet effective method for pattern recognition, RF is resistant to overfitting and provides

interpretability through feature importance analysis, while NN is adept at capturing intricate, non-linear relationships within the data.

1.   Support vector machine (SVM)

SVM determines the optimal hyperplane that separates different classes by maximizing the margin between them. This separation is achieved by minimizing a loss function while ensuring proper class distinction. To handle non-linear relationships, SVM applies kernel functions that transform data into a higher-dimensional space where linear separation may be feasible. The key hyperparameters in this study include the penalty parameter (C), the kernel function, and the kernel coefficient (gamma). A high C value may lead to overfitting, while a low C may fail to capture complex patterns. The kernel functions used in this study include radial basis function (RBF), polynomial, and sigmoid kernels. Each kernel captures different types of relationships in the data. Gamma controls the influence of individual data points. High values make the model highly sensitive to local variations, which may lead to overfitting. Low values result in underfitting by failing to capture data patterns effectively (Hastie et al., 2009; W. Zhang et al., 2023).

2.   K-nearest neighbors (KNN)

KNN predicts class labels by measuring the distance between a given data point and its nearest neighbors. Predictions are based on a majority vote. Key hyperparameters include the number of neighbors, weighting function, nearest neighbor search algorithm, and distance metric. A small number of neighbors makes the model sensitive to noise. A large number can introduce bias. Weighting options include uniform, where all neighbors have equal weight, and distance-based, where closer neighbors have higher weights. Various search algorithms are available, including Auto, ball tree, KD tree, and brute force. Euclidean and Manhattan distances serve as distance metrics to determine the similarity between data points (Ghouchan Nezhad Noor Nia et al., 2022; Z. Zhang, 2016).

3.   Random forest (RF)

RF is an ensemble learning method that combines multiple decision trees to improve prediction accuracy and reduce overfitting. Each tree is built using a subset of the data. The final classification is based on the majority vote of all trees. Important hyperparameters include the number of trees, maximum number of features, maximum tree depth, minimum samples per split, minimum samples per leaf, and bootstrap sampling. Increasing the number of trees improves model performance but also increases computational cost. The maximum number of features determines the subset of features considered at each split. A balance must be maintained to reduce the risk of overfitting while ensuring diversity. Limiting tree depth prevents overfitting. The minimum sample split and minimum sample leaf parameters define thresholds for splitting and leaf formation, controlling tree growth. When bootstrap sampling is enabled, trees are built using randomly drawn samples with replacement (Hastie et al., 2009).

4.   Neural network (NN).

NN is a feedforward deep learning model composed of an input layer, multiple hidden layers, and an output layer. Key hyperparameters include the number of hidden layers, number of neurons per layer, activation functions, loss function, optimizer, and learning rate. More hidden layers and neurons allow the model to capture complex patterns but also increase computational complexity and the risk of overfitting. Activation functions such as ReLU, sigmoid, and tanh introduce non-linearity, allowing the network to model intricate relationships in data. The loss function evaluates model performance. Optimizers such as stochastic gradient descent (SGD) and Adam adjust weights to minimize error. Learning rate controls the step size in weight updates. Low values slow training, while excessively high values can cause instability or suboptimal convergence (Dewangan et al., 2023).

3.3.2 Hyperparameter tuning

Hyperparameter tuning was conducted using a grid search combined with five-fold CV. The best performing hyperparameter values were selected based on

the highest mean CV accuracy. The key hyperparameter selection and range for each model are summarized in Appendix B1 and the optimal hyperparameter values can be found in Appendix B2.

3.4   Model training and evaluation

3.4.1   Training strategy

Each model was trained on the training dataset using the tuned hyperparameters. The models were then evaluated using the test dataset, and performance metrics were calculated.

3.4.2   Evaluation metrics

The performance of the ML models was assessed using four key metrics (Hastie et al., 2009):

1. Accuracy: Measures the proportion of correctly predicted samples.
2. Precision: Assesses the reliability of positive phase predictions.
3. Recall: Evaluates the ability to identify all relevant phase samples.
4. F1-score: A harmonic mean of precision and recall, balancing both metrics.

The formulas used for these metrics are:

$$Accuracy\ (\%) = \frac{TN + TP}{TP + TN + FN + FP} \times 100\% \tag{3.2}$$

$$Precision\ (\%) = \frac{TP}{FP + TP} \times 100\% \tag{3.3}$$

$$Recall\ (\%) = \frac{TP}{FN + TP} \times 100\% \tag{3.4}$$

$$F1 - score\ (\%) = \frac{Precision\ x\ Recall}{Precision + Recall} \times 100\% \qquad (3.5)$$

where:

True positives (TP) is an instance that is actually positive and correctly predicted,

True negatives (TN) is an instance that is actually negative and correctly predicted,

False positives (FP) is an instance that is actually negative and incorrectly predicted,

False negatives (FN) is an instance that is actually positive and incorrectly predicted.

3.4.3 Complex combination phase accuracy evaluation

To evaluate the model's capability in predicting complex multi-phase HEAs, a separate evaluation was conducted to assess the accuracy of phase combination predictions. Since traditional classification metrics do not fully capture the complexity of multi-phase structures, The strict accuracy criterion was applied. A prediction was considered correct only if all phase components matched exactly. The evaluation results were visualized using confusion matrices, demonstrating the model's effectiveness in predicting BCC, FCC, BCC+FCC, BCC+IM, FCC+IM and BCC+FCC+IM combinations.

This chapter detailed the methodology used in this study, including data collection, feature selection, preprocessing, machine learning model selection, hyperparameter tuning and performance evaluation. The models were validated using CV and various classification metrics, with additional evaluation focusing on complex combination phase accuracy. The next chapter presents the result obtained from these ML models and their analysis

CHAPTER IV

RESULTS AND DISCUSSION

This chapter presents the results obtained from ML models developed for predicting phase structures in HEAs. The performance of different ML models is analyzed based on key evaluation metrics and their predictive capabilities are compared. Additionally, a detailed data analysis is provided to understand feature correlations and their impact on phase classification. Finally, a discussion on complex combination phase accuracy and a comparison with existing studies are included.

4.1 Data analysis

4.1.1 Dataset distribution

The dataset used in this study consists of features extracted from their chemical composition and their corresponding phase classification, categorized into six distinct groups: BCC, FCC, BCC+IM, FCC+IM and BCC+FCC+IM (This information can be seen in Appendix A1). The dataset distribution is illustrated in Figure 4.1, showing the number of samples available for each phase category.
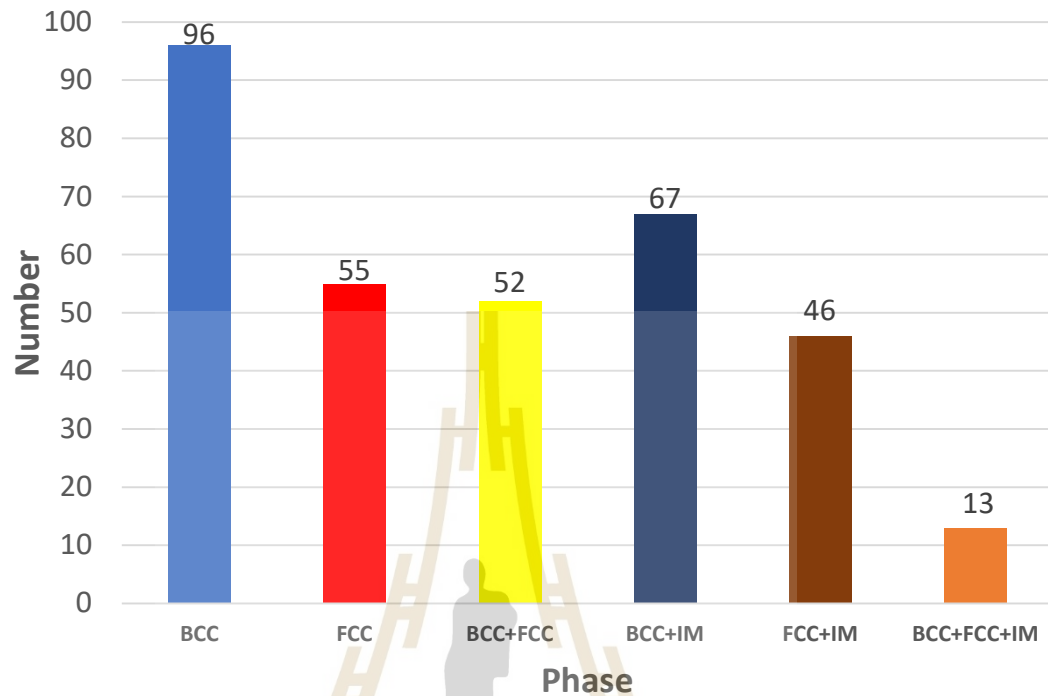
Figure 4.1 Number of samples for each phase category.

4.1.2 Feature correlation analysis

Figure 4.2 presents a heat map that visualizes the relationships or correlations between features within dataset. Each cell in the matrix represents a correlation coefficient between two features, with values ranking from -1 to 1. A coefficient of 1 signifies a strong positive correlation, while -1 indicates a strong negative correlation, and 0 represents no correlation between the features. The heat map values correspond to correlation coefficients, which quantify the strength and direction of the linear relationship between two variables. The most commonly used measure is the Pearson correlation coefficient, calculated using the formula:

$$r_{xy} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}} \tag{4.1}$$

where:

r_{xy} denotes the Pearson correlation coefficient between features x and y,

x_i and y_i are the individual data points indexed by i,

$\bar{x}$ and $\bar{y}$ represent the mean values of x and y, respectively.

The correlation values in the heat map for independent features range from -0.75 to 0.58. Notably, VEC and T_m exhibit a negative correlation, indicating that an increase in VEC is associated with a decrease in T_m. This trend aligns with phase stability, as higher VEC values are typically linked to the formation of FCC structures, while lower VEC values are more conducive to stabilizing BCC structures, given that FCC phases generally have lower T_m. In contrast, the $\Delta\chi$ and T_m display a positive correlation, suggesting that as $\Delta\chi$ increases, T_m also rises, elevated $\Delta\chi$ values tend to promote the formation of complex or IM phases, which often exhibit higher T_m compared to SS phases. Overall, no exceptionally strong positive or negative correlation were observed among the features, confirming that all selected features contribute valuable information and should be retained in the analysis.
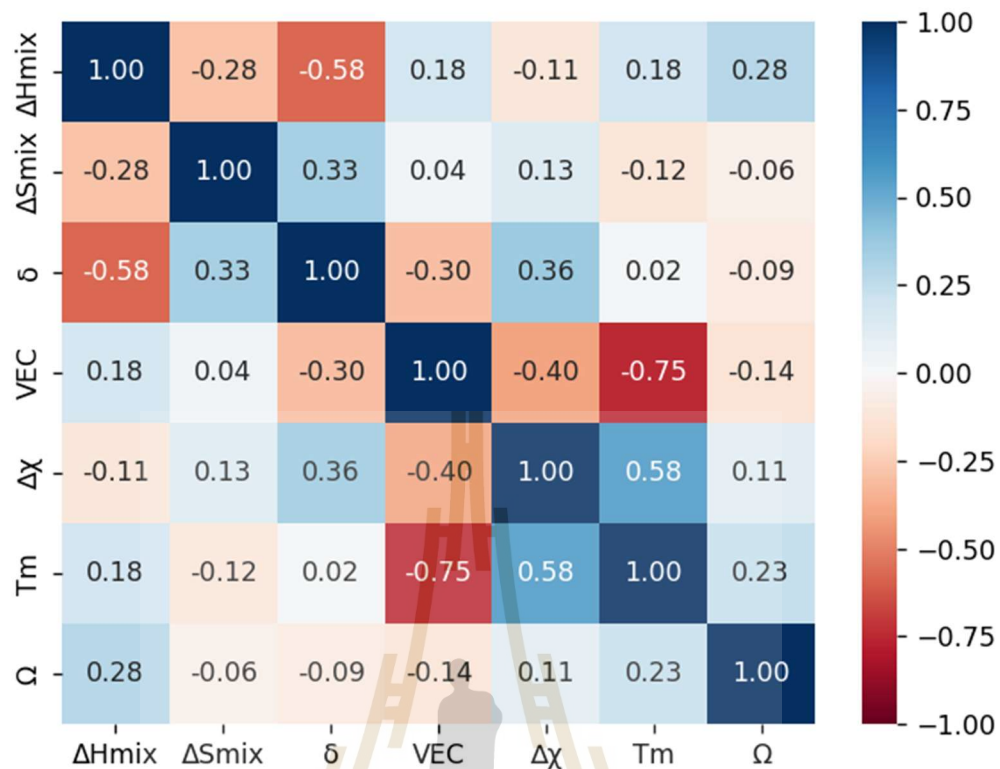
Figure 4.2 Pearson correlation coefficients between features.

4.1.3 Pair plot visualization

To further explore the relationships between individual features and phase classifications, a pair plot was generated (Figure 4.3). The scatter plots illustrate how certain features influence phase formation, with distinct clustering observed for different phase categories. The diagonal cells display kernel density plots that represent the distribution of each individual feature. This visualization aids in identifying key discriminative features for ML classification.
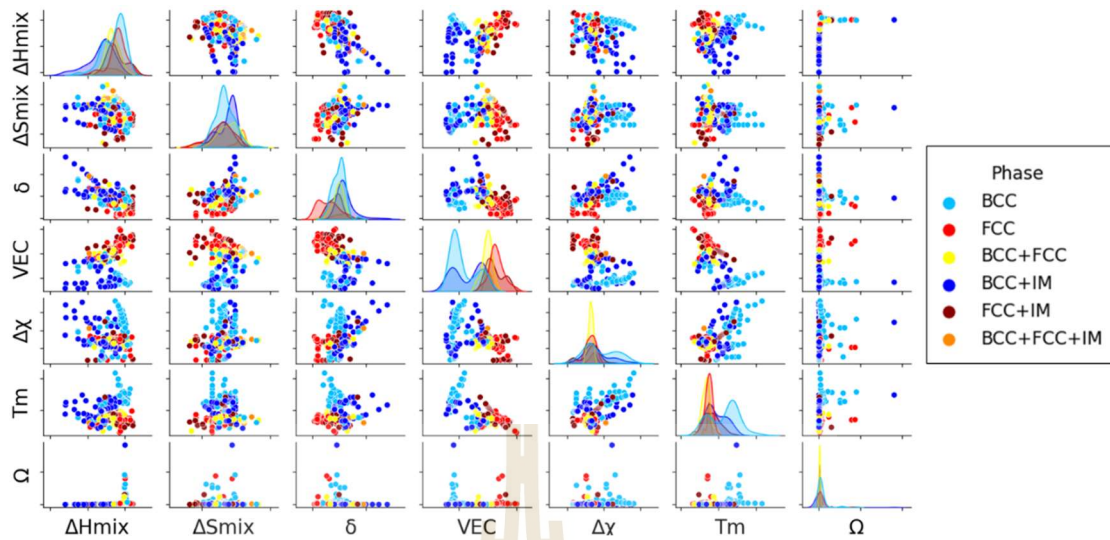
Figure 4.3 Pair plot of features for each phase category.

4.1.4 Radar plot for feature comparison

A radar plot was used to visualize the distribution of normalized feature values across different phase categories (Figure 4.4). The plot highlights key trends, such as lower δ and higher T_m in BCC phases, whereas FCC phases exhibit higher VEC values.
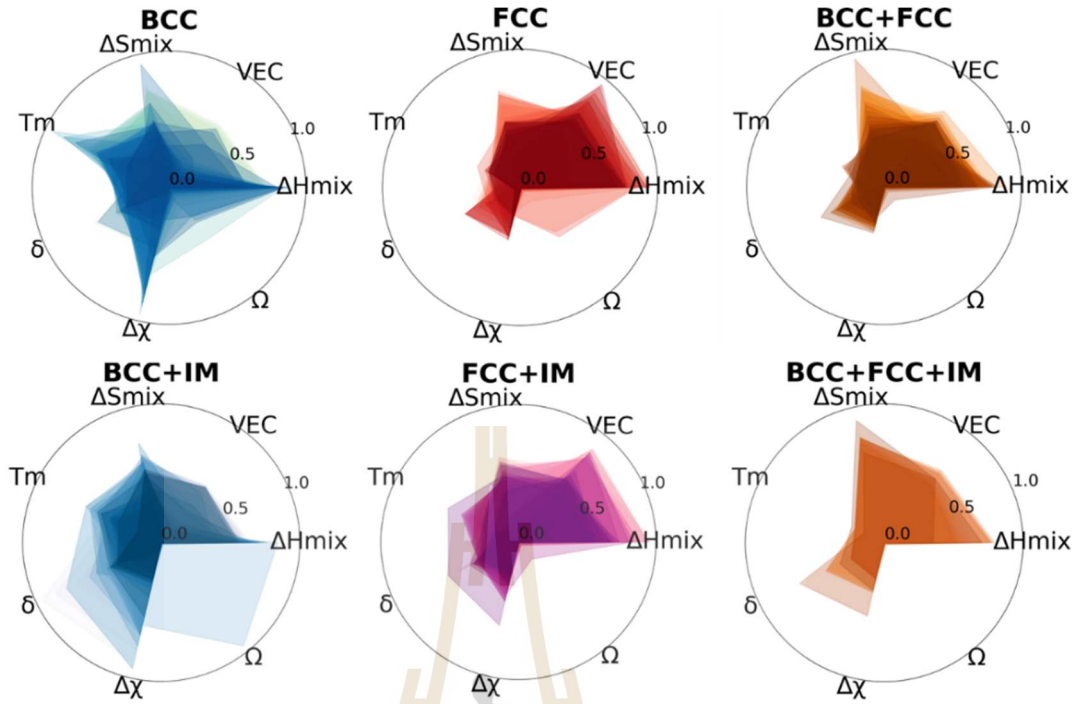
Figure 4.4 Radar plot of features for each phase categories.

4.1.5 3D scatter plot of key features

A 3D scatter plot (Figure 4.5) was generated to illustrate the distribution of HEA composition in terms of three key features: δ, ΔH_{mix} and VEC. The visualization shows distinct phase separations, reinforcing the importance of these feature in phase classification. IM phases exhibit higher δ compared to SS phases, such as BCC and FCC. The formation of IM phases is primarily influenced by atomic bonding between elements with considerable size differences. Leading to lattice distortions and structural strain, which contribute to their characteristic ordered structures. The thermodynamic stability of phases is governed by Gibbs free energy (ΔG), where a lower ΔH_{mix} corresponds to reduced ΔG, thereby promoting the formation of intermetallic compounds, In contrast, SS phases exhibit distinct trends, with higher VEC favoring the formation of FCC phases, while lower VEC is associated with the stabilization of BCC phases, Additionally, smaller δ values enhance the likelihood of

forming SS, whereas larger atomic size disparities increase the propensity for IM phase formation. These findings align with previous studies (Guo et al., 2011; Yang & Zhang, 2012; L. Zhang et al., 2020).
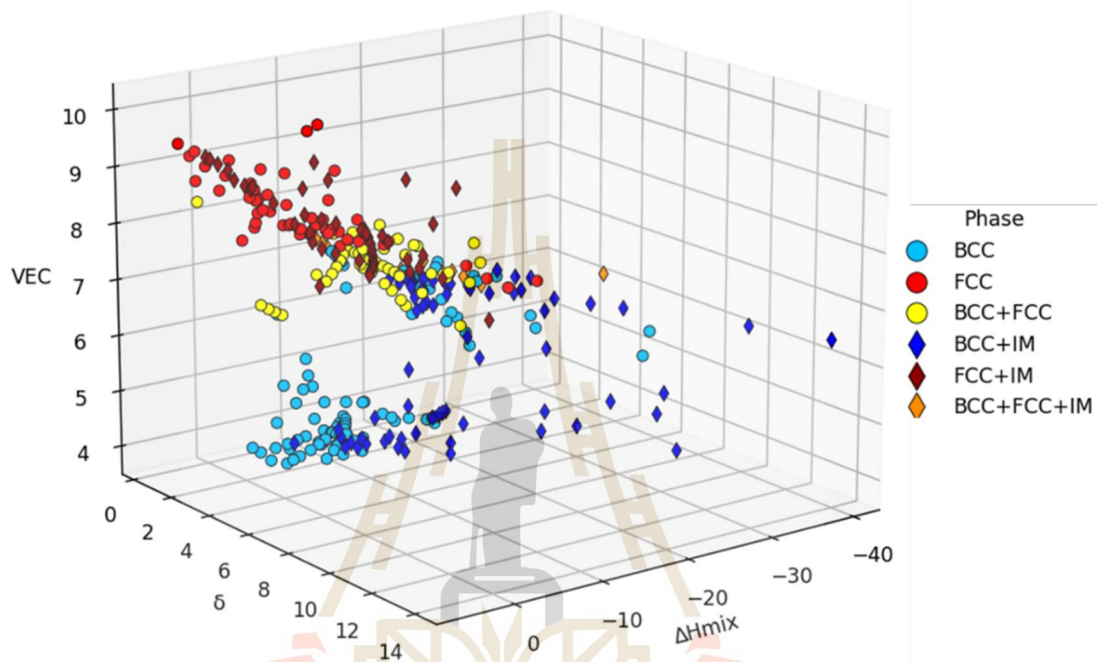
Figure 4.5 The 3D scatter plot represents phase distribution to show the influence of δ, ΔH_{mix} and VEC.

4.1.6 Feature importance analysis

Feature importance was evaluated to determine which factors most significantly influenced phase classification The RF, SVM and NN model were used to compute normalized importance score (Figure 4.6), VEC and T_m emerged as the most critical features, aligning with previous studies (He et al., 2024; Y. Zhang et al., 2014).
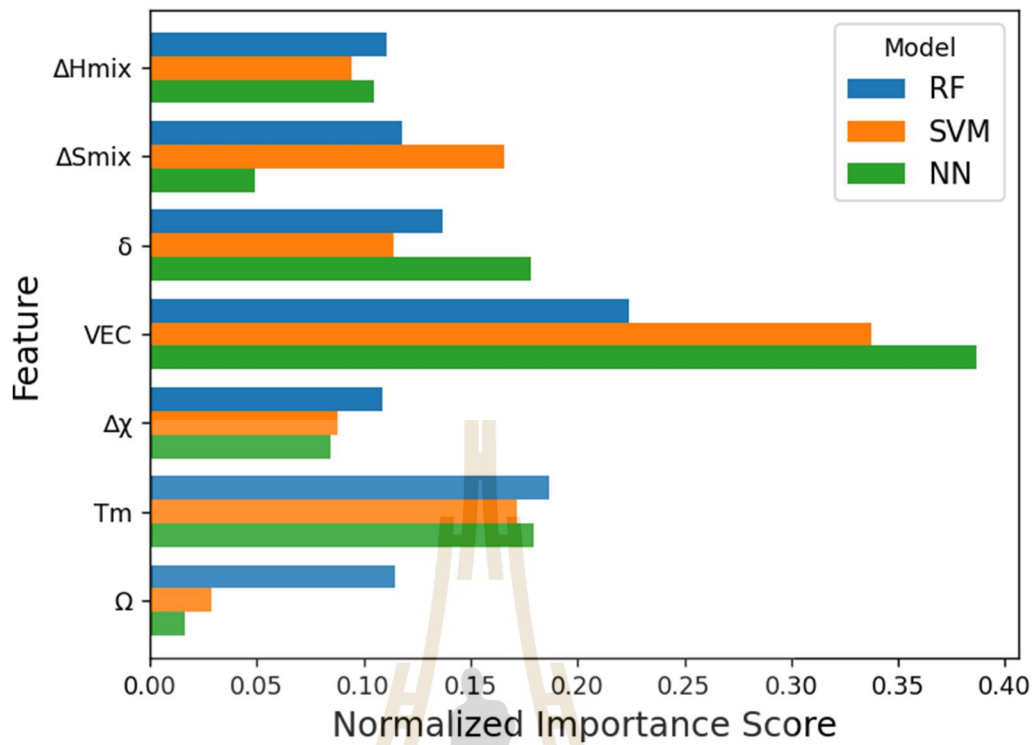
Figure 4.6 Feature importance score for HEA phase prediction by RF, SVM and NN.

4.2   ML model performance for phase prediction

4.2.1   CV and test accuracy for individual phase prediction

In this study, four machine learning algorithms (SVM, KNN, RF, and NN) were utilized to predict phase compositions in HEAs using a structured dataset. To enhance model generalizability and minimize overfitting, hyperparameter tuning was conducted through grid search in combination with layered CV. This approach involved training the models on four subsets of data while validating them on the fifth, with grid search identifying optimal parameter settings to improve performance. Table 4.1 presents a detailed evaluation of these models for each phase category (BCC, FCC, and IM) based on key performance metrics, including mean CV accuracy, test accuracy, precision, recall, and F1-score. A summary of CV and test accuracies in Figure 4.7 indicates that all models achieved an average CV accuracy exceeding 81%,

demonstrating their effectiveness in phase prediction for HEAs. The next section provides a phase-wise analysis of each model's performance.

Support vector machine (SVM)

1) BCC Phases: SVM demonstrates strong performance in classifying BCC phases, achieving a mean CV accuracy of 90.89% and a test accuracy of 92.42%. With precision at 93.15%, recall, and F1-score exceeding 90%, the model is highly effective at identifying BCC phases with minimal misclassifications.

2) FCC Phases: The SVM performs exceptionally well in distinguishing FCC phases, attaining a 96.97% test accuracy, with all key metrics reaching the same level. This suggests that the SVM model can accurately classify FCC phases, likely due to well-defined feature differences in the dataset.

3) IM Phases: Performance is slightly lower for IM phases, with a 79.81% mean CV accuracy and an 85.24% test accuracy. Precision and recall hover around 81–82%, indicating moderate classification accuracy for intermetallic phases compared to BCC and FCC.

K-Nearest Neighbors (KNN)

1) BCC Phases: KNN performs exceptionally well in classifying BCC phases, achieving a 92.79% mean CV accuracy and an outstanding 98.48% test accuracy. With precision at 97.83% and recall at 98.48%, KNN effectively differentiates BCC phases.

2) FCC Phases: The model also delivers strong results for FCC phases, attaining a 93.18% mean CV accuracy and a 96.97% test accuracy. The performance is comparable to that of SVM, confirming KNN's reliability in identifying FCC phases.

3) IM Phases: KNN outperforms SVM for IM phases, achieving a 83.27% mean CV accuracy and an 86.36% test accuracy.

Precision and F1-score range between 85–87%, indicating a well-balanced classification performance.

Random Forest (RF)

1) BCC Phases: RF achieves a 89.36% mean CV accuracy and a 95.45% test accuracy, with high precision (94.64%) and recall (96.50%). Although slightly less precise than KNN, RF remains a highly effective model for classifying BCC phases.

2) FCC Phases: With a 91.26% mean CV accuracy and 93.94% test accuracy, RF performs reliably for FCC classification. Precision, recall, and F1-score are all close to 94%, though slightly lower than the SVM and KNN models, suggesting room for improvement in learning FCC phase characteristics.

3) IM Phases: RF shows the weakest performance for IM phases, with 81.36% mean CV accuracy and 81.81% test accuracy. Precision and recall remain between 80–83%, indicating a challenge in accurately distinguishing IM phases, likely due to complex feature distributions or an imbalance in training data.

Neural Network (NN)

1) BCC Phases: NN achieves the highest mean CV accuracy for BCC phases at 94%, alongside a 95.45% test accuracy. With precision, recall, and F1-score all exceeding 94%, the model performs similarly to KNN and RF, showing strong predictive capability for BCC phases.

2) FCC Phases: NN delivers excellent performance in classifying FCC phases, with all evaluation metrics reaching 96.97%, aligning closely with the SVM and KNN models.

3) IM Phases: NN outperforms all other models in IM phase classification, achieving the highest mean CV accuracy (86%) and test accuracy (87.88%). Precision and F1-score are also high,

around 87–88%, making NN the most effective model for distinguishing IM phases among the four.

The results demonstrate that NN and KNN consistently achieve the highest accuracy for classifying BCC and FCC phases, with RF closely following. While SVM also performs well, it falls slightly behind in recall and F1-score for BCC phases. Among all models, IM phase classification exhibits the lowest accuracy, likely due to the complex and overlapping feature distributions associated with these phases. However, NN outperforms other models in this category, indicating its ability to capture intricate, nonlinear patterns in the dataset. These findings suggest that KNN and NN are highly reliable for predicting BCC and FCC phases, while NN is the most effective for distinguishing IM phases, making it a strong candidate for analyzing complex HEA compositions with multiple phase formations.
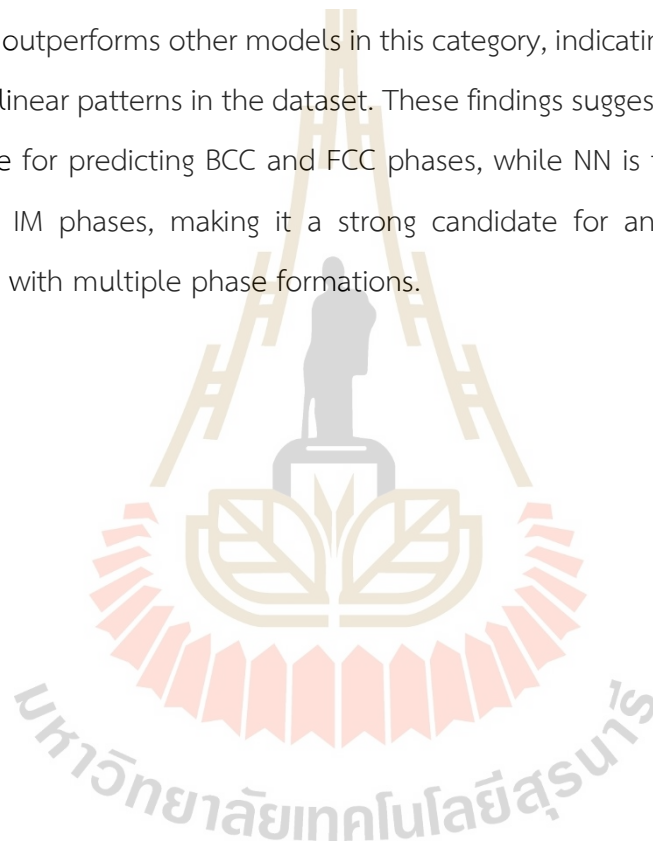
Table 4.1 Performance metrics of ML models for individual phase prediction in HEAs.

Model	Phase	Mean CV	Test	Precision (%)	Recall (%)	F1-score (%)
SVM	BCC	90.89	92.42	93.15	90.14	91.38
	FCC	92.01	96.97	96.97	96.97	96.97
	IM	79.81	85.24	81.30	82.15	82.15
KNN	BCC	92.79	98.48	97.83	98.48	98.48
	FCC	93.18	96.97	96.97	96.97	96.97
	IM	83.27	86.36	86.68	85.34	85.81
RF	BCC	89.36	95.45	94.64	95.50	95.04
	FCC	91.26	93.94	94.10	93.94	93.93
	IM	81.36	81.81	82.95	79.98	81.39
NN	BCC	94.00	95.45	95.45	94.49	94.94
	FCC	94.00	96.97	96.97	96.97	96.97
	IM	86.00	87.88	87.98	87.12	87.46
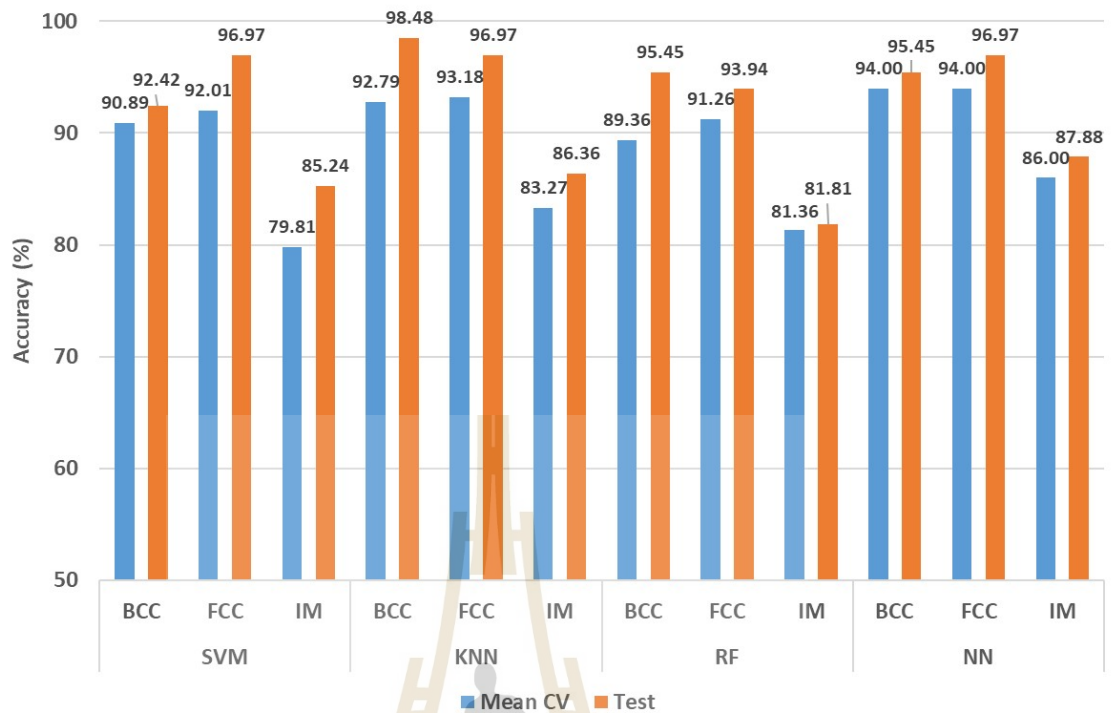
Figure 4.7 CV and test accuracy of each model for individual phase prediction.

To further evaluate the performance of the NN model, which achieved the highest CV accuracy, additional assessment metrics were utilized, including the precision–recall (PR) and receiver operating characteristic (ROC) curves. These curves provide a visual representation of the model's predictive capabilities and classification performance.

The PR curve, shown in Figure 4.8a, illustrates the relationship between precision and recall, with a larger area under the curve (AUC) indicating better predictive accuracy. For the NN model, the PR AUC values for BCC (99%) and FCC (100%) demonstrate excellent classification performance for these phases. In contrast, the IM phase achieved a PR AUC of 82%, which, while reasonable, is lower than the other two phases. This suggests that the model faces greater challenges in correctly identifying IM phases, likely due to feature overlap or variations in data representation.

Similarly, the ROC curve in Figure 4.8b is used to assess the classifier's ability to distinguish between different phases. The AUC of the ROC curve indicates the effectiveness of the model in differentiating between positive and negative classes. The diagonal line represents the baseline performance of a random classifier (50% AUC), serving as a reference point. A model performing significantly above this line indicates strong classification ability. The NN model exhibits high AUC values for BCC (99%) and FCC (100%), confirming its reliability in predicting these phases. However, the IM phase achieves a slightly lower AUC of 89%, suggesting some difficulty in its classification, possibly due to a smaller number of training samples or overlapping phase characteristics.
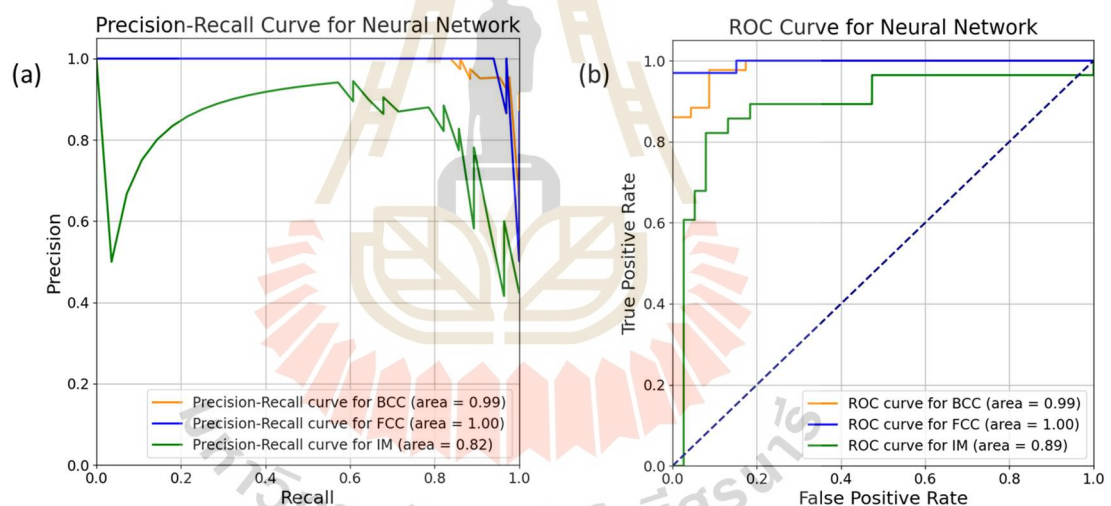
Figure 4.8 (a) PR curve for BCC, FCC and IM phases, demonstrating the trade-off between precision and recall for each phase. (b) ROC curves for BCC, FCC and IM phases, depicting the relationship between the true positive rate and false positive rate.

4.2.2 Complex combination phase prediction accuracy evaluation

The previous sections analyzed the performance of each model for individual phases; however, real-world applications require accurate predictions of complex phase combinations in HEAs. To address this, a strict evaluation criterion was implemented, where predictions were considered accurate only if all phase components in an alloy were correctly identified. Any partially or entirely incorrect predictions were excluded from the accuracy calculation. This approach was necessary due to the multi-phase nature of HEAs, where precise identification of all phases within a composition is essential.

As shown in Figure 4.9, the accuracy of predicting complex phase combinations varies across models, ranging from 75.75% to 84.85%. This stringent evaluation resulted in slightly lower test accuracies compared to the mean CV accuracy reported earlier, which is expected given the increased difficulty of predicting multiple phases simultaneously. Among the models, NN and KNN achieved the highest accuracy of 84.85%, demonstrating their strong capability in capturing nonlinear interactions among features and effectively handling multi-phase compositions. In contrast, SVM and RF exhibited moderate performance, with accuracies of 77.27% and 75.75%, respectively. This suggests that these models may struggle with complex phase predictions, likely due to overlapping feature distributions and intricate phase interactions within the dataset.
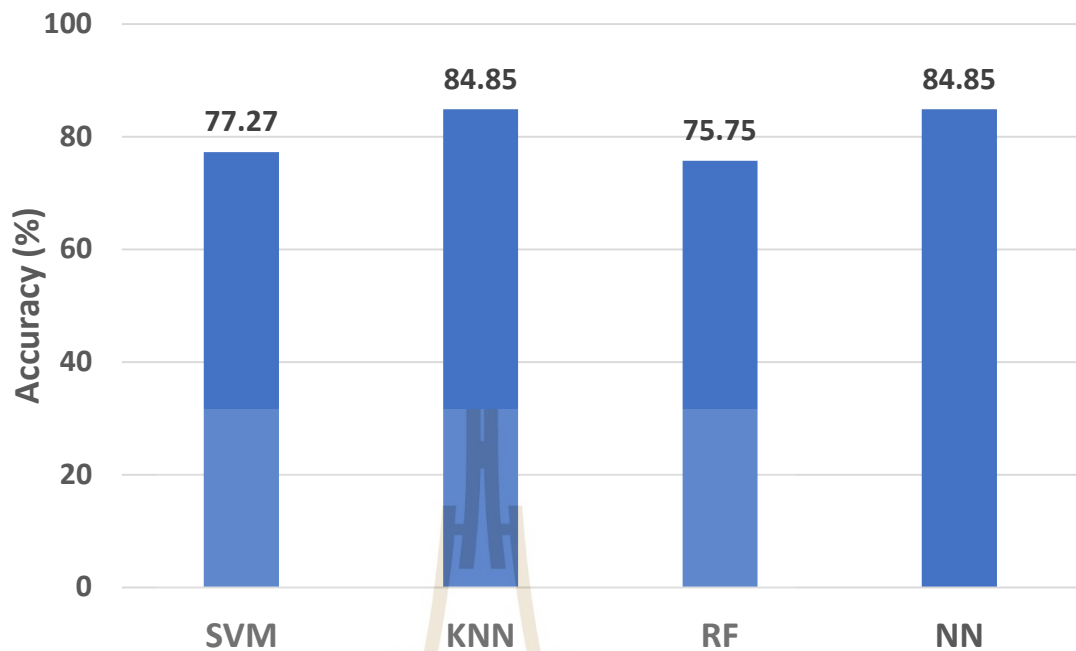
Figure 4.9 Complex combination phase prediction accuracy of each model.

The confusion matrices in Figure 4.10 provide an overview of the classification performance of all models in predicting HEA phase compositions across six categories: BCC, FCC, BCC+FCC, BCC+IM, FCC+IM, and BCC+FCC+IM. Among the models, NN and KNN achieved the highest accuracy, particularly in identifying BCC (NN: 17, KNN: 19 correct classifications) and FCC phases (NN/KNN: 9 correct classifications), with minimal misclassifications.

SVM and RF demonstrated moderate performance but struggled with multi-phase classifications due to overlapping feature distributions. The classification of BCC+FCC was most accurately predicted by KNN and NN, both achieving seven correct classifications, while SVM and RF exhibited difficulties in differentiating between binary-phase structures. NN performed exceptionally well in predicting BCC+IM and FCC+IM phases, correctly classifying 12 and 9 samples, respectively, underscoring its strength in capturing nonlinear interactions.

Misclassifications were more frequent in multi-phase categories, indicating the challenge of distinguishing overlapping phase features, which could be addressed through improved feature selection and dataset balancing. Overall, NN emerged as the most effective model for multi-phase classification, while KNN was the most reliable for single-phase predictions.
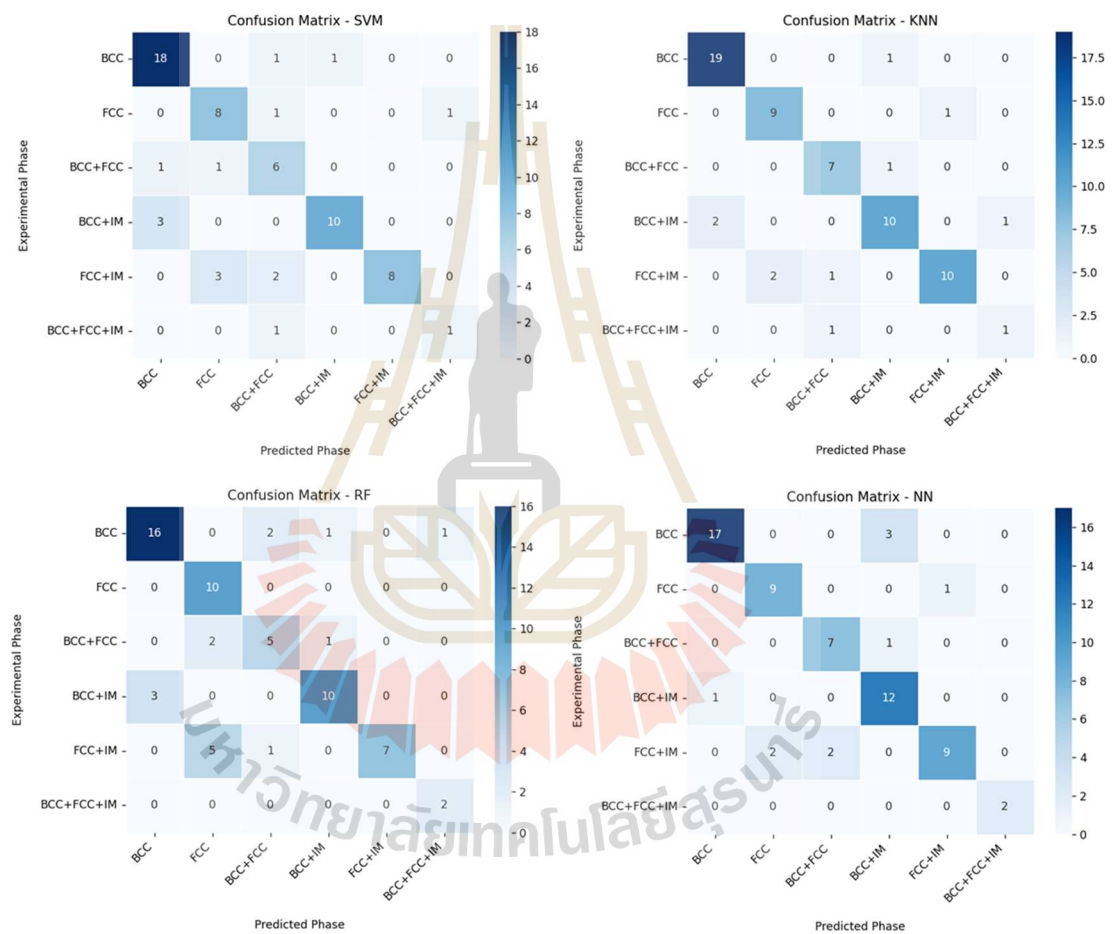
Figure 4.10 Confusion matrices for each ML model in complex combination phase prediction.

4.3   Discussion

4.3.1   Comparison with previous studies

Previous research primarily focused on classifying HEA phases into binary or ternary categories, such as distinguishing between BCC, FCC, and mixed phases. In contrast, this study expands phase classification to six distinct categories: BCC, FCC, BCC + FCC, BCC + IM, FCC + IM, and BCC + FCC + IM. Earlier studies (Dai et al., 2020; W. Huang et al., 2019; Islam et al., 2018) primarily classified HEAs into broader categories (BCC, FCC, and IM), limiting their capacity to predict complex multi-phase compositions. Additionally, RF, LDA, and SVM for SS classification was employed (Machaka, 2021) but did not explicitly consider the IM phase. This study introduces a Boolean vector encoding approach that enables multi-phase classification by treating each phase as an independent binary variable. This technique significantly improves predictive accuracy, particularly for intricate phase combinations such as BCC+IM and BCC+FCC+IM, which were often not addressed in prior research.

4.3.2   Challenges in IM phase prediction

The slightly lower accuracy for IM phases across all models suggests that IM formation is influenced by more complex and overlapping feature relationships. Future improvements could involve incorporating additional features or ensemble learning techniques.

4.3.3   Implications for HEA design

The findings indicate that ML models can significantly reduce reliance on experimental trial-and-error methods, enabling faster HEA design. The ability of NN and KNN to accurately classify both single-phase and multi-phase HEAs highlights their potential for computational alloy design.

4.4   Summary

This chapter presented the evaluation of ML models in predicting HEA phases. The NN and KNN models consistently outperformed others, particularly in predicting

complex phase combinations. These results demonstrate the feasibility of ML-driven phase prediction, providing valuable insights for HEA development. The next chapter will conclude the study and suggest future research directions.

CHAPTER V

CONCLUSION AND FUTURE WORK

5.1   Summary of findings

This study explored the application of ML models for phase prediction in high entropy alloys (HEAs). The research focused on utilizing experimentally validated datasets and computational techniques to enhance phase classification accuracy. Key contributions of this study include:

1. Feature selection and extraction

   Seven key compositional and thermodynamic features were identified and used as input variables for ML models. These features included ΔH_{mix}, ΔS_{mix}, δ, VEC, T_m, $\Delta\chi$, and Ω.

2. ML model implementation

   Four ML algorithms (SVM, KNN, RF, and NN) were developed and optimized using hyperparameter tuning and cross-validation.

3. Performance evaluation

   The models were assessed using accuracy, precision, recall, and F1-score metrics. The NN and KNN models consistently demonstrated the highest performance, particularly in distinguishing single-phase and multi-phase HEAs.

4. Complex phase prediction

   A strict accuracy criterion was applied to evaluate the models' ability to predict complex phase combinations. The NN model achieved the highest prediction accuracy (84.85%), demonstrating its capability in handling intricate phase formations in HEAs.

5. Feature importance analysis

The study identified VEC and T_m as the most influential features affecting phase stability, aligning with established phase selection criteria in HEAs.

5.2 Key contributions

The findings of this research contribute to the advancement of ML applications in computational materials science, particularly in the following aspects:

1. Development of an ML-based HEA phase prediction framework

   This study establishes a systematic approach for using ML techniques to classify and predict HEA phases, reducing dependency on experimental trial-and-error methods.

2. Boolean vector encoding for multi-phase classification

   The implementation of Boolean vector encoding for multi-label classification enables the accurate representation of complex phase compositions.

3. Performance benchmarking of ML algorithms

   Comparative evaluation of different ML models highlights the strengths and limitations of each approach in predicting HEA phase stability.

4. Guidance for HEA design

   The study provides valuable insights into the relationship between alloy composition and phase formation, assisting materials scientists in designing novel HEAs with desired properties.

5.3 Limitations

Despite its promising results, this study has several limitations:

1. Dataset size and generalizability

The dataset was compiled from published experimental studies, limiting its size and diversity. Future work should focus on expanding the dataset with more compositions and synthesis conditions.

2. Feature engineering constraints

While the selected features have demonstrated relevance, additional microstructural and thermodynamic parameters could further improve model accuracy.

3. Computational complexity

The NN model, while achieving the best performance, requires higher computational resources and longer training times compared to other ML models.

5.4 Future work

Future research should address these limitations and explore the following directions:

1. Integration of additional data sources

Expanding the dataset by incorporating high-throughput computational and experimental data will enhance model robustness.

2. Discover new features for improved IM phase prediction

Focusing on identifying and integrating new descriptors that are more sensitive to IM formation. This may include advanced thermodynamic parameters, electronic structure features.

3. Hybrid ML models

Combining multiple ML approaches, such as ensemble methods or deep learning architectures, could improve predictive accuracy and generalizability.

4. Experimental validation

   Experimental synthesis and characterization of ML-predicted HEA compositions can validate the model's predictions and refine its performance.

5. Real-Time HEA design tools

   Developing an interactive web-based platform or software integrating ML models can facilitate real-time HEA design and phase prediction for researchers and engineers.

5.5 Conclusion

This study demonstrates the potential of machine learning in accelerating HEA phase prediction and design. The results confirm that ML models, particularly NN and KNN, can effectively classify and predict phase structures based on composition-derived features. The findings contribute to the ongoing efforts to integrate computational intelligence with materials science, paving the way for more efficient and systematic HEA development. By addressing existing challenges and leveraging emerging ML advancements, future research can further optimize HEA design and discovery, driving progress in high-performance alloy applications.
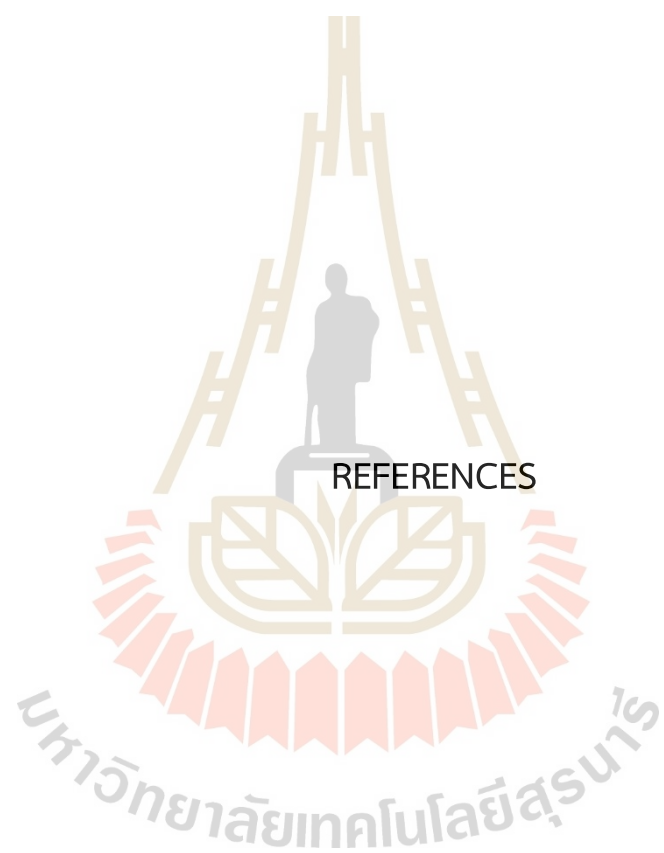
REFERENCES

REFERENCES

Chen, B., Li, X., Niu, Y., Yang, R., Chen, W., Yusupu, B., & Jia, L. (2023). A dual-phase CrFeNbTiMo refractory high entropy alloy with excellent hardness and strength. *Materials Letters*, *337*, 133958. https://doi.org/10.1016/j.matlet.2023.133958

Chen, J., Zhou, X., Wang, W., Liu, B., Lv, Y., Yang, W., Xu, D., & Liu, Y. (2018). A review on fundamental of high entropy alloys with promising high–temperature properties. *Journal of Alloys and Compounds*, *760*, 15–30. https://doi.org/10.1016/j.jallcom.2018.05.067

Dai, D., Xu, T., Wei, X., Ding, G., Xu, Y., Zhang, J., & Zhang, H. (2020). Using machine learning and feature engineering to characterize limited material datasets of high-entropy alloys. *Computational Materials Science*, *175*, 109618. https://doi.org/10.1016/j.commatsci.2020.109618

Dewangan, S. K., Nagarjuna, C., Jain, R., Kumawat, R. L., Kumar, V., Sharma, A., & Ahn, B. (2023). Review on applications of artificial neural networks to develop high entropy alloys: A state-of-the-art technique. *Materials Today Communications*, *37*, 107298. https://doi.org/10.1016/j.mtcomm.2023.107298

Gao, J., Wang, Y., Hou, J., You, J., Qiu, K., Zhang, S., & Wang, J. (2023). Phase Prediction and Visualized Design Process of High Entropy Alloys via Machine Learned Methodology. *Metals, 13*(2), 283. https://doi.org/10.3390/met13020283

Gao, M. C., Zhang, C., Gao, P., Zhang, F., Ouyang, L. Z., Widom, M., & Hawk, J. A. (2017). Thermodynamics of concentrated solid solution alloys. *Current Opinion in Solid State and Materials Science*, *21*(5), 238–251. https://doi.org/10.1016/j.cossms.2017.08.001

Ghouchan Nezhad Noor Nia, R., Jalali, M., & Houshmand, M. (2022). A Graph-Based k-Nearest Neighbor (KNN) Approach for Predicting Phases in High-Entropy Alloys. *Applied Sciences*, *12*(16), 8021. https://doi.org/10.3390/app12168021

Gong, J., Lu, W., Li, Y., Liang, S., Wang, Y., & Chen, Z. (2024). A single-phase
   Nb25Ti35V5Zr35 refractory high-entropy alloy with excellent strength-ductility
   synergy. *Journal of Alloys and Compounds*, *1006*, 176290.
   https://doi.org/10.1016/j.jallcom.2024.176290

Gorsse, S., Nguyen, M. H., Senkov, O. N., & Miracle, D. B. (2018). Database on the
   mechanical properties of high entropy alloys and complex concentrated
   alloys. *Data in Brief*, *21*, 2664–2678. https://doi.org/10.1016/j.dib.2018.11.111

Guo, S., Hu, Q., Ng, C., & Liu, C. T. (2013). More than entropy in high-entropy alloys:
   Forming solid solutions or amorphous phase. *Intermetallics*, *41*, 96–103.
   https://doi.org/10.1016/j.intermet.2013.05.002

Guo, S., Ng, C., Lu, J., & Liu, C. T. (2011). Effect of valence electron concentration on
   stability of fcc or bcc phase in high entropy alloys. *Journal of Applied Physics*,
   109(10), 103505. https://doi.org/10.1063/1.3587228

Han, P., Wang, J., & Li, H. (2024). Ultrahigh strength and ductility combination in
   Al40Cr15Fe15Co15Ni15 triple-phase high entropy alloy. *Intermetallics*, *164*,
   108118. https://doi.org/10.1016/j.intermet.2023.108118

Hastie, T., Tibshirani, R., & Friedman, J. H. (2009). *The elements of statistical learning:
   Data mining, inference, and prediction* (2nd ed). Springer.

He, Z., Zhang, H., Cheng, H., Ge, M., Si, T., Che, L., Zheng, K., Zeng, L., & Wang, Q.
   (2024). Machine learning guided BCC or FCC phase prediction in high entropy
   alloys. *Journal of Materials Research and Technology*, *29*, 3477–3486.
   https://doi.org/10.1016/j.jmrt.2024.01.257

Hu, H., Li, F., Liu, M., Yang, D., Zhu, Z., Shen, Q., & Yu, Z. (2025). High temperature
   oxidation behavior of the dual-phase AlCoCr0.5Fe2.5Ni2.5 and single phase
   Al0.25CoCrFeNi high entropy alloys. *Intermetallics*, *176*, 108553.
   https://doi.org/10.1016/j.intermet.2024.108553

Huang, W., Martin, P., & Zhuang, H. L. (2019). Machine-learning phase prediction of
   high-entropy alloys. *Acta Materialia*, *169*, 225–236.
   https://doi.org/10.1016/j.actamat.2019.03.012

Huang, X., Wang, H., Xue, W., Xiang, S., Huang, H., Meng, L., Ma, G., Ullah, A., & Zhang, G. (2020). Study on time-temperature-transformation diagrams of stainless steel using machine-learning approach. *Computational Materials Science*, *171*, 109282. https://doi.org/10.1016/j.commatsci.2019.109282

Islam, N., Huang, W., & Zhuang, H. L. (2018). Machine learning for phase selection in multi-principal element alloys. *Computational Materials Science*, *150*, 230–235. https://doi.org/10.1016/j.commatsci.2018.04.003

Jiang, S., Lin, Z., Xu, H., & Sun, Y. (2018). Studies on the microstructure and properties of AlxCoCrFeNiTi1-x high entropy alloys. *Journal of Alloys and Compounds*, *741*, 826–833. https://doi.org/10.1016/j.jallcom.2018.01.247

Kim, Y. S., Ozasa, R., Sato, K., Gokcekaya, O., & Nakano, T. (2024). Design and development of a novel non-equiatomic Ti-Nb-Mo-Ta-W refractory high entropy alloy with a single-phase body-centered cubic structure. *Scripta Materialia*, *252*, 116260. https://doi.org/10.1016/j.scriptamat.2024.116260

Li, J., Zhou, G., Han, J., Peng, Y., Zhang, H., Zhang, S., Chen, L., & Cao, X. (2023). Dynamic recrystallization behavior of single-phase BCC structure AlFeCoNiMo0.2 high-entropy alloy. *Journal of Materials Research and Technology*, *23*, 4376–4384. https://doi.org/10.1016/j.jmrt.2023.02.074

Liang, J., Li, G., Ding, X., Li, Y., Wen, Z., Zhang, T., & Qu, Y. (2024). The synergistic effect of Ni and C14 Laves phase on the hydrogen storage properties of TiVZrNbNi high entropy hydrogen storage alloy. *Intermetallics*, *164*, 108102. https://doi.org/10.1016/j.intermet.2023.108102

Liu, X., Feng, S., Xu, H., Liu, C., An, X., Chu, Z., Wei, W., Wang, D., Lu, Y., Jiang, Z., & Kong, J. (2024). A novel cast Co68Al18.2Fe6.5V4.75Cr2.55 dual-phase medium entropy alloy with superior high-temperature performance. *Intermetallics*, *169*, 108301. https://doi.org/10.1016/j.intermet.2024.108301

Liu, X., Liu, H., Wu, Y., Li, M., Xing, C., & He, Y. (2024). Tailoring phase transformation and precipitation features in a Al21Co19.5Fe9.5Ni50 eutectic high-entropy alloy to achieve different strength-ductility combinations. *Journal of Materials*

Science & Technology, *195*, 111–125.
https://doi.org/10.1016/j.jmst.2024.01.044

Liu, Y., Yen, S., Chu, S., Lin, S., & Tsai, M.-H. (2021). Mechanical and thermodynamic
data-driven design of Al-Co-Cr-Fe-Ni multi-principal element alloys. *Materials
Today Communications*, *26*, 102096.
https://doi.org/10.1016/j.mtcomm.2021.102096

Machaka, R. (2021). Machine learning-based prediction of phases in high-entropy
alloys. *Computational Materials Science*, *188*, 110244.
https://doi.org/10.1016/j.commatsci.2020.110244

Mulewicz, B., Korpala, G., Kusiak, J., & Prahl, U. (2019). Autonomous Interpretation of
the Microstructure of Steels and Special Alloys. *Materials Science Forum*, *949*,
24–31. https://doi.org/10.4028/www.scientific.net/MSF.949.24

Olorundaisi, E., Babalola, B. J., Anamu, U. S., Teffo, M. L., Kibambe, N. M., Ogunmefun,
A. O., Odetola, P., & Olubambi, P. A. (2024). Thermo-mechanical and phase
prediction of Ni25Al25Co14Fe14Ti9Mn8Cr5 high entropy alloys system using
THERMO-CALC. *Manufacturing Letters*, *41*, 160–169.
https://doi.org/10.1016/j.mfglet.2024.09.020

Potnis, G., & Das, J. (2024). Effect of cooling rate and Nb addition on the
strengthening in eutectic CoCr1.3FeMnNi0.7Nbx ($0 \leq x \leq 0.45$) high entropy
alloys comprising of low stacking fault energy FCC phase. *Materials Science
and Engineering: A*, *899*, 146467. https://doi.org/10.1016/j.msea.2024.146467

Ren, H., Chen, R. R., Gao, X. F., Liu, T., Qin, G., Wu, S. P., & Guo, J. J. (2023).
Development of wear-resistant dual-phase high-entropy alloys enhanced by
C15 Laves phase. *Materials Characterization*, *200*, 112879.
https://doi.org/10.1016/j.matchar.2023.112879

Roy, A., Babuska, T., Krick, B., & Balasubramanian, G. (2020). Machine learned feature
identification for predicting phase and Young's modulus of low-, medium-
and high-entropy alloys. *Scripta Materialia*, *185*, 152–158.
https://doi.org/10.1016/j.scriptamat.2020.04.016

Saboktakin Rizi, M., Ebrahimian, M., Minouei, H., Shim, S. H., Pouraliakbar, H., Fallah, V., Park, N., & Hong, S. I. (2024). Enhancing mechanical properties in Ti-Containing FeMn40Co10Cr10C0.5 High-Entropy alloy through Chi (χ) phase dissolution and precipitation hardening. *Materials Letters, 377,* 137516. https://doi.org/10.1016/j.matlet.2024.137516

Shafiei, A., Khani Moghanaki, S., & Amirjan, M. (2023). Effect of heat treatment on the microstructure and mechanical properties of a dual phase Al14Co41Cr15Fe10Ni20 high entropy alloy. *Journal of Materials Research and Technology, 26,* 2419–2431. https://doi.org/10.1016/j.jmrt.2023.08.071

Shi, Y., Yang, B., & Liaw, P. (2017). Corrosion-Resistant High-Entropy Alloys: A Review. *Metals, 7*(2), 43. https://doi.org/10.3390/met7020043

Singh, P., Smirnov, A. V., Alam, A., & Johnson, D. D. (2020). First-principles prediction of incipient order in arbitrary high-entropy alloys: Exemplified in Ti0.25CrFeNiAl. *Acta Materialia, 189,* 248–254. https://doi.org/10.1016/j.actamat.2020.02.063

Song, R., Ye, F., Yang, C., & Wu, S. (2018). Effect of alloying elements on microstructure, mechanical and damping properties of Cr-Mn-Fe-V-Cu high-entropy alloys. *Journal of Materials Science & Technology, 34*(11), 2014–2021. https://doi.org/10.1016/j.jmst.2018.02.026

Sun, Y., Wang, Z., Zhao, X., Liu, Z., & Cao, F. (2023). Effects of Sc addition on microstructure, phase evolution and mechanical properties of Al0.2CoCrFeNi high-entropy alloys. *Transactions of Nonferrous Metals Society of China, 33*(12), 3756–3769. https://doi.org/10.1016/S1003-6326(23)66368-X

Syarif, J., Elbeltagy, M. B., & Nassif, A. B. (2023). A machine learning framework for discovering high entropy alloys phase formation drivers. *Heliyon, 9*(1), e12859. https://doi.org/10.1016/j.heliyon.2023.e12859

Takeuchi, A., & Inoue, A. (2005). 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. *MATERIALS TRANSACTIONS*, *46*(12), 2817–2829. https://doi.org/10.2320/matertrans.46.2817

Tamuly, S., Dixit, S., Kombaiah, B., Parameswaran, V., & Khanikar, P. (2023). High strain rate deformation behavior of Al0.65CoCrFe2Ni dual-phase high entropy alloy. *Intermetallics*, *161*, 107983. https://doi.org/10.1016/j.intermet.2023.107983

Vaghari, M., & Dehghani, K. (2023). Computational and experimental investigation of a new non equiatomic FCC single-phase Cr15Cu5Fe20Mn25Ni35 high-entropy alloy. *Physica B: Condensed Matter*, *671*, 415413. https://doi.org/10.1016/j.physb.2023.415413

Wagner, C., George, E. P., & Laplanche, G. (2025). Effects of grain size and stacking fault energy on twinning stresses of single-phase Cr Mn20Fe20Co20Ni40- high-entropy alloys. *Acta Materialia*, *282*, 120470. https://doi.org/10.1016/j.actamat.2024.120470

Wang, H., Chen, W., Chu, C., Fu, Z., Jiang, Z., Yang, X., & Lavernia, E. J. (2023). Microstructural evolution and mechanical behavior of novel Ti1.6ZrNbAl lightweight refractory high-entropy alloys containing BCC/B2 phases. *Materials Science and Engineering: A, 885*, 145661. https://doi.org/10.1016/j.msea.2023.145661

Wang, H., Chen, W., Liu, S., Chu, C., Huang, L., Duan, J., Tian, Z., & Fu, Z. (2024). Exceptional combinations of tensile properties and corrosion resistance in a single-phase Ti1.6ZrNbMo0.35 refractory high-entropy alloy. *Intermetallics*, *171*, 108349. https://doi.org/10.1016/j.intermet.2024.108349

Wei, L., Liu, B., Han, X., Zhang, C., Wilde, G., & Ye, F. (2024). Effect of Al–Zr and Si–Zr atomic pairs on phases, microstructure and mechanical properties of Si-alloyed (Ti28Zr40Al20Nb12)100-Si (=1, 3, 5, 10) high entropy alloys. *Journal of Materials Research and Technology*, *32*, 2563–2577. https://doi.org/10.1016/j.jmrt.2024.08.128

Xu, C., Chen, D., Yang, X., Wang, S., Fang, H., & Chen, R. (2024). Enhancing mechanical performance of Ti2ZrNbHfVAl refractory high-entropy alloys through laves

phase. *Materials Science and Engineering: A, 918*, 147438.
https://doi.org/10.1016/j.msea.2024.147438

Xu, F., Gao, X., Cui, H., Song, Q., & Chen, R. (2023). Lightweight and high hardness
(AlNbTiVCr)100-Ni high entropy alloys reinforced by Laves phase. *Vacuum,
213*, 112115. https://doi.org/10.1016/j.vacuum.2023.112115

Yang, X., & Zhang, Y. (2012). Prediction of high-entropy stabilized solid-solution in
multi-component alloys. *Materials Chemistry and Physics, 132*(2–3), 233–238.
https://doi.org/10.1016/j.matchemphys.2011.11.021

Yao, X., Wang, W., Qi, X., Lv, Y., Yang, W., Li, T., & Chen, J. (2024). Effects of heat
treatment cooling methods on precipitated phase and mechanical properties
of CoCrFeMnNi–Mo5C0.5 high entropy alloy. *Journal of Materials Research
and Technology, 29*, 3566–3574. https://doi.org/10.1016/j.jmrt.2024.02.076

Ye, Y. F., Wang, Q., Lu, J., Liu, C. T., & Yang, Y. (2016). High-entropy alloy: Challenges
and prospects. *Materials Today, 19*(6), 349–362.
https://doi.org/10.1016/j.mattod.2015.11.026

Yeh, J. -W., Chen, S. -K., Lin, S. -J., Gan, J. -Y., Chin, T. -S., Shun, T. -T., Tsau, C. -H.,
& Chang, S. -Y. (2004). Nanostructured High-Entropy Alloys with Multiple
Principal Elements: Novel Alloy Design Concepts and Outcomes. *Advanced
Engineering Materials, 6*(5), 299–303. https://doi.org/10.1002/adem.200300567

Yu, Z., Xing, W., Liu, C., Yang, K., Shao, H., & Zhao, H. (2024). Construction of
multiscale secondary phase in Al0.25FeCoNiV high-entropy alloy and in-situ
EBSD investigation. *Journal of Materials Research and Technology, 30*, 7607–
7620. https://doi.org/10.1016/j.jmrt.2024.05.168

Záděra, A., Sopoušek, J., Buršík, J., Čupera, J., Brož, P., & Jan, V. (2024). Influence of
substitution of Cr by Cu on phase equilibria and microstructures in the Fe–Ni–
Co–Cr high-entropy alloys. *Intermetallics, 174*, 108455.
https://doi.org/10.1016/j.intermet.2024.108455

Zhang, L., Chen, H., Tao, X., Cai, H., Liu, J., Ouyang, Y., Peng, Q., & Du, Y. (2020).
Machine learning reveals the importance of the formation enthalpy and

atom-size difference in forming phases of high entropy alloys. *Materials & Design*, *193*, 108835. https://doi.org/10.1016/j.matdes.2020.108835

Zhang, W., Li, P., Wang, L., Wan, F., Wu, J., & Yong, L. (2023). Explaining of prediction accuracy on phase selection of amorphous alloys and high entropy alloys using support vector machines in machine learning. *Materials Today Communications*, *35*, 105694. https://doi.org/10.1016/j.mtcomm.2023.105694

Zhang, Y., Wen, C., Wang, C., Antonov, S., Xue, D., Bai, Y., & Su, Y. (2020). Phase prediction in high entropy alloys with a rational selection of materials descriptors and machine learning models. *Acta Materialia*, *185*, 528–539. https://doi.org/10.1016/j.actamat.2019.11.067

Zhang, Y., Zhou, Y. J., Lin, J. P., Chen, G. L., & Liaw, P. K. (2008). Solid‑Solution Phase Formation Rules for Multi‑component Alloys. *Advanced Engineering Materials*, *10*(6), 534–538. https://doi.org/10.1002/adem.200700240

Zhang, Y., Zuo, T. T., Tang, Z., Gao, M. C., Dahmen, K. A., Liaw, P. K., & Lu, Z. P. (2014). Microstructures and properties of high-entropy alloys. *Progress in Materials Science*, *61*, 1–93. https://doi.org/10.1016/j.pmatsci.2013.10.001

Zhang, Z. (2016). Introduction to machine learning: K-nearest neighbors. *Annals of Translational Medicine*, *4*(11), 218–218. https://doi.org/10.21037/atm.2016.03.37

Zhao, J., Gao, X., Zhang, J., Lu, Z., Guo, N., Ding, J., Feng, L., Zhu, G., & Yin, F. (2024). Phase formation mechanism of triple-phase eutectic AlCrFe2Ni2(MoNb)x (x = 0.2, 0.5) high-entropy alloys. *Materials Characterization*, *217*, 114449. https://doi.org/10.1016/j.matchar.2024.114449

Zhao, Q., Luo, H., Yang, Z., Pan, Z., Wang, Z., Islamgaliev, R. K., & Li, X. (2024). Hydrogen induced cracking behavior of the dual-phase Co30Cr10Fe10Al18Ni30Mo2 eutectic high entropy alloy. *International Journal of Hydrogen Energy*, *50*, 134–147. https://doi.org/10.1016/j.ijhydene.2023.09.053

Zhou, J., Liao, H., Chen, H., Feng, D., & Zhu, W. (2023). Realizing strength-ductility combination of Fe3.5Ni3.5Cr2MnAl0.7 high-entropy alloy via coherent dual-phase structure. *Vacuum*, *215*, 112297. https://doi.org/10.1016/j.vacuum.2023.112297

Zhu, C., Li, X., & Dilixiati, N. (2024). Phase evolution, mechanical properties, and corrosion resistance of Ti2NbVAl0.3Zrx lightweight refractory high-entropy alloys. *Intermetallics*, *173*, 108433. https://doi.org/10.1016/j.intermet.2024.108433
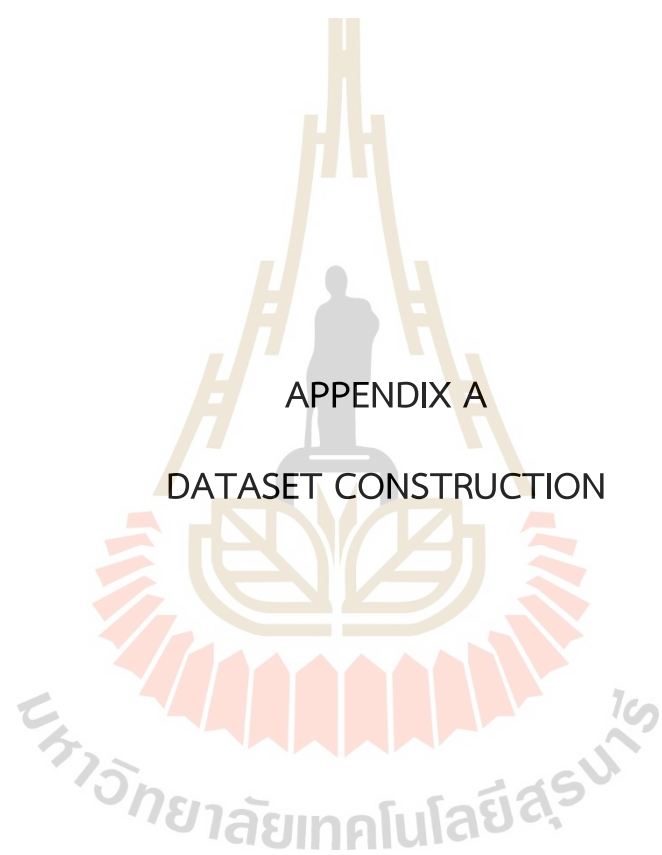
APPENDIX A

DATASET CONSTRUCTION

Table A1. Alloys and their experimental phase collected from existing literatures.

Alloys	Experimental Phase	Reference
CoFeNi	FCC	(Gorsse et al., 2018)
CoFeNiSi0.25	FCC	(Gorsse et al., 2018)
CoFeNiSi0.5	FCC+IM	(Gorsse et al., 2018)
CoFeNiSi0.75	FCC+IM	(Gorsse et al., 2018)
Al0.25CoFeNi	FCC	(Gorsse et al., 2018)
Al0.5CoFeNi	BCC+FCC	(Gorsse et al., 2018)
Al0.75CoFeNi	BCC+FCC	(Gorsse et al., 2018)
CoCrFeNi	FCC	(Gorsse et al., 2018)
CoCrFeMo0.5Ni	FCC+IM	(Gorsse et al., 2018)
CoCrFeNb0.103Ni	FCC+IM	(Gorsse et al., 2018)
CoCrFeNb0.155Ni	FCC+IM	(Gorsse et al., 2018)
CoCrFeNb0.206Ni	FCC+IM	(Gorsse et al., 2018)
CoCrFeNb0.309Ni	FCC+IM	(Gorsse et al., 2018)
CoCrFeNb0.412Ni	FCC+IM	(Gorsse et al., 2018)
CoCrFeNiTi	FCC	(Gorsse et al., 2018)
Al0.25CoCrFeNi	FCC	(Gorsse et al., 2018)
Al0.375CoCrFeNi	FCC	(Gorsse et al., 2018)
Al0.5CoCrFeNi	BCC+FCC	(Gorsse et al., 2018)
Al0.75CoCrFeNi	BCC+FCC	(Gorsse et al., 2018)
Al0.875CoCrFeNi	BCC	(Gorsse et al., 2018)

Table A1. Alloys and their experimental phase collected from existing literatures
(continued).

Alloys	Experimental Phase	Reference
AlCoCrFeNi	BCC	(Gorsse et al., 2018)
Al1.25CoCrFeNi	BCC	(Gorsse et al., 2018)
Al1.5CoCrFeNi	BCC	(Gorsse et al., 2018)
Al2CoCrFeNi	BCC	(Gorsse et al., 2018)
Al2.5CoCrFeNi	BCC	(Gorsse et al., 2018)
Al3CoCrFeNi	BCC	(Gorsse et al., 2018)
AlC0.1CoCrFeNi	BCC+IM	(Gorsse et al., 2018)
AlC0.2CoCrFeNi	BCC+IM	(Gorsse et al., 2018)
AlC0.3CoCrFeNi	BCC+IM	(Gorsse et al., 2018)
AlC0.4CoCrFeNi	BCC+IM	(Gorsse et al., 2018)
AlC0.5CoCrFeNi	BCC+IM	(Gorsse et al., 2018)
AlCCoCrFeNi	BCC+IM	(Gorsse et al., 2018)
AlC1.5CoCrFeNi	BCC+IM	(Gorsse et al., 2018)
Al0.5CoCrFeMo0.5Ni	FCC+IM	(Gorsse et al., 2018)
AlCo0.5CrFeMo0.5Ni	BCC+IM	(Gorsse et al., 2018)
AlCoCrFe0.5Mo0.5Ni	BCC+IM	(Gorsse et al., 2018)
AlCoCrFe0.6Mo0.5Ni	BCC+IM	(Gorsse et al., 2018)
AlCoCrFeMo0.1Ni	BCC	(Gorsse et al., 2018)
AlCoCrFeMo0.2Ni	BCC+IM	(Gorsse et al., 2018)
AlCoCrFeMo0.3Ni	BCC+IM	(Gorsse et al., 2018)

Table A1. Alloys and their experimental phase collected from existing literatures (continued).

Alloys	Experimental Phase	Reference
AlCoCrFeMo0.4Ni	BCC+IM	(Gorsse et al., 2018)
AlCoCrFeMo0.5Ni	BCC+IM	(Gorsse et al., 2018)
AlCoCrFeMo0.5Ni1.5	FCC+IM	(Gorsse et al., 2018)
AlCoCrFeMo0.5Ni2	FCC+IM	(Gorsse et al., 2018)
AlCo1.5CrFeMo0.5Ni	BCC+IM	(Gorsse et al., 2018)
AlCo2CrFeMo0.5Ni	BCC+FCC+IM	(Gorsse et al., 2018)
AlCoCrFe1.5Mo0.5Ni	BCC+IM	(Gorsse et al., 2018)
AlCoCrFe2Mo0.5Ni	BCC+IM	(Gorsse et al., 2018)
Al1.5CoCrFeMo0.5Ni	BCC+IM	(Gorsse et al., 2018)
Al2CoCrFeMo0.5Ni	BCC	(Gorsse et al., 2018)
AlCoCrFeNb0.1Ni	BCC	(Gorsse et al., 2018)
AlCoCrFeNb0.25Ni	BCC+IM	(Gorsse et al., 2018)
AlCoCrFeNb0.5Ni	BCC+IM	(Gorsse et al., 2018)
AlCoCrFeNb0.75Ni	BCC+IM	(Gorsse et al., 2018)
AlCoCrFeNiSi0.2	BCC	(Gorsse et al., 2018)
AlCoCrFeNiSi0.4	BCC	(Gorsse et al., 2018)
AlCoCrFeNiSi0.6	BCC	(Gorsse et al., 2018)
AlCoCrFeNiSi0.8	BCC+IM	(Gorsse et al., 2018)
AlCoCrFeNiSi	BCC+IM	(Gorsse et al., 2018)
Al0.5CoCrFeNiTi	BCC+IM	(Gorsse et al., 2018)

Table A1. Alloys and their experimental phase collected from existing literatures (continued).

Alloys	Experimental Phase	Reference
AlCoCrFeNiTi1.5	BCC+IM	(Gorsse et al., 2018)
Al1.5CoCrFeNiTi	BCC	(Gorsse et al., 2018)
Al2CoCrFeNiTi	BCC	(Gorsse et al., 2018)
CoCrFeMnNi	FCC	(Gorsse et al., 2018)
CoCrFeMnNiV0.25	FCC	(Gorsse et al., 2018)
CoCrFeMnNiV0.5	FCC	(Gorsse et al., 2018)
CoCrFeMnNiV0.75	FCC+IM	(Gorsse et al., 2018)
CoCrFeMnNiV1.0	FCC+IM	(Gorsse et al., 2018)
Al0.10CoCrFeMnNi	FCC	(Gorsse et al., 2018)
Al0.20CoCrFeMnNi	FCC	(Gorsse et al., 2018)
Al0.38CoCrFeMnNi	FCC	(Gorsse et al., 2018)
Al0.43CoCrFeMnNi	BCC+FCC	(Gorsse et al., 2018)
Al0.49CoCrFeMnNi	BCC+FCC	(Gorsse et al., 2018)
Al0.56CoCrFeMnNi	BCC+FCC	(Gorsse et al., 2018)
Al0.62CoCrFeMnNi	BCC+FCC	(Gorsse et al., 2018)
Al0.68CoCrFeMnNi	BCC+FCC	(Gorsse et al., 2018)
Al0.75CoCrFeMnNi	BCC+FCC	(Gorsse et al., 2018)
Al0.81CoCrFeMnNi	BCC+FCC	(Gorsse et al., 2018)
Al0.88CoCrFeMnNi	BCC+FCC	(Gorsse et al., 2018)
Al0.95CoCrFeMnNi	BCC+FCC	(Gorsse et al., 2018)

Table A1. Alloys and their experimental phase collected from existing literatures (continued).

Alloys	Experimental Phase	Reference
Al1.25CoCrFeMnNi	BCC	(Gorsse et al., 2018)
CoCrNi	FCC	(Gorsse et al., 2018)
CoMnNi	FCC	(Gorsse et al., 2018)
FeMnNi	FCC	(Gorsse et al., 2018)
CoCrMnNi	FCC	(Gorsse et al., 2018)
CoFeMnNi	FCC	(Gorsse et al., 2018)
Al0.5CrFe1.5MnNi0.5	BCC	(Gorsse et al., 2018)
Al0.3CrFe1.5MnNi0.5	BCC+FCC	(Gorsse et al., 2018)
AlCrFeNi	BCC+IM	(Gorsse et al., 2018)
AlCrFeNiMo0.2	BCC+IM	(Gorsse et al., 2018)
AlCrFeNiMo0.5	BCC+IM	(Gorsse et al., 2018)
CoCrCu0.5FeNi	FCC	(Gorsse et al., 2018)
CoCrCuFeNi	FCC	(Gorsse et al., 2018)
CoCrCuFeNiTi0.5	FCC	(Gorsse et al., 2018)
CoCrCuFeNiTi0.8	FCC+IM	(Gorsse et al., 2018)
CoCrCuFeNiTi	FCC	(Gorsse et al., 2018)
Al0.25CoCrCu0.5FeNiTi0.5	FCC	(Gorsse et al., 2018)
Al0.25CoCrCu0.75FeNiTi0.5	FCC	(Gorsse et al., 2018)
Al0.3CoCrCuFeNi	FCC	(Gorsse et al., 2018)
Al0.5CoCrCuFeNi	FCC	(Gorsse et al., 2018)

Table A1. Alloys and their experimental phase collected from existing literatures (continued).

Alloys	Experimental Phase	Reference
AlCoCrCuFeNi	BCC+FCC	(Gorsse et al., 2018)
AlCoCrCuFeMnNi	BCC+FCC+IM	(Gorsse et al., 2018)
AlCoCrCuFeNiTi	BCC+FCC	(Gorsse et al., 2018)
AlCoCrCuFeNiV	BCC+FCC	(Gorsse et al., 2018)
Al1.3CoCrCuFeNi	BCC+FCC	(Gorsse et al., 2018)
Al1.5CoCrCuFeNi	BCC+FCC	(Gorsse et al., 2018)
Al1.8CoCrCuFeNi	BCC+FCC	(Gorsse et al., 2018)
Al2CoCrCuFeNi	BCC+FCC	(Gorsse et al., 2018)
Al2.3CoCrCuFeNi	BCC+FCC	(Gorsse et al., 2018)
Al2.5CoCrCuFeNi	BCC+FCC	(Gorsse et al., 2018)
Al2.8CoCrCuFeNi	BCC	(Gorsse et al., 2018)
Al0.5CoCrCu0.5FeNiTi0.5	BCC+FCC	(Gorsse et al., 2018)
Al0.5CoCrCuFeNiTi0.2	FCC	(Gorsse et al., 2018)
Al0.5CoCrCuFeNiTi0.4	FCC	(Gorsse et al., 2018)
Al0.5CoCrCuFeNiTi0.6	BCC+FCC	(Gorsse et al., 2018)
Al0.5CoCrCuFeNiTi0.8	BCC+FCC	(Gorsse et al., 2018)
Al0.5CoCrCuFeNiTi	BCC+FCC+IM	(Gorsse et al., 2018)
Al0.5CoCrCuFeNiTi1.2	BCC+FCC+IM	(Gorsse et al., 2018)
Al0.5CoCrCuFeNiTi1.4	BCC+FCC+IM	(Gorsse et al., 2018)
Al0.5CoCrCuFeNiTi1.6	BCC+FCC+IM	(Gorsse et al., 2018)

Table A1. Alloys and their experimental phase collected from existing literatures (continued).

Alloys	Experimental Phase	Reference
Al0.5CoCrCuFeNiTi1.8	BCC+FCC+IM	(Gorsse et al., 2018)
Al0.5CoCrCuFeNiTi2	BCC+FCC+IM	(Gorsse et al., 2018)
Al0.5CoCrCuFeNiV0.2	FCC	(Gorsse et al., 2018)
Al0.5CoCrCuFeNiV0.4	BCC+FCC	(Gorsse et al., 2018)
Al0.5CoCrCuFeNiV0.6	BCC+FCC+IM	(Gorsse et al., 2018)
Al0.5CoCrCuFeNiV0.8	BCC+FCC+IM	(Gorsse et al., 2018)
Al0.5CoCrCuFeNiV1.0	BCC+FCC+IM	(Gorsse et al., 2018)
Al0.5CoCrCuFeNiV1.2	BCC	(Gorsse et al., 2018)
Al0.5CoCrCuFeNiV1.4	BCC	(Gorsse et al., 2018)
Al0.5CoCrCuFeNiV1.6	BCC	(Gorsse et al., 2018)
Al0.5CoCrCuFeNiV1.8	BCC	(Gorsse et al., 2018)
Al0.5CoCrCuFeNiV2.0	BCC	(Gorsse et al., 2018)
Al0.75CoCrCu0.25FeNiTi0.5	BCC+FCC	(Gorsse et al., 2018)
AlCoCrCuNiTi	BCC	(Gorsse et al., 2018)
AlCoFeNi	BCC	(Gorsse et al., 2018)
AlCoFeNiTiVZr	BCC	(Gorsse et al., 2018)
CoCuFeNi	FCC	(Gorsse et al., 2018)
CoCuFeNiSn0.02	FCC	(Gorsse et al., 2018)
CoCuFeNiSn0.04	FCC+IM	(Gorsse et al., 2018)
CoCuFeNiSn0.05	FCC+IM	(Gorsse et al., 2018)

Table A1. Alloys and their experimental phase collected from existing literatures (continued).

Alloys	Experimental Phase	Reference
CoCuFeNiSn0.07	FCC+IM	(Gorsse et al., 2018)
CoCuFeNiSn0.1	FCC+IM	(Gorsse et al., 2018)
CoCuFeNiSn0.2	FCC+IM	(Gorsse et al., 2018)
CoCuFeNiSn0.5	FCC+IM	(Gorsse et al., 2018)
AlCoCuFeNi	BCC+FCC	(Gorsse et al., 2018)
AlCoCuFeNiSi	BCC+FCC	(Gorsse et al., 2018)
AlCoCuFeNiTi	BCC+FCC	(Gorsse et al., 2018)
AlCoCuFeNiZr	BCC+FCC+IM	(Gorsse et al., 2018)
CoCuFeMnNi	FCC	(Gorsse et al., 2018)
CoCuFeMnNiSn0.03	FCC	(Gorsse et al., 2018)
CoCuFeMnNiSn0.05	FCC+IM	(Gorsse et al., 2018)
CoCuFeMnNiSn0.08	FCC+IM	(Gorsse et al., 2018)
CoCuFeMnNiSn0.10	FCC+IM	(Gorsse et al., 2018)
CoCuFeMnNiSn0.20	FCC+IM	(Gorsse et al., 2018)
CrCuFeMnNi	BCC+FCC	(Gorsse et al., 2018)
CrCuFeMoNi	FCC	(Gorsse et al., 2018)
AlCrCuFeNi0.6	BCC+FCC	(Gorsse et al., 2018)
AlCrCuFeNi0.8	BCC+FCC	(Gorsse et al., 2018)
AlCrCuFeNi	BCC+FCC	(Gorsse et al., 2018)
AlCrCuFeNi1.2	BCC+FCC	(Gorsse et al., 2018)

Table A1. Alloys and their experimental phase collected from existing literatures
(continued).

Alloys	Experimental Phase	Reference
AlCrCuFeNi1.4	BCC+FCC	(Gorsse et al., 2018)
AlCrCuFeNi2	BCC+FCC	(Gorsse et al., 2018)
AlCrCuFeNiTi	BCC+IM	(Gorsse et al., 2018)
Al0.2CrCuFeNi2	FCC	(Gorsse et al., 2018)
Al0.4CrCuFeNi2	FCC	(Gorsse et al., 2018)
Al0.6CrCuFeNi2	FCC	(Gorsse et al., 2018)
Al0.8CrCuFeNi2	FCC	(Gorsse et al., 2018)
Al1.2CrCuFeNi2	BCC+FCC	(Gorsse et al., 2018)
Al1.125CuFe0.75NiTi1.125	FCC	(Gorsse et al., 2018)
Al22.5Cu20Fe15Ni20Ti22.5	FCC	(Gorsse et al., 2018)
AlCuFeNiTi	FCC	(Gorsse et al., 2018)
AlCuNiTi	FCC	(Gorsse et al., 2018)
AlCr0.5NbTiV	BCC	(Gorsse et al., 2018)
AlCrNbTiV	BCC+IM	(Gorsse et al., 2018)
Al0.4Hf0.6NbTaTiZr	BCC	(Gorsse et al., 2018)
Al0.3HfNbTaTiZr	BCC	(Gorsse et al., 2018)
Al0.5HfNbTaTiZr	BCC	(Gorsse et al., 2018)
Al0.75HfNbTaTiZr	BCC	(Gorsse et al., 2018)
AlHfNbTaTiZr	BCC+IM	(Gorsse et al., 2018)
AlMo0.5NbTa0.5TiZr	BCC	(Gorsse et al., 2018)

Table A1. Alloys and their experimental phase collected from existing literatures (continued).

Alloys	Experimental Phase	Reference
Al0.25MoNbTiV	BCC	(Gorsse et al., 2018)
Al0.5MoNbTiV	BCC	(Gorsse et al., 2018)
Al0.75MoNbTiV	BCC	(Gorsse et al., 2018)
AlMoNbTiV	BCC	(Gorsse et al., 2018)
Al0.25NbTaTiV	BCC	(Gorsse et al., 2018)
Al0.5NbTaTiV	BCC	(Gorsse et al., 2018)
AlNbTaTiV	BCC	(Gorsse et al., 2018)
Al0.3NbTa0.8Ti1.4V0.2Zr1.3	BCC	(Gorsse et al., 2018)
Al0.5NbTa0.8Ti1.5V0.2Zr	BCC	(Gorsse et al., 2018)
Al0.3NbTaTi1.4Zr1.3	BCC	(Gorsse et al., 2018)
AlNb1.5Ta0.5Ti1.5Zr0.5	BCC	(Gorsse et al., 2018)
AlNbTiV	BCC	(Gorsse et al., 2018)
CrMo0.5NbTa0.5TiZr	BCC+IM	(Gorsse et al., 2018)
CrNbTiVZr	BCC+IM	(Gorsse et al., 2018)
CrNbTiZr	BCC+IM	(Gorsse et al., 2018)
FeMoNiTiVZr	BCC+IM	(Gorsse et al., 2018)
Hf0.5Mo0.5NbTiZr	BCC+IM	(Gorsse et al., 2018)
Hf0.5Mo0.5NbSi0.1TiZr	BCC+IM	(Gorsse et al., 2018)
Hf0.5Mo0.5NbSi0.3TiZr	BCC+IM	(Gorsse et al., 2018)
Hf0.5Mo0.5NbSi0.5TiZr	BCC+IM	(Gorsse et al., 2018)

Table A1. Alloys and their experimental phase collected from existing literatures (continued).

Alloys	Experimental Phase	Reference
Hf0.5Mo0.5NbSi0.7TiZr	BCC+IM	(Gorsse et al., 2018)
Hf0.5Mo0.5NbSi0.9TiZr	BCC+IM	(Gorsse et al., 2018)
Hf0.5Mo0.5NbTiZrC0.1	BCC+IM	(Gorsse et al., 2018)
Hf0.5Mo0.5NbTiZrC0.3	BCC+IM	(Gorsse et al., 2018)
HfMo0.25NbTaTiZr	BCC	(Gorsse et al., 2018)
HfMo0.5NbTaTiZr	BCC	(Gorsse et al., 2018)
HfMo0.75NbTaTiZr	BCC	(Gorsse et al., 2018)
HfMoNbTaTiZr	BCC	(Gorsse et al., 2018)
HfMoTaTiZr	BCC	(Gorsse et al., 2018)
HfMoNbZrTi	BCC	(Gorsse et al., 2018)
HfNbTaZr	BCC	(Gorsse et al., 2018)
Hf0.5Nb0.5Ta0.5Ti1.5Zr	BCC	(Gorsse et al., 2018)
HfNbTaTiZr	BCC	(Gorsse et al., 2018)
MoNbTaV	BCC	(Gorsse et al., 2018)
MoNbTaVW	BCC	(Gorsse et al., 2018)
MoNbTaW	BCC	(Gorsse et al., 2018)
MoNbTiV	BCC	(Gorsse et al., 2018)
Mo0.3NbTiVZr	BCC	(Gorsse et al., 2018)
Mo0.5NbTiVZr	BCC	(Gorsse et al., 2018)
Mo0.7NbTiVZr	BCC	(Gorsse et al., 2018)

Table A1. Alloys and their experimental phase collected from existing literatures (continued).

Alloys	Experimental Phase	Reference
MoNbTiVZr	BCC	(Gorsse et al., 2018)
Mo1.3NbTiVZr	BCC	(Gorsse et al., 2018)
Mo1.5NbTiVZr	BCC	(Gorsse et al., 2018)
Mo1.7NbTiVZr	BCC	(Gorsse et al., 2018)
Mo2NbTiVZr	BCC	(Gorsse et al., 2018)
MoNbTiV0.25Zr	BCC	(Gorsse et al., 2018)
MoNbTiV0.50Zr	BCC	(Gorsse et al., 2018)
MoNbTiV0.75Zr	BCC	(Gorsse et al., 2018)
MoNbTiV1.0Zr	BCC	(Gorsse et al., 2018)
MoNbTiV1.5Zr	BCC	(Gorsse et al., 2018)
MoNbTiV2.0Zr	BCC	(Gorsse et al., 2018)
MoNbTiV3.0Zr	BCC	(Gorsse et al., 2018)
MoNbTiZr	BCC	(Gorsse et al., 2018)
NbTaTiV	BCC	(Gorsse et al., 2018)
NbTaVW	BCC	(Gorsse et al., 2018)
NbTiV0.3Zr	BCC	(Gorsse et al., 2018)
NbTiV0.3Mo0.1Zr	BCC	(Gorsse et al., 2018)
NbTiV0.3Mo0.3Zr	BCC	(Gorsse et al., 2018)
NbTiV0.3Mo0.5Zr	BCC	(Gorsse et al., 2018)
NbTiV0.3Mo0.7Zr	BCC	(Gorsse et al., 2018)

Table A1. Alloys and their experimental phase collected from existing literatures (continued).

Alloys	Experimental Phase	Reference
NbTiV0.3MoZr	BCC	(Gorsse et al., 2018)
NbTiV0.3Mo1.3Zr	BCC	(Gorsse et al., 2018)
NbTiV0.3Mo1.5Zr	BCC	(Gorsse et al., 2018)
NbTiVZr	BCC	(Gorsse et al., 2018)
NbTiV2Zr	BCC	(Gorsse et al., 2018)
CoFeNiV	FCC	(Gorsse et al., 2018)
CoFeMo0.2NiV	FCC+IM	(Gorsse et al., 2018)
CoFeMo0.4NiV	FCC+IM	(Gorsse et al., 2018)
CoFeMo0.6NiV	FCC+IM	(Gorsse et al., 2018)
CoFeMo0.8NiV	FCC+IM	(Gorsse et al., 2018)
CoFeMoNiV	FCC+IM	(Gorsse et al., 2018)
CoFeMoNi1.2V	FCC+IM	(Gorsse et al., 2018)
CoFeMoNi1.4V	FCC+IM	(Gorsse et al., 2018)
CoFeMoNi1.6V	FCC+IM	(Gorsse et al., 2018)
CoFeMoNi1.8V	FCC+IM	(Gorsse et al., 2018)
CoFeMoNi2V	FCC+IM	(Gorsse et al., 2018)
CoCrCuFeMnNiTiV	BCC+FCC+IM	(Gorsse et al., 2018)
Al11.1(CoCrCuFeMnNiTiV)88.9	BCC+FCC	(Gorsse et al., 2018)
Al20(CoCrCuFeMnNiTiV)80	BCC	(Gorsse et al., 2018)
Al40(CoCrCuFeMnNiTiV)60	BCC+IM	(Gorsse et al., 2018)

Table A1. Alloys and their experimental phase collected from existing literatures (continued).

Alloys	Experimental Phase	Reference
AlFeNiTiVZr	BCC	(Gorsse et al., 2018)
(CuMnNi)75Zn25	FCC	(Gorsse et al., 2018)
(CuMnNi)80Zn20	FCC	(Gorsse et al., 2018)
(CuMnNi)90Al10	FCC+IM	(Gorsse et al., 2018)
(CuMnNi)90Sn10	FCC+IM	(Gorsse et al., 2018)
(CuMnNi)95Al5	FCC	(Gorsse et al., 2018)
(CuMnNi)95Sn5	FCC+IM	(Gorsse et al., 2018)
AlCoCrFeNi	BCC	(Jiang et al., 2018)
Al0.8CoCrFeNiTi0.2	BCC+FCC	(Jiang et al., 2018)
Al0.5CoCrFeNiTi0.5	BCC+FCC	(Jiang et al., 2018)
CrMnFeV	BCC	(Song et al., 2018)
CrMn0.3FeVCu0.06	BCC+FCC	(Song et al., 2018)
CrMn0.5FeVCu0.1	BCC+FCC	(Song et al., 2018)
CrMn0.7FeVCu0.14	BCC+FCC	(Song et al., 2018)
CrMnFeVCu0.2	BCC+FCC	(Song et al., 2018)
AlCrFe2Mo0.2Nb0.2Ni2	FCC+IM	(J. Zhao et al., 2024)
AlCrFe2Mo0.5Nb0.5Ni2	FCC+IM	(J. Zhao et al., 2024)
Al25Co14Cr5Fe14Mn8Ni25Ti9	BCC+IM	(Olorundaisi et al., 2024)
CoFeMnNiV	FCC	(J. Gao et al., 2023)
(FeCoNi)0.75Cr0.25	FCC	(Záděra et al., 2024)

Table A1. Alloys and their experimental phase collected from existing literatures (continued).

Alloys	Experimental Phase	Reference
(FeCoNi)0.75Cr0.05Cu0.2	FCC	(Záděra et al., 2024)
AlCoCr0.5Fe2.5Ni2.5	BCC+FCC	(Hu et al., 2025)
Al0.25CoCrFeNi	FCC	(Hu et al., 2025)
Ti2ZrNbHfVAl1.2	BCC+IM	(C. Xu et al., 2024)
Nb25Ti35V5Zr35	BCC	(Gong et al., 2024)
(Ti28Zr40Al20Nb12)97Si3	BCC+IM	(Wei et al., 2024)
(Ti28Zr40Al20Nb12)95Si5	BCC+IM	(Wei et al., 2024)
(Ti28Zr40Al20Nb12)90Si10	BCC+IM	(Wei et al., 2024)
Ti2NbVAl0.3Zr0.25	BCC+IM	(Zhu et al., 2024)
Ti2NbVAl0.3Zr0.5	BCC+IM	(Zhu et al., 2024)
Ti2NbVAl0.3Zr0.75	BCC+IM	(Zhu et al., 2024)
Ti2NbVAl0.3Zr	BCC	(Zhu et al., 2024)
Ti1(NbMoTa)2W0.5	BCC	(Kim et al., 2024)
FeMn40Co10Cr10C0.5Ti2	FCC+IM	(Saboktakin Rizi et al., 2024)
Co68Al18.2Fe6.5V4.75Cr2.55	FCC+IM	(X. Liu, Feng, et al., 2024)
Ti1·6ZrNbMo0.35	BCC	(Wang et al., 2024)
Cr14Mn20Fe20Co20Ni26	FCC	(Wagner et al., 2025)
Cr24Mn20Fe20Co20Ni16	FCC	(Wagner et al., 2025)
Cr26Mn20Fe20Co20Ni14	FCC	(Wagner et al., 2025)
Al21Co19.5Fe9.5Ni50	FCC+IM	(X. Liu, Liu, et al., 2024)

Table A1. Alloys and their experimental phase collected from existing literatures (continued).

Alloys	Experimental Phase	Reference
(Ti32.5V27.5Zr7.5Nb32.5)97Ni3	BCC+IM	(Liang et al., 2024)
(Ti32.5V27.5Zr7.5Nb32.5)94Ni6	BCC+IM	(Liang et al., 2024)
(Ti32.5V27.5Zr7.5Nb32.5)91Ni9	BCC+IM	(Liang et al., 2024)
Al0.2CoCrFeNi	FCC	(Sun et al., 2023)
Al0.2CoCrFeNiSc0.3	BCC+FCC	(Sun et al., 2023)
CoCrFeMnNiMo5C0.5	FCC+IM	(Yao et al., 2024)
Al0.25FeCoNiV	BCC+FCC	(Yu et al., 2024)
Co30Cr10Fe10Al18Ni30Mo2	BCC+FCC	(Q. Zhao et al., 2024)
CoCr1.3FeNi0.7MnNb0.3	FCC+IM	(Potnis & Das, 2024)
CoCr1.3FeNi0.7MnNb0.45	FCC+IM	(Potnis & Das, 2024)
Al14Co41Cr15Fe10Ni20	FCC+IM	(Shafiei et al., 2023)
Cr15Cu5Fe20Mn25Ni35	FCC	(Vaghari & Dehghani, 2023)
Al0.65CoCrFe2Ni	BCC+FCC	(Tamuly et al., 2023)
Al40Co15Cr15Fe15Ni15	BCC+IM	(Han et al., 2024)
AlFeCoNiMo0.2	BCC	(Li et al., 2023)
Ti1.6ZrNbAl0.3	BCC+IM	(Wang et al., 2023)
Ti1.6ZrNbAl0.4	BCC+IM	(Wang et al., 2023)
Ti1.6ZrNbAl0.5	BCC+IM	(Wang et al., 2023)
Ti1.6ZrNbAl0.6	BCC+IM	(Wang et al., 2023)
AlNbTiVCr	BCC	(F. Xu et al., 2023)

Table A1. Alloys and their experimental phase collected from existing literatures (continued).

Alloys	Experimental Phase	Reference
(AlNbTiVCr)99.5Ni0.5	BCC	(F. Xu et al., 2023)
(AlNbTiVCr)99Ni1	BCC+IM	(F. Xu et al., 2023)
(AlNbTiVCr)98.5Ni1.5	BCC+IM	(F. Xu et al., 2023)
(AlNbTiVCr)98Ni2	BCC+IM	(F. Xu et al., 2023)
CrFeNbTiMo	BCC+IM	(B. Chen et al., 2023)
Fe3.5Ni3.5Cr2MnAl0.7	BCC+FCC	(Zhou et al., 2023)
(AlCoCrFeNi)99.5Zr0.5	BCC+IM	(Ren et al., 2023)
(AlCoCrFeNi)99Zr1	BCC+IM	(Ren et al., 2023)
(AlCoCrFeNi)98Zr2	BCC+IM	(Ren et al., 2023)

Table A2. Atomic properties of elements (atomic radius, valence electron count, Pauling electronegativity, and melting temperature) used for feature extraction in HEA phase prediction.

Element	Atomic radius (A)	Valence electron	Pauling electronegativity	Melting temperature (K)
H	-	-	2.2	13.81
He	-	-	-	0.95
Li	1.56	1	0.98	453.7
Be	1.13	2	1.57	1560
B	0.98	3	2.04	2365
C	0.92	4	2.55	3825
N	-	5	3.04	63.15
O	-	6	3.44	54.8
F	-	-	3.98	53.55
Ne	-	-	-	24.55
Na	1.91	1	0.93	371
Mg	1.6	2	1.31	922
Al	1.43	3	1.61	933.5
Si	1.32	4	1.9	1683
P	-	5	2.19	317.3
S	-	6	2.58	392.2
Cl	-	-	3.16	172.17
Ar	-	-	-	83.95
K	2.38	1	0.82	336.8
Ca	1.98	2	1	1112

Table A2. Atomic properties of elements (atomic radius, valence electron count, Pauling electronegativity, and melting temperature) used for feature extraction in HEA phase prediction (continued).

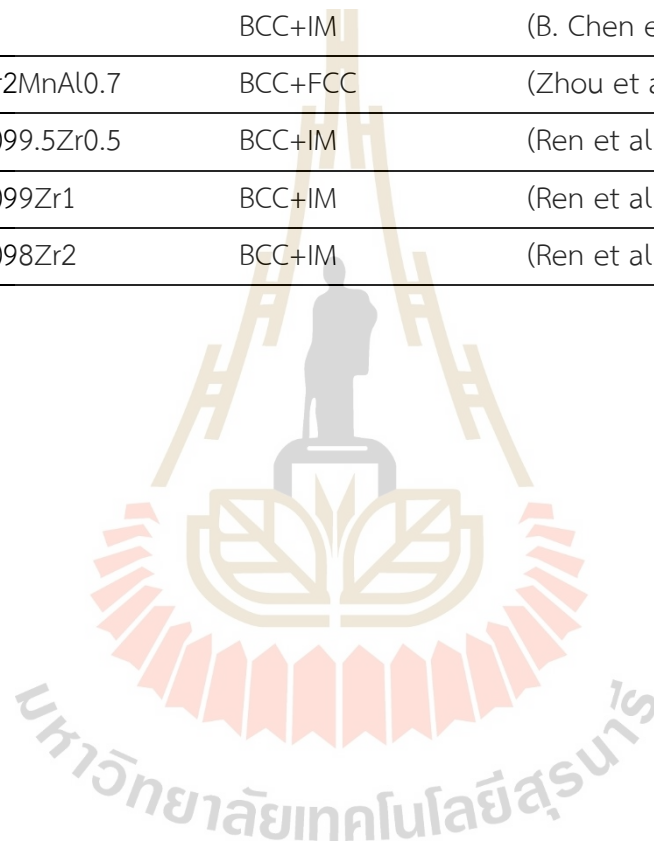| Element	Atomic radius (A)	Valence electron	Pauling electronegativity	Melting temperature (K)
Sc	1.64	3	1.36	1814
Ti	1.46	4	1.54	1945
V	1.35	5	1.63	2163
Cr	1.28	6	1.66	2130
Mn	1.26	7	1.55	1518
Fe	1.27	8	1.83	1808
Co	1.25	9	1.88	1768
Ni	1.25	10	1.91	1726
Cu	1.28	11	1.9	1356.6
Zn	1.39	12	1.65	692.73
Ga	1.41	3	1.81	302.92
Ge	1.37	4	2.01	1211.5
As	1.39	5	2.18	1090
Se	-	6	2.55	494
Br	-	-	2.96	265.95
Kr	-	-	3	116
Rb	2.55	1	0.82	312.63
Sr	2.15	2	0.95	1042
Y	1.8	3	1.22	1795
Zr	1.6	4	1.33	2128

Table A2. Atomic properties of elements (atomic radius, valence electron count, Pauling electronegativity, and melting temperature) used for feature extraction in HEA phase prediction (continued).

Element	Atomic radius (A)	Valence electron	Pauling electronegativity	Melting temperature (K)
Nb	1.47	5	1.6	2742
Mo	1.4	6	2.16	2896
Tc	1.36	7	1.9	2477
Ru	1.34	8	2.2	2610
Rh	1.35	9	2.28	2236
Pd	1.38	10	2.2	1825
Ag	1.45	11	1.93	1235.08
Cd	1.57	12	1.69	594.26
In	1.66	3	1.78	429.78
Sn	1.55	4	1.96	505.12
Sb	1.59	5	2.05	903.91
Te	-	6	2.1	722.72
I	-	-	2.66	386.7
Xe	-	-	2.6	161.39
Cs	2.73	1	0.79	301.54
Ba	2.24	2	0.89	1002
La	1.88	3	1.1	1191
Ce	1.82	3	1.12	1071
Pr	1.83	3	1.13	1204
Nd	1.82	3	1.14	1294

Table A2. Atomic properties of elements (atomic radius, valence electron count, Pauling electronegativity, and melting temperature) used for feature extraction in HEA phase prediction (continued).

Element	Atomic radius (A)	Valence electron	Pauling electronegativity	Melting temperature (K)
Pm	1.81	-	1.13	1315
Sm	1.8	3	1.17	1347
Eu	1.8	3	1.2	1095
Gd	1.8	3	1.2	1585
Tb	1.78	3	1.1	1629
Dy	1.77	3	1.22	1685
Ho	1.77	3	1.23	1747
Er	1.76	3	1.24	1802
Tm	1.75	3	1.25	1818
Yb	1.74	3	1.1	1092
Lu	-	3	1.27	1936
Hf	1.58	4	1.3	2504
Ta	1.47	5	1.5	3293
W	1.41	6	2.36	3695
Re	1.38	7	1.9	3455
Os	1.35	8	2.2	3300
Ir	1.36	9	2.2	2720
Pt	1.39	10	2.28	2042.1
Au	1.44	11	2.54	1337.58
Hg	1.57	2	2	234.31

Table A2. Atomic properties of elements (atomic radius, valence electron count, Pauling electronegativity, and melting temperature) used for feature extraction in HEA phase prediction (continued).

Element	Atomic radius (A)	Valence electron	Pauling electronegativity	Melting temperature (K)
Tl	1.72	3	1.62	577
Pb	1.75	4	1.87	600.65
Bi	1.7	5	2.02	544.59
Po	1.76	-	2	-
At	-	-	2.2	575
Rn	-	-	2.2	202
Fr	-	-	0.7	300
Ra	-	-	0.9	973
Ac	-	-	1.1	-
Th	1.8	-	1.3	2028
Pa	1.63	-	1.5	1845
U	1.56	-	1.38	1408
Np	1.56	-	1.36	912
Pu	1.64	-	1.28	913
Am	1.81	-	1.13	1449
Cm	-	-	1.28	-

Table A3. The values of ΔH_{ij} for atomic pairs between the element (Takeuchi & Inoue, 2005).

	Pd 46	Ag 47	Cd 48	In 49	Sn 50	Sb 51	Cs 55	Ba 56	La 57	Ce 58	Pr 59	Nd 60	Pm 61	Sm 62	Eu 63	Gd 64	Tb 65	Dy 66	Ho 67	Er 68	Tm 69	Yb 70	Lu 71	Hf 72	Ta 73	W 74	Re 75	Os 76	Ir 77	Pt 78	Au 79	Hg 80	Tl 81	Pb 82	Bi 83	Th 90	U 92	Pu 94
46 Pd	-7	-26	-31	-34	-28	-9	-62	-82	-82	-83	-83	-86	-84	-61	-84	-86	-84	-83	-85	-85	-62	-87	-80	-52	-6	6	8	6	2	0	-18	-21	-18	-21	-91	-59	-58	
45 Rh	2	10	-6	-8	-13	-8	34	-21	-50	-52	-53	-55	-54	-24	-54	-55	-54	-57	-55	-28	-69	-63	-45	-9	1	2	1	-2	2	5	6	3		-63	-46	42		
44 Ru	6	23	9	10	4	9	54	6	-28	-30	-32	-35	-34	2	-34	-35	-35	-37	-37	-4	-40	-52	-39	-10	-1	0	-1	-1	15	18	24	25	23		-45	-36	-30	
43 Tc	4	24	10	11	5	8	69	11	-23	-25	-27	-30	-29	6	-29	-30	-29	-32	-32	1	-35	-47		-3	-7	0	0	-2	3	14	18	25	26	23	-39	-31	-26	
42 Mo	-15	37	28	33	20	17	128	73	31	29	26	25	23	24	53	24	22	22	22	20	19	55	17	4	-5	0	-7	-14	-21	-28	3	32	46	42	38	13	2	8
41 Nb	-53	16	11	15	-1	-11	135	81	36	34	32	32	29	30	69	30	28	27	28	26	25	61	23	4	0	-8	-26	-39	-53	-57	-32	11	26	17	12	20	4	9
40 Zr	-91	-20	-26	-25	-43	-60	101	52	13	12	10	10	9	9	42	9	8	8	7	7	36	6	0	3	-9	-35	-55	-76	-100	-74	-31	-19	-33	-40	4	-3	-3	
39 Y	-84	-29	-35	-36	-51	-68	62	20	0	0	0	0	0	14	0	0	0	0	0	10	0	11	27	24	4	-28	-53	-83	-74	43	-35	-48	-54	1	11	4		
38 Sr	-61	-27	-33	-37	-46	-66	14	0	14	15	16	18	17	0	17	18	18	20	20	1	21	50	73	70	39	13	-16	-59	-59	-45	-44	-56	-61	27	48	33		
37 Rb	9	7	1	4	-7	-24	0	9	52	53	55	55	57	56	13	56	58	58	57	59	59	16	61	98	125	129	98	72	44	9	-10	-11	-14	-23	-28	72	99	77
33 As	36	-8	4	-3	1	3	-12	-66	-81	-81	-80	-81	-80	-60	-80	-80	-80	-79	-79	-79	-59	-80	-75	-45	9	-5	-11	-19	-31	0	3	6	7		-82	-60	-62	
32 Ge	43.5	-17.5	-14.5	-13.5	-12.5	-10.5	-19.5	-63.5	-73.5	-73.5	-72.5	-101.5	-73.5	-72.5	-58.5	-72.5	-72.5	-71.5	-71.5	-71.5	-58.5	-71.5	-65.5	-37.5	-7.5	-14.5	-24.5	-37.5	-21.5	-11.5	-9.5	-7.5	-7.5		-72.5	-52.5	-56.5	
31 Ga	42	-5	-3	1	-1	14	-30	-41	-41	-41	-40	-41	-40		-40	-40	-40	-39	-39	-27	-34	-10	11	3	-7	-21	-19	1	6	5	4	-39	-29	-28				
30 Zn	-33	4	1	3	-1	15	-23	-31	-31	-31	-32	-31	-21	-31	-31	-30	-31	-30	-21	-31	-24	-3	15	8	-1	-13	-29	-16	1	6	5	4	-30	-15	-20			
29 Cu	-14	2	6	10	7	7	28	-9	-21	-21	-22	-23	-22	-10	-22	-23	-22	-23	-23	-12	-24	-17	2	22	18	10	0	-12	-9	8	15	15	15	-24	-7	-9		
28 Ni	0	15	2	4	-1		48	0	-27	-28	-30	-32	-31	-3	-31	-32	-31	-34	-34	-7	-36	-42	-29	-3	2	1	-2	-5	7	8	13	13	10	-39	-25	-26		
27 Co	-1	19	6	7	0	58	11	-17	-18	-20	-22	-23	-22	7	-22	-23	-23	-22	-24	-24	2	-27	-35	-24	-1	2	0	-3	-7	7	12	18	17	14	-30	-23	-19	
26 Fe	4	28	17	19	11	10	85	37	5	3	1	1	2	-1	30	-1	-3	-3	-2	5	-5	25	-7	-21	-15	0	0	4	9	-13	8	22	31	29		-11	-6	
25 Mn	-23	13	2	3	-7	-11	71	29	3	1	0	0	-2	23	-3	-3	-2	4	4	19	-6	-12	4	-1	6	-1	-9	-18	-21	4	11	7	3	9	-5	4		
24 Cr	-15	27	17	20	10	7	97	50	17	15	13	13	10	11	43	11	9	9	10	8	8	5	9	-7	-11	-18	-24	0	21	31	28	24	2	-3	2			
23 V	-35	17	9	12	3	-8	103	57	22	20	18	16	17	49	17	15	16	14	13	43	11	-1	-1	-13	-23	-34	45	-19	10	22	15	10	9	1	4			
22 Ti	-65	-2	-8	-5	-21	-33	104	57	20	18	17	15	15	48	15	14	14	13	12	41	11	0	1	-6	-25	-41	-57	-74	-47	-74	-10	2	8	-14	8	0	2	
21 Sc	-86	-28	-30	-30	-61	70	28	2	2	1	1	1	21	1	1	1	0	0	16	0	5	16	9	-17	-39	-62	-89	-57	-37	-28	-40	46	0	3	-1			
20 Ca	-63	-28	-32	-32	-45	19	1	8	9	10	10	11	0	11	12	11	13	13	0	14	39	57	28	4	-23	-55	-90	-43	-40	-52	-56	19	37	24				
19 K	9	7	1	4	-22	0	6	46	47	49	49	51	50	10	50	52	51	51	53	52	12	55	92	119	124	95	70	42	9	-9	-10	-13	-21	-24	65	93	72	
15 P	-36.5	-18.5	-11.5	-10.5	-7.5	-24.5	-86.5	-112.5	-112.5	-112.5	-114.5	-113.5	-79.5	-113.5	-113.5	-111.5	-113.5	-112.5	-80.5	-114.5	-117.5	-46.5	-32.5	-29.5	-30.5	-34.5	-4.5	-1.5	4.5	5.5	-119.5	-98.5	-95.5					
14 Si	-55	-20	-13	-10	-11	-8	-3	-52	-73	-73	-73	-74	-49	-74	-74	-74	-74		-56	-31	-31	-36	-43	-53	-30	-10	-4	-2	-2	-77	-66	-66						
13 Al	46	-4	3	7	4	2	28	-20	-38	-38	-38	-39	-19	-39	-38	-38	-38	-20	-39	-19	-2	9	-18	-30	-44	-22	4	11	10	10	-40	-33						
12 Mg	40	-10	-6	4	-9	-16	25	4	-7	-7	-6	-6	-6	-6	-5	-6	-5	-6	10	30	38	21	5	-13	-35	-32	-10	-3	-8	-10	-3	14	6					
11 Na	-15	0	-3	5	-8	-20	3	-3	24	25	26	26	28	28	0	28	29	28	30	30	2	31	89	97	73	52	26	-1	-14	-11	-11	-18	-20	65	49			
7 N	-62	-94	-91	-98	-90	-74	-155	-212	-235	-234	-233	-233	-202	-231	-232	-232	-231	-230	-229	-199	-231	-218	-173	-103	-72	-60	-54	-52	-81	-91	-82	-80	-237	-188				
6 C	-32	-32	-27	-27	-23	-13	-43	-90	-116	-116	-117	-116	-87	-117	-118	-116	-118	-117	-88	-119	-123	-101	-60	-42	-35	-32	-20	-20	-19	-13	-12	-123	-105	-102				
5 B	24	5	13	18	18	23	29	-19	-47	-48	-49	-51	-19	-50	-51	-50	-52	-52	-22	-54	-66	-31	-25	-24	26	-28	-2	19	27	30	31	-57	-50					
4 Be	8	6	11	16	15	18	-10	-29	-30	-31	-31	-33	-12	-32	-33	-32	-33	-14	-35	-37	-24	-3	0	-2	-5	-10	0	15	23	25	26	-37	-25					
3 Li	40	-16	-13	-12	1	-28	16	0	6	7	7	8	8	-1	8	9	8	9	9	-1	10	30	48	50	29	11	-9	-33	-19	-15	-21	-23	14	21				
1 H	-27	-10	-6	-6	-4	1	-16	-49	-60	-61	-61	-61	-45	-61	-61	-60	-61	-61	-24	-61	-63	-46	-18	-19	-20	-24	-8	-3	-1	0	-61	-53	-54					

Table A3. The values of ΔH_{ij} for atomic pairs between the element (Takeuchi & Inoue, 2005) (continued).

#	El	H	Li	Be	B	C	N	Na	Mg	Al	Si	P	K	Ca	Sc	Ti	V	Cr	Mn	Fe	Co	Ni	Cu	Zn	Ga	Ge	As	Rb	Sr	Y	Zr	Nb	Mo	Tc	Ru	Rh	(El)	(#)
45	Rh	-23	-14	-6	-25	-35	-63	19	-17	-32	-46	-34.5	31	-28	-61	-52	-29	-13	-16	-5	-2	-1	-2	-17	-25	-29.5	-24	33	-22	-54	-72	-46	-15	0	1	Rh	Pu	94
44	Ru	-20	5	-3	-24	-35	-61	44	0	-21	-38	-30.5	60	4	-44	-43	-25	-12	-11	-5	-1	0	7	-5	-11	-18.5	-14	62	5	-34	-39	-41	-14	0	Ru		U	92
43	Tc	-21	8	3	-25	-39	-68	47	3	-20	-38	-33.5	65	1	-39	-39	-21	9	8	-3	0	1	8	4	-10	-18.5	-15	67	10	-29	-53	-35	-11	Tc			Th	90
42	Mo	-28	49	-7	-34	-57	-115	93	36	-5	-35	-53.5	120	56	11	4	0	0	5	-2	-5	-7	7	12	7	-13.5	-16	125	69	24	-6	-6	Mo			Bi	83
41	Nb	-46	51	-25	-54	-102	-174	93	32	-18	-56	-89.5	123	63	18	2	-1	-7	4	-16	-25	-30	3	-1	-8	-36.5	-44	130	76	30	4	Nb			Pb	82
40	Zr	-69	27	-43	-71	-131	-233	59	6	-44	-84	-127.5	88	37	4	0	4	-12	-15	-25	-41	-49	-23	-29	-40	-72.5	-85	95	48	9	Zr			Tl	81
39	Y	-61	8	-32	-50	-117	-232	28	-6	-38	-73	-113.5	50	11	1	15	17	11	-1	-1	-22	-31	-22	-31	-40	-72.5	-80	56	17	Y			Hg	80
38	Sr	-45	0	-10	-18	-87	-206	-2	4	-18	-49	-81.5	7	1	25	53	54	47	27	34	10	-1	-9	-21	-27	-59.5	-61	10	Sr			Au	79
37	Rb	-16	13	28	28	-44	-154	2	23	25	-4	-24.5	0	15	64	100	100	94	69	83	57	47	27	14	13	-19.5	-12	Rb			Pt	78
33	As	-1	-29	7	10	-14	-59	-14	-21	-6	-17	-25	-51	-61	-77	-60	-36	-19	-31	-14	-18	-19	-3	-6	-6	-12.5	As				Ir	77
32	Ge	-14.5	-34.5	-3.5	-0.5	-29.5	-78.5	-14.5	-21.5	-26.5	-14.5	-17	-19.5	-59.5	-69.5	-51.5	-31.5	-18.5	-31.5	-21.5	-23.5	-11.5	-15.5	-15.5	-15.5	Ge				Os	76
31	Ga	-8	-9	5	6	-33	-95	5	4	1	-17	-18.5	-28	-38	-23	-8	-1	-13	-2	-11	-15	-1	0	4	Ga				Re	75
30	Zn	-7	-7	4	4	-32	-88	6	4	1	-18	-17.5	13	-22	-29	-15	-2	5	-6	4	-5	-9	1	Zn				W	74
29	Cu	-6	-5	0	0	-33	-84	16	-3	-1	-19	-17.5	25	-13	-24	9	5	12	4	13	6	4	Cu				Ta	73
28	Ni	-23	1	4	-24	-39	-69	32	4	-22	-40	-34.5	45	-7	-39	-35	-18	-7	-8	-2	0	Ni				Hf	72
27	Co	-23	8	4	-24	-42	-75	41	3	-19	-38	-35.5	55	2	-30	-28	-14	4	-5	-1	Co				Lu	71
26	Fe	-23	26	4	-26	-50	-87	62	18	-11	-35	-39.5	81	25	-11	-17	-7	-1	0	Fe				Yb	70
25	Mn	-34	19	-10	-32	-66	-119	49	10	-19	-45	-57.5	66	9	-8	-8	-1	2	Mn				Tm	69
24	Cr	-28	35	-7	-31	-61	-107	71	24	-10	-37	-49.5	81	38	7	-7	0	Cr				Er	68
23	V	-39	37	-16	-42	-82	-143	73	23	-16	-48	-73	96	44	7	-2	V				Ho	67
22	Ti	-54	34	-30	-58	-109	-190	68	16	-30	-66	-100.5	94	43	8	Ti				Dy	66
21	Sc	-60	12	-36	-55	-118	-224	34	-3	-38	-74	-112.5	58	17	Sc				Tb	65
20	Ca	-46	-1	-14	-22	-89	-201	1	-6	-51	-81.5	-24.5	12	Ca				Gd	64
19	K	-16	11	27	27	43	-152	1	20	13	-20.5	-24.5	K				Eu	63
15	P	0.5	-46.5	-3.5	0.5	-45	-24.5	-39.5	-20.5	-26.5	-26.5	P				Sm	62
14	Si	-19	-30	-15	-14	-39	-81	-26	-19	Si				Pm	61
13	Al	-8	4	0	0	-36	-92	13	Al				Nd	60
12	Mg	-19	0	-3	4	-55	-134	10	Mg				Pr	59
11	Na	-16	4	18	18	-45	-141	Na				Ce	58
7	N	-18	-145	-39	-28	-2	N				La	57
6	C	-3	-61	-15	-10	C				Cs	55
5	B	5	-6	0	B				Sb	51
4	Be	2	-5	Be				Sn	50
3	Li	-25	Li				In	49
1	H	H				Cd	48
(a)	Cd	0	0	-2	1	-36	-36	-36	-35	-35	-32	-35	-34	-34	-33	-33	-33	-31	-19	9	33	25	14	0	-18	-11	0	2	2	-1	-32	-8	-16	Cd	48
(b)	Ag	-2	-2	-3	4	8	-28	-30	-30	-30	-29	-30	-29	-29	-29	-29	-28	-30	-30	-13	15	43	38	28	16	-1	-6	-1	3	3	2	0	-6	Ag	47

| (b) | | 47 | 48 | 49 | 50 | 51 | 55 | 56 | 57 | 58 | 59 | 60 | 61 | 62 | 63 | 64 | 65 | 66 | 67 | 68 | 69 | 70 | 71 | 72 | 73 | 74 | 75 | 76 | 77 | 78 | 79 | 81 | 82 | 83 | 90 | 92 | 94 |
| | | Ag | Cd | In | Sn | Sb | Cs | Ba | La | Ce | Pr | Nd | Pm | Sm | Eu | Gd | Tb | Dy | Ho | Er | Tm | Yb | Lu | Hf | Ta | W | Re | Os | Ir | Pt | Au | Tl | Pb | Bi | Th | U | Pu |

Table A3. The values of ΔH_{ij} for atomic pairs between the element (Takeuchi & Inoue, 2005) (continued).

(a)		1	3	4	5	⋯	42	43	44	45	46	47	48	49	50	⋯	83	90	92	94	(a)	
(b)		H	Li	Be	B	⋯	Mo	Tc	Ru	Rh	Pd	Ag	Cd	In	Sn	⋯	Bi	Th	U	Pu		(b)
1	H		-25	2	5		-28	-21	-20	-23	-27	-10	-6	-6	-4		0	-61	-53	-54	H	1
3	Li	-25		-5	-6		49	8	5	-14	-40	-16	-13	-12	-18		-23	14	30	21	Li	3
4	Be	2			0		-7	-3	-3	-6	-8	6	11	16	15		26	-37	-27	-25	Be	4
5	B	5	-6	0			-34	-25	-24	-25	-24	5	13	18	18		31	-57	-54	-50	B	5
⋮	⋮																				⋮	⋮
42	Mo	-28	49	-7	-34			-11	-14	-15	-15	37	28	33	20		38	13	2	8	Mo	43
43	Tc	-21	8	-3	-25		-11		0	0	4	24	10	11	5		23	-39	-31	-26	Tc	44
44	Ru	-20	5	-3	-24		-14	0		1	6	23	9	10	4		23	-45	-36	-30	Ru	44
45	Rh	-23	-14	-6	-25		-15	0	1		2	10	-6	-8	-13		3	-63	-46	-42	Rh	45
46	Pd	-27	-40	-8	-24		-15	4	6	2		-7	-26	-31	-34		-21	-91	-59	-58	Pd	46
47	Ag	-10	-16	6	5		37	24	23	10	-7		-2	-2	-3		2	-29	0	-6	Ag	47
48	Cd	-6	-13	11	13		28	10	9	-6	-26	-2		0	0		1	-32	-8	-16	Cd	48
49	In	-6	-12	16	18		33	11	10	-8	-31	-2	0		0		-1	-31	-7	-16	In	49
50	Sn	-4	-18	15	18		20	5	4	-13	-34	-3	0	0			1	-48	-23	-30	Sn	50
⋮	⋮																				⋮	⋮
83	Bi	0	-23	26	31		38	23	23	3	-21	2	1	-1	1			-48	-15	-26	Bi	83
90	Th	-61	14	-37	-57		13	-39	-45	-63	-91	-29	-32	-31	-48		-48		4	0	Th	90
92	U	-53	30	-27	-54		2	-31	-36	-46	-59	0	-8	-7	-23		-15	4		1	U	92
94	Pu	-54	21	-25	-50		8	-26	-30	-42	-58	-6	-16	-16	-30		-26	0	1		Pu	94
(b)		H	Li	Be	B	⋯	Mo	Tc	Ru	Rh	Pd	Ag	Cd	In	Sn	⋯	Bi	Th	U	Pu		(b)
	(a)	1	3	4	5	⋯	42	43	44	45	46	47	48	49	50	⋯	83	90	92	94	(a)	

Table A4. Extracted features of each alloy.

Alloys	ΔH_{mix} (kJ/mol)	ΔS_{mix} J/(mol.K)	δ	VEC	$\Delta\chi$	T_m (K)	Ω
CoFeNi	-1.33	9.13	0.75	9.00	0.03	1767.33	12.11
CoFeNiSi0.25	-11.83	10.69	1.52	8.62	0.03	1760.85	1.59
CoFeNiSi0.5	-19.43	11.24	1.88	8.29	0.03	1755.29	1.02
CoFeNiSi0.75	-24.96	11.47	2.10	8.00	0.03	1750.47	0.80
Al0.25CoFeNi	-6.06	10.69	3.71	8.54	0.08	1703.19	3.00
Al0.5CoFeNi	-9.47	11.24	4.78	8.14	0.10	1648.21	1.96
Al0.75CoFeNi	-11.95	11.47	5.41	7.80	0.11	1600.57	1.54
CoCrFeNi	-3.75	11.53	1.03	8.25	0.10	1858.00	5.71
CoCrFeMo0.5Ni	-4.35	13.15	3.52	8.00	0.14	1973.33	5.97
CoCrFeNb0.103Ni	-5.47	12.21	2.75	8.17	0.10	1880.19	4.20
CoCrFeNb0.155Ni	-6.28	12.42	3.25	8.13	0.10	1890.98	3.74
CoCrFeNb0.206Ni	-7.02	12.59	3.66	8.09	0.11	1901.30	3.41
CoCrFeNb0.309Ni	-8.42	12.84	4.30	8.02	0.11	1921.39	2.93
CoCrFeNb0.412Ni	-9.69	13.03	4.81	7.95	0.11	1940.55	2.61
CoCrFeNiTi	-16.32	13.38	6.13	7.40	0.14	1875.40	1.54
Al0.25CoCrFeNi	-6.75	12.71	3.25	7.94	0.11	1803.62	3.39
Al0.375CoCrFeNi	-7.99	12.97	3.80	7.80	0.11	1778.76	2.89
Al0.5CoCrFeNi	-9.09	13.15	4.22	7.67	0.11	1755.28	2.54
Al0.75CoCrFeNi	-11.40	12.80	4.92	7.20	0.12	1711.96	1.92
Al0.875CoCrFeNi	-11.66	13.37	5.06	7.31	0.12	1692.06	1.94

Table A4. Extracted features of each alloy (continued).

Alloys	ΔH_{mix} (kJ/mol)	ΔS_{mix} J/(mol.K)	δ	VEC	$\Delta\chi$	T_m (K)	Ω
AlCoCrFeNi	-12.32	13.38	5.25	7.20	0.12	1673.10	1.82
Al1.25CoCrFeNi	-13.42	13.34	5.55	7.00	0.12	1637.88	1.63
Al1.5CoCrFeNi	-14.28	13.25	5.76	6.82	0.12	1605.86	1.49
Al2CoCrFeNi	-15.44	12.98	6.04	6.50	0.13	1549.83	1.30
Al2.5CoCrFeNi	-16.09	12.63	6.19	6.23	0.13	1502.42	1.18
Al3CoCrFeNi	-16.41	12.26	6.26	6.00	0.13	1461.79	1.09
AlC0.1CoCrFeNi	-15.35	13.92	6.61	7.14	0.16	1715.29	1.56
AlC0.2CoCrFeNi	-18.14	14.22	7.68	7.08	0.19	1755.87	1.38
AlC0.3CoCrFeNi	-20.70	14.43	8.56	7.02	0.21	1794.91	1.25
AlC0.4CoCrFeNi	-23.07	14.59	9.32	6.96	0.23	1832.50	1.16
AlC0.5CoCrFeNi	-25.26	14.70	9.99	6.91	0.25	1868.73	1.09
AlCCoCrFeNi	-33.89	14.90	12.43	6.67	0.31	2031.75	0.89
AlC1.5CoCrFeNi	-39.67	14.78	14.00	6.46	0.34	2169.69	0.81
Al0.5CoCrFeMo0.5Ni	-8.68	14.53	4.83	7.50	0.15	1869.35	3.13
AlCo0.5CrFeMo0.5Ni	-11.72	14.53	5.54	6.90	0.17	1785.90	2.21
AlCoCrFe0.5Mo0.5Ni	-12.56	14.53	5.65	7.00	0.17	1781.90	2.06
AlCoCrFe0.6Mo0.5Ni	-12.32	14.61	5.61	7.02	0.17	1782.41	2.11
AlCoCrFeMo0.1Ni	-12.13	13.92	5.30	7.18	0.13	1697.08	1.95
AlCoCrFeMo0.2Ni	-11.95	14.22	5.35	7.15	0.14	1720.13	2.05
AlCoCrFeMo0.3Ni	-11.78	14.43	5.40	7.13	0.15	1742.32	2.14

Table A4. Extracted features of each alloy (continued).

Alloys	ΔH_{mix} (kJ/mol)	ΔS_{mix} J/(mol.K)	δ	VEC	$\Delta\chi$	T_m (K)	Ω
AlCoCrFeMo0.4Ni	-11.60	14.59	5.44	7.11	0.15	1763.69	2.22
AlCoCrFeMo0.5Ni	-11.44	14.70	5.47	7.09	0.16	1784.27	2.29
AlCoCrFeMo0.5Ni1.5	-11.53	14.53	5.39	7.33	0.15	1779.42	2.24
AlCoCrFeMo0.5Ni2	-11.46	14.23	5.29	7.54	0.15	1775.31	2.21
AlCo1.5CrFeMo0.5Ni	-11.08	14.53	5.39	7.25	0.15	1782.92	2.34
AlCo2CrFeMo0.5Ni	-10.70	14.23	5.29	7.38	0.15	1781.77	2.37
AlCoCrFe1.5Mo0.5Ni	-10.50	14.53	5.30	7.17	0.15	1786.25	2.47
AlCoCrFe2Mo0.5Ni	-9.70	14.23	5.15	7.23	0.15	1787.92	2.62
Al1.5CoCrFeMo0.5Ni	-13.19	14.53	5.82	6.75	0.16	1713.38	1.89
Al2CoCrFeMo0.5Ni	-14.30	14.23	6.01	6.46	0.16	1653.38	1.65
AlCoCrFeNb0.1Ni	-13.32	13.92	5.50	7.16	0.12	1694.06	1.77
AlCoCrFeNb0.25Ni	-14.66	14.34	5.83	7.10	0.12	1724.00	1.69
AlCoCrFeNb0.5Ni	-16.53	14.70	6.24	7.00	0.13	1770.27	1.57
AlCoCrFeNb0.75Ni	-18.03	14.85	6.55	6.91	0.13	1812.52	1.49
AlCoCrFeNiSi0.2	-16.39	14.22	5.15	7.08	0.12	1673.48	1.45
AlCoCrFeNiSi0.4	-19.84	14.59	5.07	6.96	0.12	1673.83	1.23
AlCoCrFeNiSi0.6	-22.76	14.78	4.98	6.86	0.12	1674.16	1.09
AlCoCrFeNiSi0.8	-25.23	14.87	4.90	6.76	0.12	1674.47	0.99
AlCoCrFeNiSi	-27.33	14.90	4.82	6.67	0.12	1674.75	0.91
Al0.5CoCrFeNiTi	-19.50	14.53	6.33	7.25	0.14	1784.46	1.33

Table A4. Extracted features of each alloy (continued).

Alloys	ΔH_{mix} (kJ/mol)	ΔS_{mix} J/(mol.K)	δ	VEC	$\Delta\chi$	T_m (K)	Ω
AlCoCrFeNiTi1.5	-23.91	14.78	6.84	6.46	0.15	1735.85	1.07
Al1.5CoCrFeNiTi	-22.72	14.78	6.64	6.38	0.14	1658.04	1.08
Al2CoCrFeNiTi	-23.35	14.53	6.64	6.14	0.14	1606.29	1.00
CoCrFeMnNi	-4.16	13.38	0.92	8.00	0.14	1790.00	5.76
CoCrFeMnNiV0.25	-5.30	14.34	1.73	7.86	0.14	1807.76	4.89
CoCrFeMnNiV0.5	-6.21	14.70	2.18	7.73	0.14	1823.91	4.31
CoCrFeMnNiV0.75	-6.96	14.85	2.48	7.61	0.14	1838.65	3.93
CoCrFeMnNiV1.0	-7.56	14.90	2.70	7.50	0.14	1852.17	3.65
Al0.10CoCrFeMnNi	-5.24	13.92	2.05	7.90	0.14	1773.21	4.71
Al0.20CoCrFeMnNi	-6.24	14.22	2.70	7.81	0.14	1757.06	4.00
Al0.38CoCrFeMnNi	-7.85	14.56	3.49	7.65	0.14	1729.50	3.21
Al0.43CoCrFeMnNi	-8.25	14.62	3.66	7.60	0.14	1722.17	3.05
Al0.49CoCrFeMnNi	-8.72	14.69	3.85	7.55	0.14	1713.55	2.89
Al0.56CoCrFeMnNi	-9.23	14.75	4.05	7.50	0.14	1703.73	2.72
Al0.62CoCrFeMnNi	-9.65	14.79	4.20	7.45	0.14	1695.51	2.60
Al0.68CoCrFeMnNi	-10.05	14.82	4.34	7.40	0.14	1687.46	2.49
Al0.75CoCrFeMnNi	-10.50	14.85	4.49	7.35	0.14	1678.28	2.38
Al0.81CoCrFeMnNi	-10.86	14.87	4.60	7.30	0.14	1670.59	2.29
Al0.88CoCrFeMnNi	-11.25	14.89	4.73	7.25	0.14	1661.82	2.20
Al0.95CoCrFeMnNi	-11.63	14.90	4.85	7.20	0.14	1653.25	2.12

Table A4. Extracted features of each alloy (continued).

Alloys	ΔH_{mix} (kJ/mol)	ΔS_{mix} J/(mol.K)	δ	VEC	$\Delta\chi$	T_m (K)	Ω
Al1.25CoCrFeMnNi	-13.03	14.87	5.25	7.00	0.14	1618.70	1.85
CoCrNi	-4.89	9.13	1.12	8.33	0.11	1874.67	3.50
CoMnNi	-5.78	9.13	0.38	8.67	0.16	1670.67	2.64
FeMnNi	-4.44	9.13	0.65	8.33	0.15	1684.00	3.46
CoCrMnNi	-5.50	11.53	0.97	8.00	0.15	1785.50	3.74
CoFeMnNi	-4.00	11.53	0.66	8.50	0.14	1705.00	4.91
Al0.5CrFe1.5MnNi0.5	-7.26	12.66	4.04	7.00	0.13	1708.83	2.98
Al0.3CrFe1.5MnNi0.5	-5.51	12.32	3.32	7.19	0.13	1744.90	3.90
AlCrFeNi	-13.25	11.53	5.47	6.75	0.12	1649.38	1.43
AlCrFeNiMo0.2	-12.65	12.57	5.53	6.71	0.15	1708.74	1.70
AlCrFeNiMo0.5	-11.85	13.15	5.57	6.67	0.17	1787.89	1.98
CoCrCu0.5FeNi	0.49	13.15	1.06	8.56	0.09	1802.29	47.98
CoCrCuFeNi	3.20	13.38	1.07	8.80	0.09	1757.72	7.35
CoCrCuFeNiTi0.5	-3.70	14.70	4.46	8.36	0.12	1774.75	7.04
CoCrCuFeNiTi0.8	-6.75	14.87	5.27	8.14	0.13	1783.55	3.93
CoCrCuFeNiTi	-8.44	14.90	5.65	8.00	0.14	1788.93	3.16
Al0.25CoCrCu0.5FeNiTi0.5	-9.09	15.43	5.14	7.86	0.13	1774.51	3.01
Al0.25CoCrCu0.75FeNiTi0.5	-7.28	15.55	5.02	8.00	0.13	1755.51	3.75
Al0.3CoCrCuFeNi	0.16	14.43	3.15	8.47	0.10	1711.07	157.65
Al0.5CoCrCuFeNi	-1.52	14.70	3.82	8.27	0.11	1682.79	16.26

Table A4. Extracted features of each alloy (continued).

Alloys	ΔH_{mix} (kJ/mol)	ΔS_{mix} J/(mol.K)	δ	VEC	$\Delta\chi$	T_m (K)	Ω
AlCoCrCuFeNi	-4.78	14.90	4.82	7.83	0.12	1620.35	5.05
AlCoCrCuFeMnNi	-5.63	16.18	4.57	7.71	0.14	1605.73	4.61
AlCoCrCuFeNiTi	-13.80	16.18	6.23	7.29	0.14	1666.73	1.95
AlCoCrCuFeNiV	-7.76	16.18	4.69	7.43	0.12	1697.87	3.54
Al1.3CoCrCuFeNi	-6.24	14.85	5.19	7.60	0.12	1587.64	3.78
Al1.5CoCrCuFeNi	-7.05	14.78	5.38	7.46	0.12	1567.52	3.29
Al1.8CoCrCuFeNi	-8.08	14.64	5.60	7.26	0.13	1539.54	2.79
Al2CoCrCuFeNi	-8.65	14.53	5.71	7.14	0.13	1522.23	2.56
Al2.3CoCrCuFeNi	-9.38	14.35	5.84	6.97	0.13	1498.03	2.29
Al2.5CoCrCuFeNi	-9.78	14.21	5.91	6.87	0.13	1482.98	2.16
Al2.8CoCrCuFeNi	-10.28	14.01	5.99	6.72	0.13	1461.85	1.99
Al0.5CoCrCu0.5FeNiTi0.5	-10.84	15.74	5.47	7.64	0.13	1736.28	2.52
Al0.5CoCrCuFeNiTi0.2	-4.15	15.45	4.52	8.12	0.12	1691.99	6.30
Al0.5CoCrCuFeNiTi0.4	-6.42	15.76	5.04	7.98	0.13	1700.57	4.17
Al0.5CoCrCuFeNiTi0.6	-8.40	15.92	5.45	7.85	0.13	1708.58	3.24
Al0.5CoCrCuFeNiTi0.8	-10.11	16.00	5.76	7.73	0.14	1716.09	2.72
Al0.5CoCrCuFeNiTi	-11.60	16.01	6.02	7.62	0.14	1723.13	2.38
Al0.5CoCrCuFeNiTi1.2	-12.89	15.97	6.22	7.51	0.14	1729.75	2.14
Al0.5CoCrCuFeNiTi1.4	-14.02	15.91	6.39	7.41	0.15	1735.99	1.97
Al0.5CoCrCuFeNiTi1.6	-15.00	15.82	6.53	7.31	0.15	1741.88	1.84

Table A4. Extracted features of each alloy (continued).

Alloys	ΔH_{mix} (kJ/mol)	ΔS_{mix} J/(mol.K)	δ	VEC	$\Delta \chi$	T_m (K)	Ω
Al0.5CoCrCuFeNiTi1.8	-15.86	15.72	6.65	7.22	0.15	1747.45	1.73
Al0.5CoCrCuFeNiTi2	-16.60	15.60	6.74	7.13	0.15	1752.71	1.65
Al0.5CoCrCuFeNiV0.2	-2.50	15.45	3.87	8.16	0.11	1699.64	10.50
Al0.5CoCrCuFeNiV0.4	-3.34	15.76	3.91	8.05	0.12	1715.35	8.09
Al0.5CoCrCuFeNiV0.6	-4.07	15.92	3.94	7.95	0.12	1730.02	6.76
Al0.5CoCrCuFeNiV0.8	-4.71	16.00	3.97	7.86	0.12	1743.77	5.93
Al0.5CoCrCuFeNiV1.0	-5.25	16.01	3.98	7.77	0.12	1756.67	5.35
Al0.5CoCrCuFeNiV1.2	-5.73	15.97	3.99	7.69	0.12	1768.80	4.93
Al0.5CoCrCuFeNiV1.4	-6.14	15.91	4.00	7.61	0.12	1780.22	4.61
Al0.5CoCrCuFeNiV1.6	-6.50	15.82	4.00	7.54	0.12	1791.01	4.36
Al0.5CoCrCuFeNiV1.8	-6.81	15.72	3.99	7.47	0.12	1801.20	4.16
Al0.5CoCrCuFeNiV2.0	-7.08	15.60	3.99	7.40	0.12	1810.85	3.99
Al0.75CoCrCu0.25FeNiTi0.5	-14.39	15.55	5.83	7.27	0.13	1717.05	1.86
AlCoCrCuNiTi	-16.67	14.90	6.50	7.17	0.15	1643.18	1.47
AlCoFeNi	-13.75	11.53	5.81	7.50	0.12	1558.88	1.31
AlCoFeNiTiVZr	-31.43	16.18	8.87	6.14	0.19	1781.64	0.92
CoCuFeNi	5.00	11.53	1.03	9.50	0.03	1664.65	3.84
CoCuFeNiSn0.02	5.02	11.73	1.90	9.47	0.03	1658.88	3.88
CoCuFeNiSn0.04	5.04	11.87	2.47	9.45	0.03	1653.17	3.90
CoCuFeNiSn0.05	5.05	11.94	2.71	9.43	0.03	1650.33	3.90

Table A4. Extracted features of each alloy (continued).

Alloys	ΔH_{mix} (kJ/mol)	ΔS_{mix} J/(mol.K)	δ	VEC	$\Delta\chi$	T_m (K)	Ω
CoCuFeNiSn0.07	5.07	12.05	3.12	9.41	0.03	1644.71	3.91
CoCuFeNiSn0.1	5.09	12.20	3.64	9.37	0.03	1636.37	3.92
CoCuFeNiSn0.2	5.17	12.57	4.90	9.24	0.03	1609.43	3.91
CoCuFeNiSn0.5	5.33	13.15	7.04	8.89	0.04	1535.81	3.79
AlCoCuFeNi	-5.28	13.38	5.25	8.20	0.11	1518.42	3.85
AlCoCuFeNiSi	-20.44	14.90	4.82	7.50	0.11	1545.85	1.13
AlCoCuFeNiTi	-16.89	14.90	6.58	7.50	0.15	1589.52	1.40
AlCoCuFeNiZr	-23.89	14.90	9.59	7.50	0.21	1620.02	1.01
CoCuFeMnNi	1.76	13.38	0.92	9.00	0.13	1635.32	12.43
CoCuFeMnNiSn0.03	1.77	13.60	1.98	8.97	0.13	1628.58	12.50
CoCuFeMnNiSn0.05	1.78	13.71	2.43	8.95	0.13	1624.13	12.51
CoCuFeMnNiSn0.08	1.79	13.84	2.97	8.92	0.14	1617.52	12.50
CoCuFeMnNiSn0.10	1.80	13.92	3.28	8.90	0.14	1613.16	12.48
CoCuFeMnNiSn0.20	1.83	14.22	4.44	8.81	0.14	1591.85	12.34
CrCuFeMnNi	2.72	13.38	0.92	8.40	0.14	1707.72	8.40
CrCuFeMoNi	4.64	13.38	4.10	8.20	0.16	1983.32	5.72
AlCrCuFeNi0.6	-2.68	13.24	5.03	7.39	0.12	1579.07	7.79
AlCrCuFeNi0.8	-3.40	13.35	5.01	7.50	0.12	1585.19	6.22
AlCrCuFeNi	-4.00	13.38	4.99	7.60	0.12	1590.82	5.32
AlCrCuFeNi1.2	-4.50	13.36	4.96	7.69	0.12	1596.02	4.74

Table A4. Extracted features of each alloy (continued).

Alloys	ΔH_{mix} (kJ/mol)	ΔS_{mix} J/(mol.K)	δ	VEC	$\Delta\chi$	T_m (K)	Ω
AlCrCuFeNi1.4	-4.91	13.30	4.93	7.78	0.12	1600.83	4.33
AlCrCuFeNi2	-5.78	12.98	4.82	8.00	0.12	1613.35	3.62
AlCrCuFeNiTi	-13.67	14.90	6.29	7.00	0.14	1649.85	1.80
Al0.2CrCuFeNi2	0.12	12.01	2.69	8.77	0.10	1717.94	174.27
Al0.4CrCuFeNi2	-1.70	12.45	3.51	8.56	0.11	1688.89	12.36
Al0.6CrCuFeNi2	-3.27	12.72	4.08	8.36	0.12	1661.91	6.47
Al0.8CrCuFeNi2	-4.61	12.88	4.50	8.17	0.12	1636.79	4.57
Al1.2CrCuFeNi2	-6.78	13.02	5.08	7.84	0.13	1591.42	3.05
Al1.125CuFe0.75NiTi1.125	-19.96	13.30	6.68	6.98	0.16	1535.38	1.02
Al22.5Cu20Fe15Ni20Ti22.5	-19.96	13.30	6.68	6.98	0.16	1535.38	1.02
AlCuFeNiTi	-17.60	13.38	6.61	7.20	0.15	1553.82	1.18
AlCuNiTi	-23.25	11.53	6.73	7.00	0.17	1490.28	0.74
AlCr0.5NbTiV	-15.41	13.15	4.55	4.44	0.04	1966.33	1.68
AlCrNbTiV	-14.56	13.38	5.19	4.60	0.04	1982.70	1.82
Al0.4Hf0.6NbTaTiZr	-6.33	14.50	4.11	4.32	0.12	2396.76	5.49
Al0.3HfNbTaTiZr	-3.99	14.43	4.13	4.32	0.12	2432.46	8.80
Al0.5HfNbTaTiZr	-7.67	14.70	4.18	4.27	0.12	2377.95	4.56
Al0.75HfNbTaTiZr	-11.86	14.85	4.30	4.57	0.26	2383.33	2.99
AlHfNbTaTiZr	-14.78	14.90	4.27	4.17	0.12	2257.58	2.28
AlMo0.5NbTa0.5TiZr	-16.84	14.53	4.34	4.30	0.22	2168.60	1.87

Table A4. Extracted features of each alloy (continued).

Alloys	ΔH_{mix} (kJ/mol)	ΔS_{mix} J/(mol.K)	δ	VEC	$\Delta\chi$	T_m (K)	Ω
Al0.25MoNbTiV	-6.26	12.71	3.31	4.88	0.24	2348.09	4.77
Al0.5MoNbTiV	-8.99	13.15	3.22	4.78	0.24	2269.50	3.32
Al0.75MoNbTiV	-11.12	13.33	3.14	4.68	0.23	2199.18	2.64
AlMoNbTiV	-11.56	14.90	3.05	4.67	0.22	2328.75	3.00
Al0.25NbTaTiV	-4.82	12.71	3.42	4.65	0.05	2441.50	6.44
Al0.5NbTaTiV	-8.40	13.15	3.33	4.56	0.05	2357.72	3.69
AlNbTaTiV	-13.44	13.38	3.16	4.40	0.05	2215.30	2.21
Al0.3NbTa0.8Ti1.4V0.2Zr1.3	-4.86	13.46	4.50	4.34	0.11	2315.69	6.41
Al0.5NbTa0.8Ti1.5V0.2Zr	-8.62	13.78	4.23	4.30	0.10	2264.25	3.62
Al0.3NbTaTi1.4Zr1.3	-4.41	12.63	4.06	4.34	0.10	2360.89	6.76
AlNb1.5Ta0.5Ti1.5Zr0.5	-15.12	12.51	3.07	4.20	0.08	2134.90	1.77
AlNbTiV	-16.25	11.53	3.30	4.25	0.03	1945.88	1.38
CrMo0.5NbTa0.5TiZr	-4.92	14.53	7.13	4.90	0.22	2407.90	7.11
CrNbTiVZr	-4.64	13.38	7.67	4.80	0.12	2221.60	6.41
CrNbTiZr	-5.00	11.53	7.84	4.75	0.12	2236.25	5.15
FeMoNiTiVZr	-19.78	14.90	8.56	6.17	0.27	2111.00	1.59
Hf0.5Mo0.5NbTiZr	0.10	14.43	4.57	4.56	0.24	2480.33	362.28
Hf0.5Mo0.5NbSi0.1TiZr	-6.47	13.60	5.09	4.49	0.26	2361.78	4.96
Hf0.5Mo0.5NbSi0.3TiZr	-17.22	14.17	5.61	4.47	0.27	2330.21	1.92
Hf0.5Mo0.5NbSi0.5TiZr	-26.07	14.43	6.01	4.44	0.27	2301.44	1.27

Table A4. Extracted features of each alloy (continued).

Alloys	ΔH_{mix} (kJ/mol)	ΔS_{mix} J/(mol.K)	δ	VEC	$\Delta\chi$	T_m (K)	Ω
Hf0.5Mo0.5NbSi0.7TiZr	-33.39	14.53	6.32	4.43	0.27	2275.13	0.99
Hf0.5Mo0.5NbSi0.9TiZr	-39.45	14.55	6.57	4.41	0.27	2250.96	0.83
Hf0.5Mo0.5NbTiZrC0.1	-10.64	13.60	7.70	4.49	0.30	2414.02	3.09
Hf0.5Mo0.5NbTiZrC0.3	-28.58	14.17	11.22	4.47	0.36	2479.65	1.23
HfMo0.25NbTaTiZr	1.56	14.34	4.26	4.48	0.19	2540.19	23.34
HfMo0.5NbTaTiZr	0.60	14.70	4.45	4.55	0.23	2556.36	63.14
HfMo0.75NbTaTiZr	-0.21	14.85	4.59	4.61	0.26	2571.13	180.40
HfMoNbTaTiZr	-0.89	14.90	4.70	4.67	0.28	2584.67	43.32
HfMoTaTiZr	-1.92	13.38	5.06	4.60	0.31	2553.20	17.79
HfMoNbZrTi	-1.60	13.38	5.06	4.60	0.31	2443.00	20.43
HfNbTaZr	3.50	11.53	3.95	4.50	0.12	2666.75	8.78
Hf0.5Nb0.5Ta0.5Ti1.5Zr	1.88	12.42	4.17	4.25	0.11	2328.75	15.43
HfNbTaTiZr	2.72	13.38	4.01	4.40	0.12	2522.40	12.41
MoNbTaV	-3.25	11.53	3.56	5.25	0.26	2773.50	9.84
MoNbTaVW	-4.64	13.38	3.21	5.40	0.34	2957.80	8.53
MoNbTaW	-6.50	11.53	2.27	5.50	0.36	3156.50	5.60
MoNbTiV	-2.75	11.53	3.41	5.00	0.25	2436.50	10.21
Mo0.3NbTiVZr	-1.25	12.83	5.96	4.60	0.20	2289.95	23.41
Mo0.5NbTiVZr	-1.78	13.15	5.91	4.67	0.23	2316.89	17.13
Mo0.7NbTiVZr	-2.21	13.31	5.85	4.72	0.25	2341.53	14.11

Table A4. Extracted features of each alloy (continued).

Alloys	ΔH_{mix} (kJ/mol)	ΔS_{mix} J/(mol.K)	δ	VEC	$\Delta\chi$	T_m (K)	Ω
MoNbTiVZr	-2.72	13.38	5.77	4.80	0.27	2374.80	11.68
Mo1.3NbTiVZr	-3.10	13.33	5.69	4.87	0.29	2404.30	10.32
Mo1.5NbTiVZr	-3.31	13.25	5.63	4.91	0.30	2422.18	9.71
Mo1.7NbTiVZr	-3.47	13.15	5.58	4.95	0.31	2438.81	9.24
Mo2NbTiVZr	-3.67	12.98	5.50	5.00	0.31	2461.67	8.71
MoNbTiV0.25Zr	-2.60	12.71	5.24	4.76	0.30	2412.18	11.78
MoNbTiV0.50Zr	-2.67	13.15	5.48	4.78	0.29	2398.33	11.82
MoNbTiV0.75Zr	-2.70	13.33	5.65	4.79	0.28	2385.95	11.77
MoNbTiV1.0Zr	-2.72	13.38	5.77	4.80	0.27	2374.80	11.68
MoNbTiV1.5Zr	-2.71	13.25	5.93	4.82	0.26	2355.55	11.52
MoNbTiV2.0Zr	-2.67	12.98	6.00	4.83	0.25	2339.50	11.38
MoNbTiV3.0Zr	-2.53	12.26	6.01	4.86	0.23	2314.29	11.22
MoNbTiZr	-2.50	11.53	4.92	4.75	0.31	2427.75	11.19
NbTaTiV	-0.25	11.53	3.53	4.75	0.05	2535.75	116.90
NbTaVW	-4.50	11.53	3.49	5.25	0.34	2973.25	7.62
NbTiV0.3Zr	1.43	10.84	5.10	4.39	0.12	2261.79	17.11
NbTiV0.3Mo0.1Zr	0.80	11.62	5.15	4.44	0.16	2280.44	33.30
NbTiV0.3Mo0.3Zr	-0.28	12.32	5.22	4.53	0.21	2314.64	102.64
NbTiV0.3Mo0.5Zr	-1.14	12.65	5.26	4.61	0.25	2345.24	26.12
NbTiV0.3Mo0.7Zr	-1.83	12.80	5.29	4.68	0.27	2372.78	16.64

Table A4. Extracted features of each alloy (continued).

Alloys	ΔH_{mix} (kJ/mol)	ΔS_{mix} J/(mol.K)	δ	VEC	$\Delta\chi$	T_m (K)	Ω
NbTiV0.3MoZr	-2.62	12.83	5.30	4.77	0.30	2409.28	11.80
NbTiV0.3Mo1.3Zr	-3.19	12.72	5.28	4.85	0.31	2441.02	9.72
NbTiV0.3Mo1.5Zr	-3.49	12.61	5.26	4.90	0.32	2459.98	8.89
NbTiVZr	-0.25	11.53	6.03	4.50	0.12	2244.50	103.48
NbTiV2Zr	-1.28	11.08	6.41	4.60	0.11	2228.20	19.28
CoFeNiV	-10.50	11.53	3.22	8.00	0.11	1866.25	2.05
CoFeMo0.2NiV	-10.16	12.57	3.71	7.90	0.13	1915.29	2.37
CoFeMo0.4NiV	-9.83	13.01	4.05	7.82	0.14	1959.86	2.59
CoFeMo0.6NiV	-9.53	13.24	4.31	7.74	0.16	2000.57	2.78
CoFeMo0.8NiV	-9.24	13.35	4.50	7.67	0.16	2037.88	2.95
CoFeMoNiV	-8.96	13.38	4.64	7.60	0.17	2072.20	3.09
CoFeMoNi1.2V	-8.70	13.36	4.75	7.54	0.17	2103.88	3.23
CoFeMoNi1.4V	-8.45	13.30	4.84	7.48	0.18	2133.22	3.36
CoFeMoNi1.6V	-8.21	13.21	4.90	7.43	0.18	2160.46	3.47
CoFeMoNi1.8V	-7.99	13.10	4.95	7.38	0.18	2185.83	3.58
CoFeMoNi2V	-7.78	12.98	4.99	7.33	0.19	2209.50	3.69
CoCrCuFeMnNiTiV	-8.00	17.29	5.19	7.50	0.15	1801.83	3.89
Al11.1(CoCrCuFeMnNiTiV)88.9	-12.05	18.27	5.75	7.00	0.15	1705.34	2.59
Al20(CoCrCuFeMnNiTiV)80	-14.88	17.99	6.01	6.60	0.14	1628.16	1.97
Al40(CoCrCuFeMnNiTiV)60	-17.97	15.97	6.09	5.70	0.13	1454.50	1.29

Table A4. Extracted features of each alloy (continued).

Alloys	ΔH_{mix} (kJ/mol)	ΔS_{mix} J/(mol.K)	δ	VEC	$\Delta\chi$	T_m (K)	Ω
AlFeNiTiVZr	-31.33	14.90	8.60	5.67	0.19	1783.92	0.85
(CuMnNi)75Zn25	-3.50	11.53	4.32	10.00	0.16	1323.33	4.36
(CuMnNi)80Zn20	-2.98	11.47	4.02	9.87	0.16	1365.54	5.25
(CuMnNi)90Al10	-5.16	9.35	3.40	9.19	0.15	1568.21	2.84
(CuMnNi)90Sn10	-1.69	9.35	5.61	9.25	0.14	1541.43	8.54
(CuMnNi)95Al5	-2.66	10.33	3.01	9.02	0.17	1503.56	5.84
(CuMnNi)95Sn5	-0.25	10.33	4.98	9.07	0.17	1482.16	60.48
AlCoCrFeNi	-12.32	13.38	5.25	7.20	0.12	1673.10	1.82
Al0.8CoCrFeNiTi0.2	-13.89	14.21	5.44	7.24	0.13	1713.56	1.75
Al0.5CoCrFeNiTi0.5	-17.10	14.69	5.99	7.06	0.13	1726.66	1.48
CrMnFeV	-2.25	11.53	2.74	6.50	0.10	1904.75	9.76
CrMn0.3FeVCu0.06	-2.77	11.39	2.74	6.48	0.10	1975.53	8.11
CrMn0.5FeVCu0.1	-1.94	11.98	2.73	6.56	0.10	1943.24	11.97
CrMn0.7FeVCu0.14	-1.28	12.32	2.72	6.63	0.11	1914.98	18.48
CrMnFeVCu0.2	-0.50	12.57	2.68	6.71	0.11	1878.65	47.33
AlCrFe2Mo0.2Nb0.2Ni2	-12.52	12.67	5.43	7.38	0.13	1759.23	1.78
AlCrFe2Mo0.5Nb0.5Ni2	-14.00	13.71	5.94	7.21	0.15	1850.07	1.81
Al25Co14Cr5Fe14Mn8Ni25Ti9	-21.87	15.07	6.53	6.85	0.15	1568.51	1.08
CoFeMnNiV	-8.96	13.38	2.96	7.80	0.14	1796.60	2.68
(FeCoNi)0.75Cr0.25	-3.75	11.53	1.03	8.25	0.10	1858.00	5.71

Table A4. Extracted features of each alloy (continued).

Alloys	ΔH_{mix} (kJ/mol)	ΔS_{mix} J/(mol.K)	δ	VEC	$\Delta\chi$	T_m (K)	Ω
(FeCoNi)0.75Cr0.05Cu0.2	3.73	12.57	1.03	9.25	0.06	1703.32	5.74
AlCoCr0.5Fe2.5Ni2.5	-14.68	8.48	6.52	6.44	0.14	1368.59	0.79
Al0.25CoCrFeNi	-6.75	12.71	3.25	7.94	0.11	1803.62	3.39
Ti2ZrNbHfVAl1.2	-16.31	14.56	5.36	4.11	0.13	2020.44	1.80
Nb25Ti35V5Zr35	1.63	10.24	4.87	4.30	0.12	2219.20	13.94
(Ti28Zr40Al20Nb12)97Si3	-27.51	11.60	5.28	3.92	0.14	1904.68	0.80
(Ti28Zr40Al20Nb12)95Si5	-31.29	11.91	5.52	3.92	0.15	1900.11	0.72
(Ti28Zr40Al20Nb12)90Si10	-39.94	12.42	6.05	3.93	0.17	1888.69	0.59
Ti2NbVAl0.3Zr0.25	-6.28	11.36	4.13	4.37	0.07	2111.44	3.82
Ti2NbVAl0.3Zr0.5	-6.22	11.87	4.66	4.35	0.08	2112.30	4.03
Ti2NbVAl0.3Zr0.75	-6.13	12.13	5.03	4.34	0.10	2113.08	4.18
Ti2NbVAl0.3Zr	-6.04	12.26	5.29	4.32	0.10	2113.78	4.29
Ti1(NbMoTa)2W0.5	-4.43	13.17	2.03	5.05	0.34	2785.26	8.29
FeMn40Co10Cr10C0.5Ti2	-4.92	9.03	3.73	7.06	0.15	1689.93	3.10
Co68Al18.2Fe6.5V4.75Cr2.55	-13.03	8.22	5.40	7.58	0.11	1646.71	1.04
Ti1·6ZrNbMo0.35	0.19	10.61	4.36	4.43	0.21	2277.37	124.03
Cr14Mn20Fe20Co20Ni26	-4.40	13.23	0.88	8.24	0.14	1765.76	5.31
Cr24Mn20Fe20Co20Ni16	-3.89	13.31	0.94	7.84	0.14	1806.16	6.18
Cr26Mn20Fe20Co20Ni14	-3.72	13.23	0.95	7.76	0.14	1814.24	6.45
Al21Co19.5Fe9.5Ni50	-13.68	10.12	5.63	8.15	0.12	1575.56	1.16

Table A4. Extracted features of each alloy (continued).

Alloys	ΔH_{mix} (kJ/mol)	ΔS_{mix} J/(mol.K)	δ	VEC	$\Delta\chi$	T_m (K)	Ω
(Ti32.5V27.5Zr7.5Nb32.5)97Ni3	-3.62	11.44	5.16	4.76	0.10	2261.15	7.15
(Ti32.5V27.5Zr7.5Nb32.5)94Ni6	-6.86	11.89	5.59	4.92	0.11	2244.60	3.89
(Ti32.5V27.5Zr7.5Nb32.5)91Ni9	-9.88	12.20	5.96	5.09	0.12	2228.05	2.75
Al0.2CoCrFeNi	-6.21	12.57	2.98	8.00	0.10	1813.98	3.67
Al0.2CoCrFeNiSc0.3	-10.54	13.77	7.66	7.67	0.15	1813.98	2.37
CoCrFeMnNiMo5C0.5	-13.33	13.45	8.44	6.86	0.25	2413.57	2.44
Al0.25FeCoNiV	-13.07	12.71	4.14	7.71	0.12	1811.38	1.76
Co30Cr10Fe10Al18Ni30Mo2	-12.46	13.05	5.36	7.76	0.13	1667.95	1.75
CoCr1.3FeNi0.7MnNb0.3	-6.52	14.29	3.85	7.60	0.14	1866.75	4.09
CoCr1.3FeNi0.7MnNb0.45	-7.68	14.51	4.52	7.53	0.14	1890.84	3.57
Al14Co41Cr15Fe10Ni20	-10.49	12.28	4.75	7.81	0.11	1701.07	1.99
Cr15Cu5Fe20Mn25Ni35	-3.29	12.22	0.90	8.30	0.15	1732.53	6.44
Al0.65CoCrFe2Ni	-8.33	12.77	4.22	7.60	0.11	1742.79	2.67
Al40Co15Cr15Fe15Ni15	-16.23	12.51	6.22	6.15	0.13	1488.20	1.15
AlFeCoNiMo0.2	-13.33	12.57	5.88	7.43	0.14	1622.55	1.53
Ti1.6ZrNbAl0.3	-6.79	10.48	4.13	4.18	0.11	2118.47	3.27
Ti1.6ZrNbAl0.4	-9.20	10.72	4.14	4.15	0.11	2088.85	2.43
Ti1.6ZrNbAl0.5	-11.37	10.91	4.15	4.12	0.11	2060.67	1.98
Ti1.6ZrNbAl0.6	-13.33	11.05	4.15	4.10	0.11	2033.83	1.69
AlNbTiVCr	-14.56	13.38	5.19	4.60	0.04	1982.70	1.82

Table A4. Extracted features of each alloy (continued).

Alloys	ΔH_{mix} (kJ/mol)	ΔS_{mix} J/(mol.K)	δ	VEC	$\Delta\chi$	T_m (K)	Ω
(AlNbTiVCr)99.5Ni0.5	-14.86	13.58	5.23	4.63	0.04	1981.42	1.81
(AlNbTiVCr)99Ni1	-15.16	13.71	5.27	4.65	0.05	1980.13	1.79
(AlNbTiVCr)98.5Ni1.5	-15.45	13.83	5.31	4.68	0.05	1978.85	1.77
(AlNbTiVCr)98Ni2	-15.74	13.93	5.35	4.71	0.06	1977.57	1.75
CrFeNbTiMo	-9.28	13.38	6.24	5.80	0.22	2304.20	3.32
Fe3.5Ni3.5Cr2MnAl0.7	-7.43	12.01	3.34	7.86	0.13	1757.05	2.84
(AlCoCrFeNi)99.5Zr0.5	-12.88	13.58	5.48	7.18	0.12	1675.37	1.77
(AlCoCrFeNi)99Zr1	-13.43	13.71	5.71	7.17	0.13	1677.65	1.71
(AlCoCrFeNi)98Zr2	-14.51	13.93	6.12	7.14	0.13	1682.20	1.61
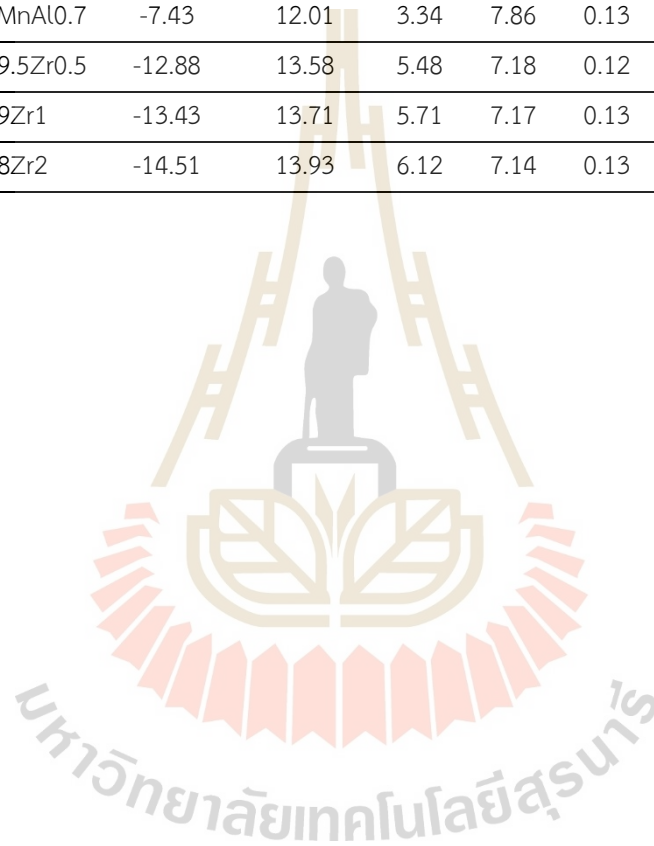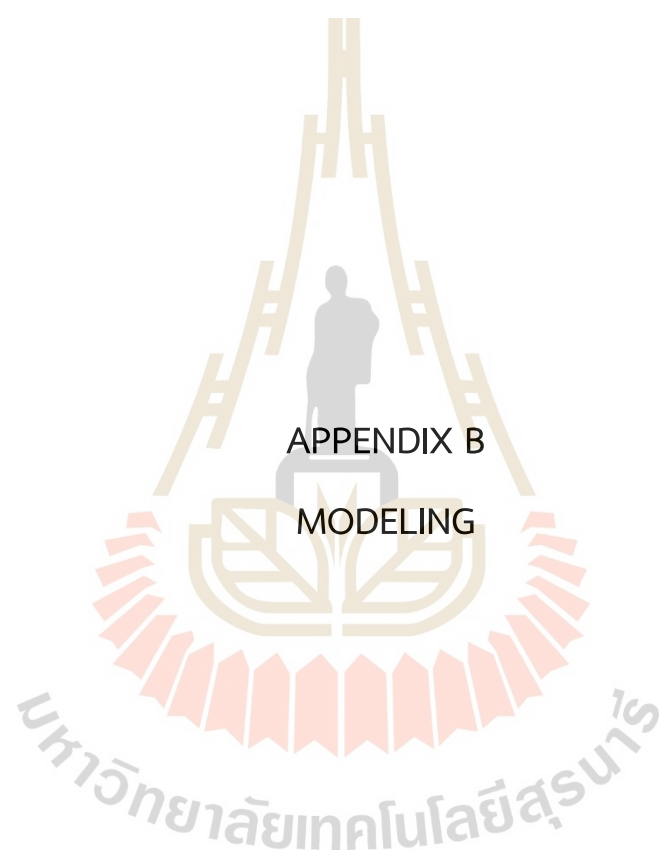
APPENDIX B

MODELING

Table B1. Hyperparameter selection and their ranges for each model.

Model	Hyperparameter	Range of Hyperparameter
SVM	C	[1,10, 50, 100]
	kernel	rbf, poly, sigmoid
	gamma	[1, 10, 100]
KNN	N_neighbors	range (1, 50)
	weights	uniform, distance
	P	manhattan, Euclidean
	Algorithm	auto, ball_tree, kd_tree, brute
RF	n_estimators	[50, 100, 200]
	max_depth	[None, 5, 10, 20]
	min_samples_split	[2, 5, 10]
	min_samples_leaf	[1, 2, 4]
NN	hidden_layer_sizes	[50, 100, 200]
	activation	logistic, tanh, relu
	solver	lbfgs, sgd, adam
	alpha	[0.0001, 0.001, 0.01]

Table B2. Optimal hyperparameter for each phase prediction in HEAs

Model	Phase	Optimal hyperparameter
SVM	BCC	C': 10, 'gamma': 1, 'kernel': 'rbf'
	FCC	C': 10, 'gamma': 1, 'kernel': 'poly'
	IM	C': 10, 'gamma': 1, 'kernel': 'rbf'
KNN	BCC	algorithm': 'auto', 'n_neighbors': 1, 'p': 1, 'weights': 'uniform'
	FCC	algorithm': 'auto', 'n_neighbors': 4, 'p': 1, 'weights': 'distance'
	IM	algorithm': 'auto', 'n_neighbors': 1, 'p': 2, 'weights': 'uniform'
RF	BCC	max_depth': 10, 'min_samples_leaf': 1, 'min_samples_split': 2, 'n_estimators': 50
	FCC	max_depth': 10, 'min_samples_leaf': 4, 'min_samples_split': 2, 'n_estimators': 50
	IM	max_depth': 10, 'min_samples_leaf': 1, 'min_samples_split': 2, 'n_estimators': 50
NN	BCC	activation': 'logistic', 'alpha': 0.0001, 'hidden_layer_sizes': (100,), 'solver': 'lbfgs'
	FCC	activation': 'tanh', 'alpha': 0.001, 'hidden_layer_sizes': (300,), 'solver': 'lbfgs'
	IM	activation': 'tanh', 'alpha': 0.01, 'hidden_layer_sizes': (300,), 'solver': 'lbfgs'

Table B3. Phase prediction results for the test dataset using different ML model. The table compares the experimentally observed phase with the predicted phase from each ML model.

Alloys	Experimental phase	Predicted phase			
		SVM	KNN	RF	NN
(CuMnNi)95Sn5	FCC+IM	BCC+FCC	FCC+IM	FCC	FCC
MoNbTaW	BCC	BCC	BCC	BCC	BCC+IM
CoCrFeNi	FCC	FCC	FCC	FCC	FCC
(Ti32.5V27.5Zr7.5Nb32.5)97Ni3	BCC+IM	BCC+IM	BCC+IM	BCC	BCC+IM
Al0.5CoCrCuFeNiV1.4	BCC	BCC	BCC	BCC+FCC	BCC
AlNb1.5Ta0.5Ti1.5Zr0.5	BCC	BCC	BCC	BCC	BCC
(Ti28Zr40Al20Nb12)97Si3	BCC+IM	BCC+IM	BCC+IM	BCC+IM	BCC+IM
CoCrFeMnNiV0.5	FCC	FCC	FCC	FCC	FCC
CoCuFeNiSn0.2	FCC+IM	FCC+IM	FCC+IM	FCC+IM	FCC+IM
CrNbTiZr	BCC+IM	BCC+IM	BCC+IM	BCC+IM	BCC+IM
Al0.5CoCrCuFeNiTi0.6	BCC+FCC	FCC	BCC+FCC	FCC	BCC+FCC
CoCrCuFeMnNiTiV	BCC+FCC+IM	BCC+FCC	BCC+FCC	BCC+FCC+IM	BCC+FCC+IM
CoFeMoNi1.6V	FCC+IM	FCC+IM	FCC+IM	FCC+IM	FCC+IM
Al1.8CoCrCuFeNi	BCC+FCC	BCC+FCC	BCC+FCC	BCC+FCC	BCC+FCC
AlC0.5CoCrFeNi	BCC+IM	BCC+IM	BCC+IM	BCC+IM	BCC+IM
CoCuFeNiSn0.1	FCC+IM	FCC+IM	FCC+IM	FCC+IM	FCC+IM
Al0.10CoCrFeMnNi	FCC	FCC	FCC	FCC	FCC
Al0.25NbTaTiV	BCC	BCC	BCC	BCC	BCC
CoCuFeMnNiSn0.20	FCC+IM	FCC	FCC+IM	FCC+IM	FCC+IM
Al0.5CoCrCuFeNiV2.0	BCC	BCC	BCC	BCC+FCC	BCC

Table B3. Phase prediction results for the test dataset using different ML model. The table compares the experimentally observed phase with the predicted phase from each ML model (continued).

Alloys	Experimental phase	Predicted phase			
		SVM	KNN	RF	NN
CoCrMnNi	FCC	FCC	FCC	FCC	FCC
Al22.5Cu20Fe15Ni20Ti22.5	FCC	FCC	FCC	FCC	FCC
Hf0.5Mo0.5NbSi0.3TiZr	BCC+IM	BCC	BCC	BCC+IM	BCC+IM
CrCuFeMoNi	FCC	BCC+FCC+IM	FCC+IM	FCC	FCC+IM
Al0.65CoCrFe2Ni	BCC+FCC	BCC+FCC	BCC+FCC	BCC+FCC	BCC+FCC
CoFeNiSi0.75	FCC+IM	FCC+IM	FCC+IM	FCC+IM	FCC+IM
CrFeNbTiMo	BCC+IM	BCC+IM	BCC	BCC+IM	BCC+IM
HfMoNbZrTi	BCC	BCC	BCC	BCC	BCC
AlCrCuFeNi0.8	BCC+FCC	BCC+FCC	BCC+FCC	BCC+FCC	BCC+FCC
MoNbTiV	BCC	BCC	BCC	BCC	BCC
HfNbTaTiZr	BCC	BCC	BCC	BCC	BCC
CoCrFeMnNiV1.0	FCC+IM	FCC	FCC	FCC	BCC+FCC
Al14Co41Cr15Fe10Ni20	FCC+IM	BCC+FCC	BCC+FCC	BCC+FCC	BCC+FCC
CoCrFeNb0.412Ni	FCC+IM	FCC+IM	FCC+IM	FCC	FCC+IM
AlMoNbTiV	BCC	BCC	BCC	BCC	BCC
Al0.81CoCrFeMnNi	BCC+FCC	BCC+FCC	BCC+FCC	BCC+FCC	BCC+FCC
AlC0.4CoCrFeNi	BCC+IM	BCC+IM	BCC+IM	BCC+IM	BCC+IM
AlCoCrFeNiSi0.4	BCC	BCC	BCC	BCC	BCC
CoCrFeMnNiV0.75	FCC+IM	FCC	FCC	FCC	FCC
Al0.5CoCrFeNiTi0.5	BCC+FCC	BCC	BCC+IM	BCC+IM	BCC+IM

Table B3. Phase prediction results for the test dataset using different ML model. The table compares the experimentally observed phase with the predicted phase from each ML model (continued).

Alloys	Experimental phase	Predicted phase			
		SVM	KNN	RF	NN
AlCrCuFeNi0.6	BCC+FCC	BCC+FCC	BCC+FCC	BCC+FCC	BCC+FCC
(AlCoCrFeNi)99Zr1	BCC+IM	BCC	BCC+IM	BCC	BCC+IM
Mo1.3NbTiVZr	BCC	BCC	BCC	BCC	BCC
CoCrCu0.5FeNi	FCC	BCC+FCC	FCC	FCC	FCC
AlCrCuFeNi2	BCC+FCC	BCC+FCC	BCC+FCC	FCC	BCC+FCC
AlCoCrFe0.5Mo0.5Ni	BCC+IM	BCC+IM	BCC+IM	BCC+IM	BCC+IM
Ti1.6ZrNbAl0.4	BCC+IM	BCC+IM	BCC+IM	BCC+IM	BCC+IM
MoNbTiV0.50Zr	BCC	BCC	BCC	BCC	BCC
CoCrFeNb0.155Ni	FCC+IM	FCC+IM	FCC+IM	FCC+IM	FCC+IM
Al0.5HfNbTaTiZr	BCC	BCC	BCC	BCC	BCC
Ti1.6ZrNbAl0.6	BCC+IM	BCC+IM	BCC+IM	BCC+IM	BCC+IM
NbTiV2Zr	BCC	BCC	BCC	BCC	BCC
Cr15Cu5Fe20Mn25Ni35	FCC	FCC	FCC	FCC	FCC
Al3CoCrFeNi	BCC	BCC+IM	BCC+IM	BCC+IM	BCC+IM
CoFeMoNiV	FCC+IM	FCC+IM	FCC+IM	FCC+IM	FCC+IM
Cr26Mn20Fe20Co20Ni14	FCC	FCC	FCC	FCC	FCC
Al0.5CoCrCuFeNiV1.0	BCC+FCC+IM	BCC+FCC+IM	BCC+FCC+IM	BCC+FCC+IM	BCC+FCC+IM
MoNbTiV1.0Zr	BCC	BCC	BCC	BCC	BCC
Al40(CoCrCuFeMnNiTiV)60	BCC+IM	BCC	BCC+FCC+IM	BCC	BCC
Al0.5CoCrCuFeNiV1.2	BCC	BCC+FCC	BCC	BCC+FCC+IM	BCC+IM

STable B3. Phase prediction results for the test dataset using different ML model. The table compares the experimentally observed phase with the predicted phase from each ML model (continued).

Alloys	Experimental phase	Predicted phase			
		SVM	KNN	RF	NN
Al0.5CoCrCuFeNi	FCC	FCC	FCC	FCC	FCC
NbTiV0.3Mo0.3Zr	BCC	BCC	BCC	BCC	BCC
CoCrFeNb0.309Ni	FCC+IM	FCC+IM	FCC+IM	FCC	FCC+IM
Ti1.6ZrNbAl0.5	BCC+IM	BCC+IM	BCC+IM	BCC+IM	BCC+IM
MoNbTaVW	BCC	BCC	BCC	BCC	BCC
MoNbTiVZr	BCC	BCC	BCC	BCC	BCC
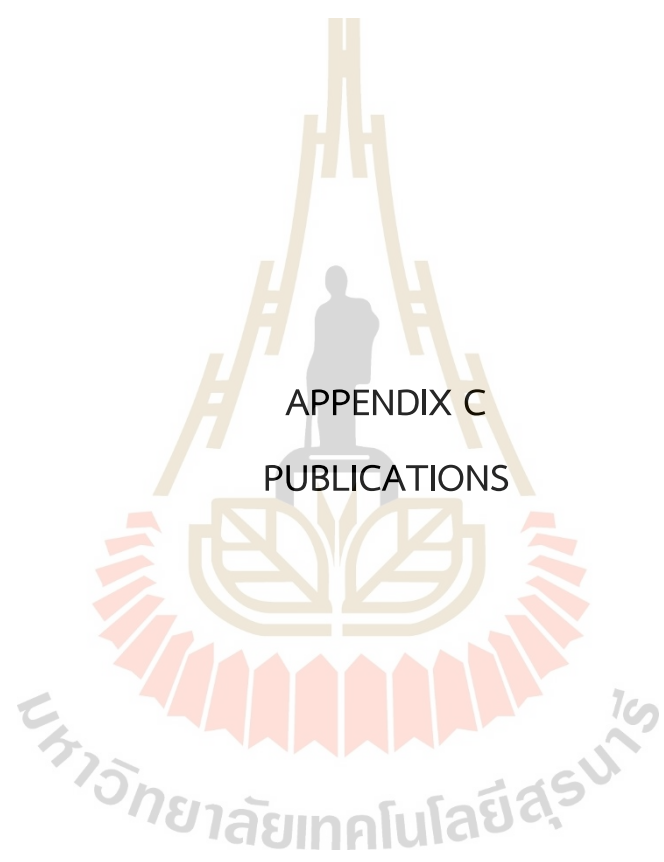
APPENDIX C

PUBLICATIONS

List of Publication

INTERNATIONAL JOURNAL PAPER

Thampiriyanon, J., & Khumkoa, S. (2025). Machine Learning–Based Prediction of Complex Combination Phases in High-Entropy Alloys. *Metals*, *15*(3), 227. https://doi.org/10.3390/met15030227

CONFERENCE PAPER

Thampiriyanon, J., & Khumkoa, S. (2022). Mechanical properties prediction of austenitic stainless steel using machine learning, The 13th Thailand Metallurgy Conference.

INTERNATIONAL CONFERENCE PAPER

Thampiriyanon, J., Laungsakulthai, K., Laokhen, P., Thongnak, S., & Khumkoa, S. (2020, September). A Feasibility Study on Metallurgical Slag Classification by Microstructure Recognition. In *Materials Science Forum* (Vol. 1009, pp. 107-113). Trans Tech Publications Ltd.

metals

MDPI

Article

Machine Learning–Based Prediction of Complex Combination Phases in High-Entropy Alloys

Jirapracha Thampiriyanon and Sakhob Khumkoa *

School of Metallurgical Engineering, Institute of Engineering, Suranaree University of Technology, Nakhon Ratchasima 30000, Thailand; jirapracha@gmail.com

* Correspondence: sakhob@sut.ac.th

Abstract: High-entropy alloys (HEAs) are a novel class of materials that exhibit exceptional mechanical, thermal and corrosion-resistant properties, making them highly promising for various industrial applications. The complex and variable compositions of HEAs present significant challenges in accurately predicting their phase structures, which is crucial for tailoring their properties for specific applications. This study proposes a machine learning (ML)–based approach to predict HEA phase structures using experimentally validated datasets containing chemical composition–derived features. A Boolean vector technique was employed to encode multiphase classifications, allowing the model to accurately capture complex phase relationships. Four ML algorithms consisting of support vector machine (SVM), k-nearest neighbors (KNN), random forest (RF) and neural network (NN) were utilized to develop predictive models. The models were evaluated using rigorous cross-validation (CV) and tested on unseen data samples. The results demonstrate that NN and KNN outperform other models, achieving an impressive test accuracy of 84.85%. Feature importance analysis reveals that valence election concentration and melting temperature are the most influential factors governing phase formation. This study highlights the effectiveness of ML in accelerating HEA design by significantly reducing reliance on experimental trial-and-error approaches. The predictive models can be applied in alloy design to rapidly screen compositions with desired phase structures, guiding experimental synthesis and optimizing material properties for advanced application. The findings underscore the potential of ML in computational materials science and open new avenues for integration artificial intelligence–driven approaches with conventional metallurgical techniques to enhance HEA development.

Keywords: high-entropy alloys; machine learning; phase prediction

1. Introduction

High-entropy alloys (HEAs) are a type of metallic alloy that usually contains five or more principal elements in approximately equimolar proportions, as opposed to conventional alloys that typically consist of one or two dominant elements. These alloys are also known as complex metallic alloys or multi-principal element alloys (MPEAs) [1]. The combination of multiple elements in near-equiatomic ratios results in unique and superior properties [2] including enhanced corrosion resistance [3,4], high wear resistance [5,6], excellent fatigue resistance [7] and exceptional mechanical strengths [8] that make HEAs the material of choice for a wide range of applications. The burgeoning interest in HEAs is driven by their remarkable attributes that offer new ways to develop materials with tailored characteristics. The discovery and development of HEAs have opened up exciting possibilities for creating novel materials with customized characteristics and have the potential

to revolutionize various industries in the future. The phase structure of HEAs includes body-centered cubic (BCC), face-centered cubic (FCC) and intermetallic (IM) phases. These phases critically influence the material properties. BCC structures are synonymous with high strength and hardness, while FCC structures enhance ductility and formability. IM phases arising from specific element combinations can enhance or detract from the alloy's mechanical properties depending on their nature and distribution within the material [9]. Due to the complexity of HEAs's phase structures, advanced methods are required for phase prediction. Common approaches for predicting phases in HEAs have traditionally relied on trial-and-error experimentation. While this method provides essential validation and real-world data, it is often time-consuming and resource-intensive. Consequently, it is challenging to thoroughly explore the extensive compositional space of HEAs using this approach alone. Various empirical design approaches [10,11] have been used to identify the phase formation in solid solution (SS), IM and amorphous (AM) phases. These approaches include parameters such as entropy of mixing (ΔS_{mix}), enthalpy of mixing (ΔH_{mix}), melting temperature (T_m), atomic size difference (δ), valence electron concentration (VEC) and the parameter for predicting solid solution formation (Ω). Additionally, the VEC and electronegativity difference ($\Delta\chi$) has been proposed as an empirical rule to differentiate between BCC and FCC structures of SS phases [12]. While these models can indicate certain phase formation trends, their prediction performance is often unreliable, making them not very robust. The Calculation of Phase Diagrams (CALPHAD) [13] and density functional theory (DFT) [14] are efficient computational methods for investigating the thermodynamics and phase stability of HEAs. However, CALPHAD relies on static databases and approximations, which may not fully capture the complexities of systems with higher-order elements [3,15]. On the other hand, DFT does not depend on pre-existing databases; instead, it computes material properties from first principles. While this allows DFT to avoid the limitations of static data, it demands significant computational resources, making it impractical for modeling disordered or compositionally complex systems such as HEAs [16]. In contrast, machine learning (ML) offers innovative and efficient approaches for predictive modeling in materials science, enabling autonomous microstructure interpretation steels and special alloys [17], phase and mechanical property prediction in entropy alloys [18] and accelerated analysis of time–temperature–transformation diagrams in stainless steels [19]. These advancements significantly enhance the speed and accuracy of material discovery and design. This method enables the rapid screening of vast compositional space, predicting material properties based on large datasets. ML has the potential to streamline the discovery and optimization of HEAS by leveraging large, diverse datasets to identify complex patterns and make accurate predictions. Unlike CALPHAD, which relies on static databases with limited flexibility, ML dynamically learns from diverse datasets and can generalize to new compositions, enabling it to handle the complexities of HEAs more effectively. This approach is particularly advantageous given the growing repository of experimental HEA data enabling the development of predictive models that can navigate the complex interplay of compositional and processing parameters to accurately predict phase outcomes.

Several notable recent studies on the application of ML for HEA phase prediction have showcased the efficacy of various algorithms [20–22]. The corresponding phase data from experimental observations and some features derived from Gibbs free energy and Hume-Rothery parameters serve as valuable training data for ML algorithms. Islam et al. [23] utilized a neural network (NN) incorporating five input features to categorize phase selection in HEAs as SS, IM or AM. They achieved an average cross-validation accuracy exceeding 80%. Huang et al. [24] proposed ML to effectively investigate phase selection rules using a comprehensive experimental dataset comprising 401 samples. Three diverse

ML algorithms—k-nearest neighbors (KNN), support vector machine (SVM) and neural network (NN)—were selected for classifying phases into SS, IM and multiphase SS + IM. The NN achieved the highest testing accuracy of 74.3%. Dai et al. [25] applied six different algorithms consisting of SVM, Ada boost (AB), decision tree (DT), RF, gradient boosting (GB) and logistic regression (LR). They selected 9 initial features and created 36 additional features through dimensionality augmentation. Two distinct feature selection techniques, namely Least Absolute Shrinkage and Selection Operator (LASSO) and Recursive Feature Elimination (RFE), were applied to classify phases as FCC, BCC, hexagonal close packed (HCP), multiphase (MP) and AM. The findings highlighted that feature engineering led to enhanced predictive accuracy in phase identification compared to conventional methods. Zhang et al. [26] employed NN, SVM and GB and optimized features with feature selection and feature variable transformation based on kernel principal component analysis (KPCA) to classify phases into SS, IM and SS + IM. The accuracy of the testing set predicted by the SVM was 97.43%. Machaka et al. [27] trained DT, linear discriminant analysis (LDA), naïve Bayes (NB), generalized linear regression (GLMNET), RF, NN, KNN and SVM to classify solid solution phases into BCC, FCC and BCC + FCC. A total of 36 metallurgy specific features were reduced to 13 features by using feature selection. The RF outperformed the other algorithms with an accuracy rate of 97.5%. Syarif et al. [28] implemented NN to discover the set of element phase formation drivers that can stabilize or destabilize the phase formation of BCC, FCC and IM in HEAs based on the concentration of the alloy constituent element. Nia et al. [29] proposed KNN with an HEA interaction network to categorize FCC, BCC, HCP, MP and AM structures. The results show that the accuracy of the proposed algorithm was 88.88%. Gao et al. [30] employed four ensemble models including RF, XGboost, Voting and Stacking to identify the phases of BCC, FCC and BCC + FCC. Among these algorithms, Voting and Stacking stand out with a predictive accuracy of over 92%. He et al. [31] distinguished BCC, FCC, BCC + FCC and AM phases using five ML algorithms. RF showed the best performance of the tested algorithms, with an accuracy of 87%. ΔH_{mix} and VEC parameters play an important role in prediction. The experimental results validated that the phase structure of CoCrFeNiAlx alloys with the increase in Al content is consistent with that obtained by ML prediction. These recent studies highlight ML's capability not only to achieve high predictive accuracies but also to uncover the underlying phase selection rules that govern the formation of specific phases in HEAs.

Previous research has primarily focused on distinguishing FCC, BCC and FCC + BCC phases in SS [27,30,31] or identifying SS, IM, SS + IM or AM phases [23,24,26]. However, no study has accurately identified complex combinations such as BCC + IM, FCC + IM or BCC + FCC + IM. The current study aims to address this gap by accurately identifying these specific phase combinations in HEAs through the development of a comprehensive model capable of classifying HEAs into a broader spectrum of six distinct phase categories including BCC, FCC, BCC + FCC, BCC + IM, FCC + IM and BCC + FCC + IM. This endeavor not only addresses the intricate challenge posed by multiphase and complex phase structures in HEAs but also makes the following key contributions: (1) introduction of a novel Boolean vector encoding technique to effectively represent complex phase combinations, (2) demonstrating that NN and KNN outperform other models with a maximum accuracy of 84.85% for complex phase predictions and (3) identification of critical features such as VEC and T_m as major determinants in HEA phase prediction. These contributions collectively enhance the granularity and accuracy of phase prediction, significantly advancing the design and development of HEAs with tailored properties for innovative and cutting-edge applications.

2. Materials and Methods

Figure 1 illustrates the comprehensive strategy employed in this study to predict the phase structures of HEAs using an ML approach. This strategy encompasses the typical steps of an ML process as applied in materials science [32]. The ML models were implemented using scikit-learn (version 1.2.0, Python Software Foundation, Wilmington, DE, USA). The phase structures considered in this study are categorized into six groups: BCC, FCC, BCC + FCC, BCC + IM, FCC + IM and BCC + FCC + IM.
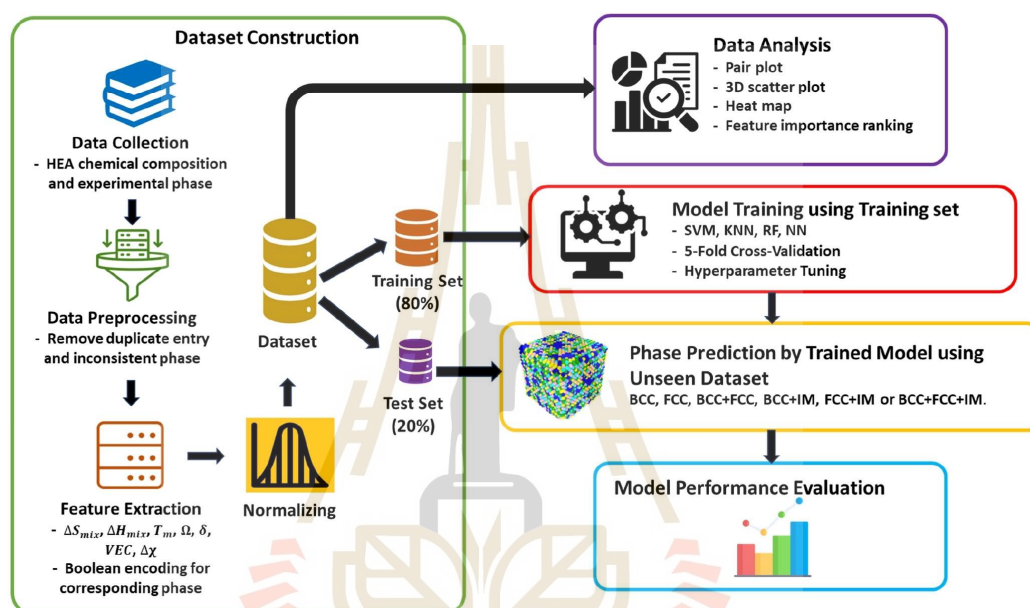
Figure 1. The comprehensive strategy for predicting the phases of high-entropy alloys in this work.

2.1. Dataset Construction

2.1.1. Data Collection and Feature Extraction

The data for this study were sourced from the existing literature [33–64]. The dataset underwent a manual refinement process that included removing duplicate entries, addressing missing and inconsistent phase data and limiting the scope to alloys synthesized via arc melting. Alloys produced by different synthesis methods such as powder metallurgy, mechanical alloying, laser melting deposition and additive manufacturing were excluded to avoid the influence of production techniques on the results [9]. After this refinement process, 329 entries remained including 96 BCC, 55 FCC, 52 BCC + FCC, 67 BCC + IM, 46 FCC + IM and 13 BCC + FCC + IM phases.

The ML models in this study utilize input features derived from chemical compositions of HEAs, selected based on their established correlation with phase stability. The dataset comprises numerical features extracted from the chemical composition of each alloy and their corresponding phase, calculated using the equations presented in Table 1. The seven input features include ΔS_{mix} (mixing entropy), affecting the tendency to form solid solutions; ΔH_{mix} (mixing enthalpy), which governs thermodynamic stability; T_m (melting temperature), which influences phase stability under thermal conditions; (parameter for predicting solid solution formation); (atomic size difference), impacting lattice distortions; (valence electron concentration), which differentiates BCC and FCC phases; and (elec-

tronegativity difference), affecting phase separation and intermetallic formation. Each HEA composition is represented as a numerical vector containing these seven features, and the model predicts the phase classification into one of six categories: BCC (body-centered cubic), FCC (face-centered cubic), BCC + FCC (dual-phase BCC and FCC), BCC + IM (BCC with intermetallic phases), FCC + IM (FCC with intermetallic phases) and BCC + FCC + IM (a combination of all three phases). Given the inherent imbalance in the dataset, a Boolean vector approach was introduced to improve predictive accuracy [65]. Instead of treating phase combinations as distinct categorical labels, each phase was encoded separately using a binary representation, allowing the model to learn the presence or absence of each phase independently. The target variable consists of three binary values corresponding to BCC, FCC and IM, where 1 indicates the presence and 0 indicates the absence of particular phase. For instance, a sample with a BCC phase is represented as [1, 0, 0], a BCC + FCC phase as [1, 1, 0] and a complex BCC + FCC + IM phase as [1, 1, 1]. This encoding strategy ensures that the model effectively captures phase interactions without being constrained by the uneven distribution of specific phase combination. Table 2 provides an example of input feature values and the corresponding output phase classification using this Boolean encoding scheme.

Table 1. Equations for feature extraction from chemical composition of alloys.

Equation	Feature Description	Unit	Reference
$\Delta S_{mix} = -R \sum_{i=1}^{n} c_i \ln c_i$	Mixing entropy	J/(mol·K)	[66]
$\Delta H_{mix} = \sum_{i=1,\, i \neq j}^{n} 4\Delta H_{ij} c_i c_j$	Mixing enthalpy	kJ/mol	[26]
$T_m = \sum_{i=1}^{n} c_i T_{mi}$	Melting temperature	K	[10]
$\Omega = \frac{T_m \Delta S_{mix}}{|\Delta H_{mix}|}$	Parameter for predicting solid solution formation	-	[10]
$\delta = \sqrt{\sum_{i=1}^{n} c_i \left(1 - \frac{r_i}{\bar{r}}\right)}$	Atom size difference	%	[66]
$VEC = \sum_{i=1}^{n} c_i VEC_i$	Valence electron concentration	-	[12]
$\Delta \chi = \sqrt{\sum_{i=1}^{n} c_i (\chi_i - \bar{\chi})}$	Electronegativity difference	-	[67]

Table 2. Example of input features and corresponding phase classifications of high-entropy alloys.

Alloys	Experimental Phase	ΔH_{mix}	ΔS_{mix}	δ	VEC	$\Delta \chi$	T_m	Ω	BCC	FCC	IM
HfMoNbTaTiZr	BCC	−0.89	14.90	4.70	4.67	0.28	2584.67	43.32	1	0	0
CoCrFeMnNiV0.5	FCC	−6.21	14.70	2.18	7.73	0.14	1823.91	4.31	0	1	0
CoCrFeMnNiV0.75	FCC + IM	−6.96	14.85	2.48	7.61	0.14	1838.65	3.93	0	1	1
AlCrCuFeNi0.8	BCC + FCC	−3.40	13.35	5.01	7.50	0.12	1585.19	6.22	1	1	0
AlCoCuFeNiZr	BCC + FCC + IM	−23.89	14.90	9.59	7.50	0.21	1620.02	1.01	1	1	1

Footnote: In these equations, R represents the gas constant; c_i and c_j are atomic concentrations of the ith and jth components, respectively; ΔH_{ij} is an interaction parameter between the ith and jth elements; T_{mi} is the melting temperature of the ith component; r_i is the atomic radius of the ith component; $\bar{r}$ is the average atomic radius; VEC_i is the valence

electron concentration of the *i*th component; χ_i is the Pauling electronegativity of the *i*th component; and $\overline{\chi}$ is the average Pauling electronegativity.

2.1.2. Data Normalization

Preprocessing data in ML is an essential step to enhance data quality and ensure their suitability for ML algorithms. One common technique is normalization, which standardizes the range of feature data, ensuring that each feature contributes approximately equally to the final prediction as described by the following formula [68]:

$$x_{new} = \frac{x_i - \overline{x}}{\sigma} \tag{1}$$

where x_{new} is the standardized value of the feature, x_i is the original value of the feature, x_i is the mean value of the feature and σ is the standard deviation of the feature. The purpose of normalization is to produce dimensionless numerical features so that each data point is on the same numerical scale. This process is crucial for improving the accuracy of the ML model by ensuring that no single feature dominates the learning process due to its scale.

2.2. ML Algorithms

ML algorithms have become a power tool for predicting complex material properties including the phase structure of HEAs. By leveraging large datasets and sophisticated algorithms, researchers can identify patterns and relationships that traditional methods may overlook. In this study, four ML algorithms were employed to predict the phases of HEAs: SVM, KNN, RF and NN. These algorithms were implemented using the scikit-learn library and were selected based on their frequent application in similar studies and their demonstrated effectiveness in handling classification tasks in material science. Each algorithm offers unique strengths: SVM excels in distinguishing nonlinear relationships, KNN is simple yet effective for pattern recognition, RF is robust against overfitting and interpretable due to feature importance analysis and NN is well-suited for capturing complex, nonlinear interactions within the dataset.

- Support Vector Machine (SVM)

The SVM aims to find the hyperplane that best separates different classes of data points. This separation is achieved by minimizing the loss function, which involves finding the hyperplane that maximizes the margin between the classes. It can handle nonlinear boundaries by using the kernel function [68,69], which transforms the data into a higher dimensional space where separation might be possible. The key hyperparameters used for the SVM in this study are penalty (C), the kernel function and the kernel coefficient (gamma). When C is too large, the model may overfit the training data, leading to poor generalization to new data. When C is too small, the model might not capture outliers well resulting in high training set errors and limiting to fit the training data. Different kernel functions serve different purposes and can capture various types of relationship in the data. The radial basis function (RBF), polynomials and sigmoid kernels are kernel functions used for the SVM in this study. Gamma affects the sensitivity to differences in feature vectors. If gamma is set too high, the model might overfit as it would become sensitive to individual data points. If gamma is chosen too low, the model may not capture the underlying patterns in the data.

- K-Nearest Neighbor (KNN)

The KNN makes predictions based on the minimum distance between data points and a simple majority vote from the nearest neighbors. Its hyperparameters are the number of neighbors to consider for prediction, the weight function, the algorithm for computing the

nearest neighbors and a distance metric [29,70]. Smaller values of nearest neighbors can make the model sensitive to noise, while larger values can make it biased. For this study, two common weight function options are uniform (equal weights) and distance (inverse of distance as weight). The algorithms used for efficiently computing the nearest neighbors are Auto, ball tree, KD tree and brute force. Euclidean and Manhattan are used to measure the distance between data points.

- Random Forest (RF)

RF is an ensemble technique that combines multiple decision trees to make predictions. Each tree relies on values from a random vector sampled independently and from the same distribution for all trees. The majority predicted class across trees become the final predicted class. The hyperparameters for tuning this model include the number of trees, the maximum feature, the maximum tree depth, the minimum sample split, the minimum sample leaf and bootstrap sampling. The number of trees determine the decision tree in the forest. Increasing the number of trees can improve model performance, but it also increases computation time. The maximum feature specifies the maximum number of features considered for splitting a node in each decision tree. This parameter can impact the diversity and randomness of the trees, reducing overfitting. The maximum tree depth defines the maximum number of levels a decision tree can have. Restricting tree depth can help prevent overfitting and enhance generalization. The minimum sample split sets the minimum number of data points required in a node before it is eligible for further splitting. The minimum sample leaf specifies the minimum number of data points that a leaf node must contain. If bootstrap sampling is set to true, sampling is performed with replacements when constructing individual decision trees [68].

- Neural Network (NN)

The NN is a type of feedforward neural network with multiple layers including an input layer, one or more hidden layers and an output layer. The key hyperparameters associated with training the NN are the number of hidden layers, the size, activation functions, the loss function, the optimizer and the learning rate. The number of hidden layers and the number of neurons in each hidden layer are crucial parameters that define the depth and width of the neural network. More layers and neurons allow the network to capture more complex patterns but also increase the risk of overfitting and the computational complexity [71]. Activation functions like rectified linear unit (ReLU), sigmoid or tanh are used to introduce nonlinearities in the model, allowing it to learn more complex relationships. Different activation functions have different properties and are suitable for different types of data. The loss function is a measure of how well the neural network is performing. The optimizer is an algorithm that adjusts the weights of the network to minimize the loss. Common optimizers include stochastic gradient descent (SGD) and adaptive moment estimation (Adam). The learning rate is a hyperparameter that controls how to change the model in response to the estimated error when the model weights are updated. A small value can make the training process very slow while a large value can cause the model to converge too quickly to the optimal solution or even diverge.

The dataset was randomly divided into two subsets at an 80:20 ratio for training and testing. This means that 80% of the data were used to train the models while the remaining 20% were reserved for testing the model performance. This split helps ensure that the models are trained on a comprehensive portion of the data while also providing an unbiased evaluation of their performance on unseen data. Grid search is a hyperparameter tuning technique used in ML to systematically search for the optimal combination of hyperparameters that yield the best model performance. The hyperparameter search space for each model was defined based on prior literature, empirical validation, preliminary

model performance assessments and iterative tuning based on a validation set. Table 3 summarizes the hyperparameter options and their ranges for the SVM, KNN, RF and NN used for the phase prediction of HEAs in this study. Each hyperparameter is crucial for tuning each model to achieve optimal performance. The evaluation strategy used was five-fold cross validation (CV), where the dataset was randomly split into equal parts. In each iteration, four parts were used for training, while one was reserved for validation, ensuring robust generalization and reducing overfitting risks. The optimal hyperparameter were selected based on the highest average CV accuracy [72].

Table 3. Hyperparameter selection and their ranges for the different models in this study.

Model	Hyperparameter	Range of Hyperparameter
SVM	C	[1, 10, 50, 100]
	kernel	rbf, poly, sigmoid
	gamma	[1, 10, 100]
KNN	N_neighbors	range (1, 50)
	weights	uniform, distance
	P	manhattan, Euclidean
	Algorithm	auto, ball_tree, kd_tree, brute
RF	n_estimators	[50, 100, 200]
	max_depth	[None, 5, 10, 20]
	min_samples_split	[2, 5, 10]
	min_samples_leaf	[1, 2, 4]
NN	hidden_layer_sizes	[50, 100, 200]
	activation	logistic, tanh, relu
	solver	lbfgs, sgd, adam
	alpha	[0.0001, 0.001, 0.01]

2.3. Evaluation of Model Performance

During model training, five-fold CV was applied to assess the mean CV accuracy of each model, providing an initial measure of model reliability and generalizability. For testing with an unseen dataset, multiple evaluation metrics were calculated to provide a comprehensive performance assessment. These metrics included accuracy, precision, recall and F1-score. The accuracy score measures the ratio of correctly predicted instances to total instances. The precision score calculates the proportion of true positives among all predicted positives. The recall score assesses the model's ability to identify all true positives. F1-score combines precision and recall to find their harmonic mean for a balanced assessment [73]. These metrics are defined as follows [68]:

$$Accuracy\ (\%) = \frac{TN + TP}{TP + TN + FN + FP} \times 100\% \qquad (2)$$

$$Precision\ (\%) = \frac{TP}{FP + TP} \times 100\% \qquad (3)$$

$$Recall\ (\%) = \frac{TP}{FN + TP} \times 100\% \qquad (4)$$

where a true negative (TN) is an instance that is actually negative and correctly predicted as negative, a true positive (TP) is an instance that is actually positive and correctly predicted as positive, a false positive (FP) is an instance that is actually negative but incorrectly

predicted as positive and a false negative (*FN*) is an instance that is actually positive but incorrectly predicted as negative. Additionally, the receiver operating characteristic (ROC) curve and the precision–recall (PR) curve were analyzed to provide a better understanding of model performance. The ROC curve provides insight into the trade-off between the true positive and false positive rate, while the PR curve highlights the balance between precision and recall, which is especially valuable for assessing models on imbalanced datasets. The evaluation metrics used in this study are essential for assessing model performance in predicting HEA phases. From a material science perspective, high accuracy, precision and recall in phase prediction models indicate reliable and efficient classification of HEA phases. High accuracy ensures that the predicted phase compositions closely math experimental observations, while high precision signifies that the model rarely misclassifies phases, making it valuable for applications where incorrect phase classification could negatively impact mechanical strength, corrosion resistance or thermal stability. High recall ensures that all relevant phases are correctly identified reducing the risk of missing critical phase formations in complex multiphase HEAs. These high-performance metrics have significant metallurgical implications, including enhanced reliability in alloy design, reduced experimental costs, improved process optimization and better prediction of final material properties.

The complex combination phase prediction accuracy offers a rigorous evaluation standard for applications requiring phase identification in HEAs. This metric, alongside the other performance indicators, provides a comprehensive overview of each model's suitability for predicting HEA phase compositions in both individual and multiphase scenarios.

3. Results and Discussion

3.1. Data Analysis

The pair plot generated using the Seaborn library (version 0.12.2, MWaskom, San Francisco, CA, USA), as depicted in Figure 2, showcases the relationships between pairs of features. Each scatter plot point corresponds to a specific phase in the dataset, with its position representing the values of the two features for that phase. The diagonal cells contain kernel density plots illustrating the distribution of each individual features. The pair plot reveals that certain feature pairs, such as VEC and ΔH_{mix}, exhibit noticeable trends, where specific phase categories cluster within distinct regions. For instance, FCC phases tend to align with higher VEC values, while BCC phases correlate with lower VEC values and more negative ΔH_{mix}. These relationships indicate that some features play a critical role in determining phase stability and composition, offering valuable insights for phase prediction.

The radar plot in Figure 3 provides a comparative visualization of the normalized feature values for each phase category. The plot highlights distinct patterns across the phases, with BCC phases exhibiting lower δ values and higher T_m while FCC phases show relatively higher δ values. IM phases, on the other hand, demonstrate a unique combination of high δ and more negative ΔH_{mix}, reflecting their ordered structures and complex bonding characteristics. These differences emphasize the significance of these features in distinguishing between phase categories.
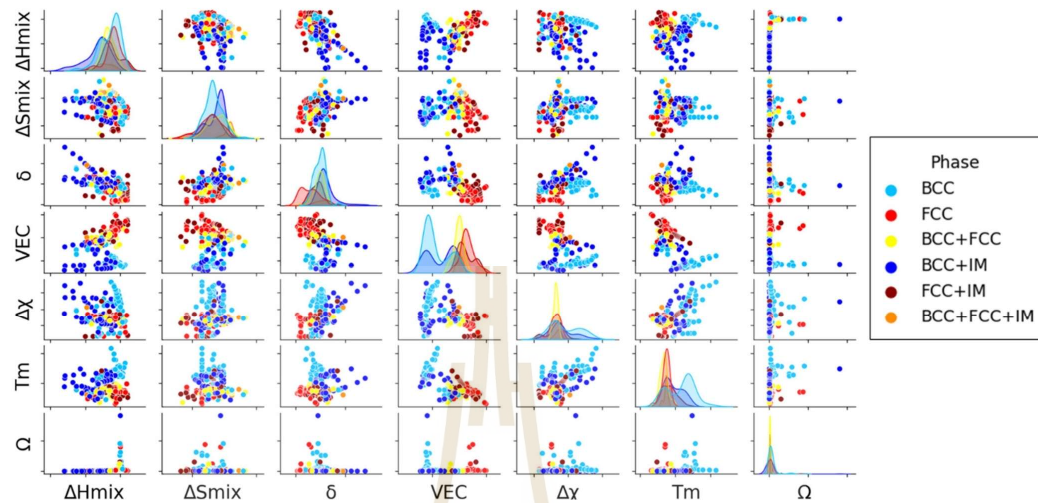
Figure 2. Pair plot of features for phase categories. Abbreviations: entropy of mixing (ΔS_{mix}), enthalpy of mixing (ΔH_{mix}), melting temperature (T_m), atomic size difference (δ), valence electron concentration (VEC), parameter for predicting solid solution formation (Ω), electronegativity difference ($\Delta\chi$), BCC (body-centered cubic), FCC (face-centered cubic), IM (intermetallic).
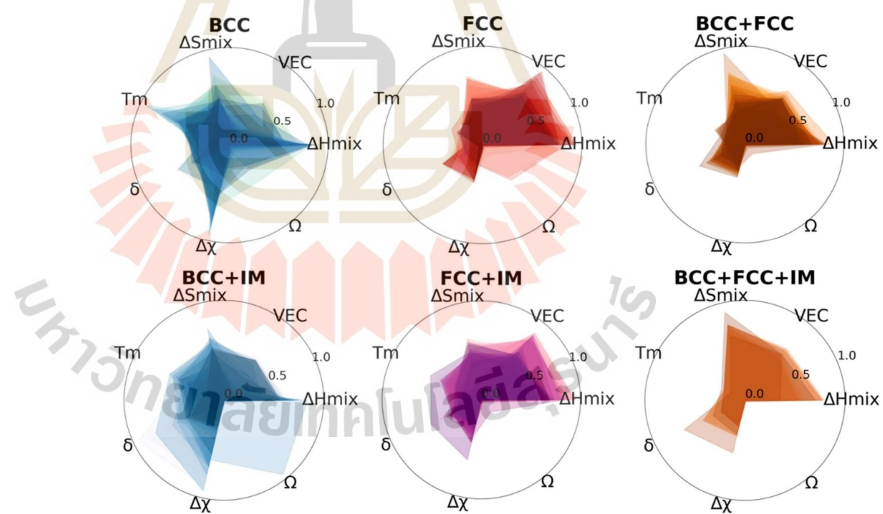
Figure 3. Radar plot of features for phase categories of dataset. Abbreviations: entropy of mixing (ΔS_{mix}), enthalpy of mixing (ΔH_{mix}), melting temperature (T_m), atomic size difference (δ), valence electron concentration (VEC), parameter for predicting solid solution formation (Ω), electronegativity difference ($\Delta\chi$), BCC (body-centered cubic), FCC (face-centered cubic), IM (intermetallic).

Figure 4 shows a heat map illustrating the relationships or correlations between features in the dataset. Each cell in the matrix represents the correlation coefficient between two features, ranging from −1 to 1. A value of 1 indicates a strong positive correlation, −1 indicates a strong negative correlation and 0 indicates no correlation between features. The values in the heatmap are correlation coefficients that measure the strength and direc-

tion of the linear relationship between two features. The most commonly used correlation coefficient is the Pearson correlation coefficient calculated as follows [68]:

$$r_{xy} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}} \tag{5}$$

where r_{xy} is the Pearson correlation coefficient between feature x and y; x_i and y_i are the individual sample points indexed with i; and $\bar{x}$ and $\bar{y}$ are the means of the x and y features, respectively.


Figure 4. Correlation heat map of feature relationships in dataset. Abbreviations: entropy of mixing (ΔS_{mix}), enthalpy of mixing (ΔH_{mix}), melting temperature (T_m), atomic size difference (δ), valence electron concentration (VEC), parameter for predicting solid solution formation (Ω), electronegativity difference ($\Delta \chi$).

The correlation values in the heat map of two independent features range from -0.75 to 0.60. Specifically, VEC and T_m exhibit a negative correlation implying that as VEC increases, T_m tends to decrease. Higher VEC values favor FCC structures, whereas lower VEC values stabilize BCC structures, since FCC phases often have lower T_m values. Conversely, the electronegativity difference and T_m show a positive correlation, meaning that as $\Delta \chi$ increases, so does the melting temperature. Larger $\Delta \chi$ values can promote the formation of complex phases or IM phases, which often have higher T_m than SS phases [74]. Overall, there are no strong positive or negative correlations between the features. Thus, all the features can be retained.

The 3D scatter plot was created to illustrate the phase distribution across three axes: δ, ΔH_{mix} and VEC. The relationship between these features and the corresponding phases is displayed in Figure 5. From the plot, it is apparent that IM phases have higher δ and lower ΔH_{mix} values compared to solid solution phases (BCC and FCC). IM phases often involve the bonding of atoms with significantly different atomic sizes, creating lattice distortions and stains within the crystal structure leading to distinctive and ordered structures. Gibbs free energy (ΔG) determines the thermodynamic stability of phases. A lower ΔH_{mix} results in a lower ΔG, which can favor the formation of intermetallic compounds. When examining solid solution phases, a higher VEC tends to result in the formation of FCC phases, while a lower VEC tends to lead to the formation of BCC phases. Lower δ values

generally encourage the formation of SS phases, whereas significant atomic size disparities are more likely to result in the formation of IM phases. The results are consistent with earlier research [10,12,26].
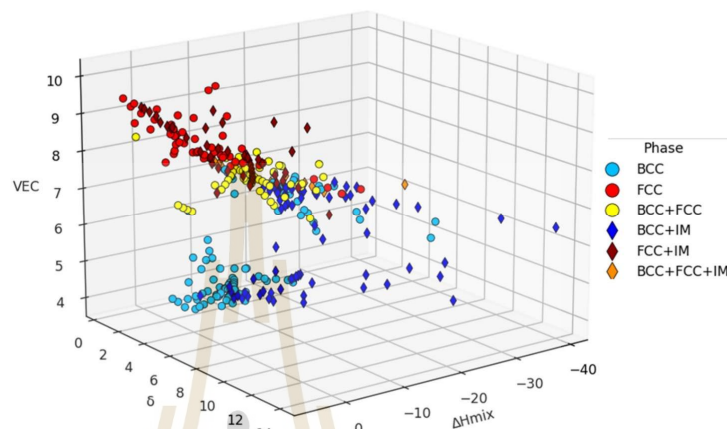
Figure 5. The 3D scatter plot showing phase distribution to compare the effect of δ, ΔH_{mix} and VEC. Abbreviations: entropy of mixing (ΔS_{mix}), enthalpy of mixing (ΔH_{mix}), melting temperature (T_m), atomic size difference (δ), valence electron concentration (VEC), parameter for predicting solid solution formation (Ω), electronegativity difference ($\Delta \chi$), BCC (body-centered cubic), FCC (face-centered cubic), IM (intermetallic).

Figure 6 presents the normalized importance scores of key features used for phase prediction in HEAs across three ML models: RF, SVM and NN. The results indicate that VEC consistently ranks as the most significant feature across all models, highlighting its strong correlation with phase stability. This aligns with metallurgical principles, as VEC has been widely recognized for differentiating between BCC and FCC structure in HEAs. The T_m also exhibits high importance, suggesting that thermal stability plays a crucial role in phase formation. The results are consistent with earlier research [30,74], which also identified VEC and T_m as significant predictors of phases in HEAs.
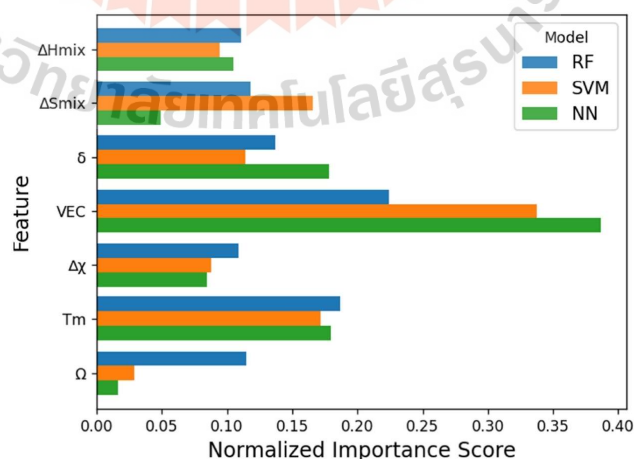
Figure 6. Feature importance ranking for phase prediction in HEAs by RF, SVM and NN models.

3.2. Model Performance for Phase Prediction

In this study, four ML algorithms (SVM, KNN, RF and NN) were employed to predict phase compositions in HEAs based on a structured dataset. To ensure model generalizability and avoid overfitting, hyperparameters were optimized using grid search combined with layered cross-validation. This process involved training the models on four sets of data and validating of them on the fifth, with grid search identifying the best parameter combinations to enhance model performance. Table 4 provides a comprehensive evaluation of these models for each phase (BCC, FCC and IM) using metrics such as mean CV accuracy, test accuracy, precision, recall and F1-score. A summary of the CV and test accuracies in Figure 7 shows that all the models achieve an average CV accuracy above 81%, confirming their viability for phase prediction in HEAs. The following section discusses each model's performance by phase.

1.   SVM

  •   BCC Phases: The SVM achieves a mean CV accuracy of 90.89% and a test accuracy of 92.42%. Precision, recall and F1-score are all above 90%, with precision at 93.15 being the highest. This suggests that the SVM is highly reliable for correctly identifying BCC phases with only minor classifications.

  •   FCC Phases: The SVM performs exceptionally well for the FCC phases with test accuracy and all other metrics at 96.97%. This indicates that the SVM effectively classifies the FCC phase with near-perfect accuracy, likely due to clear distinguishing features in the dataset for this phase.

  •   IM Phases: For IM phases, the SVM shows a lower performance compared to BCC and FCC phases, with a mean CV accuracy of 79.81% and a test accuracy of 85.24%. Precision and recall are both around 81–82%, indicating a moderate performance.

2.   KNN

  •   BCC Phases: The KNN shows excellent performance for BCC phases with a high mean CV accuracy of 92.79% and an impressive test accuracy of 98.48%. Both precision and recall are very high at 97.83% and 98.48%, respectively, indicating that the KNN is particularly effective in distinguish BCC phases.

  •   FCC Phases: The KNN's performance for FCC phases is similarly strong, with a mean CV accuracy of 93.18%, a perfect test accuracy and other metrics at 96.97%. This suggests that the KNN can reliably identify FCC phases, similar to SVM.

  •   IM Phases: The KNN also performs well for IM phases with a mean CV accuracy of 83.27% and a test accuracy of 86.36%. The precision and F1-score are around 85–87%, indicating a balanced performance with a slight improvement over the SVM.

3.   RF

  •   BCC Phases: The RF achieves a mean CV accuracy of 89.36% and a strong test accuracy of 95.45% along with high precision (94.64%) and recall (96.50%). These metrics suggest that RF is highly effective in identifying BCC phases although it slightly lags behind the KNN in precision.

  •   FCC Phases: The RF performs well for FCC phases with a mean CV accuracy of 91.26% and a test accuracy of 93.94%. The precision, recall and F1-scores are all close to 94%, indicating a solid performance but slightly lower than the performance of the SVM and KNN. The RF is still reliable for FCC phases but the slightly lower metrics suggest that FCC features are well-learned but not perfectly.

- IM Phases: The RF's performance for IM phases is moderate with a mean CV accuracy of 81.36% and a test accuracy of 81.81%. Precision and recall are around 80–83%, which are the lowest among the models. This lower performance for IM phases could indicate that RF struggles to differentiate IM from other phases, possibly due to the complex feature space of IM phases or insufficient training samples.

4. NN

- BCC Phases: The NN has the highest mean CV accuracy for BCC phases at 94% and a high test accuracy of 95.45%. Precision, recall and F1-score are all above 94%, showing that the NN performs comparably to the KNN and RF for BCC phases. This suggests that the NN has a strong predictive capability for BCC phases and handles the features well.

- FCC Phases: The NN shows a strong performance for FCC phases with all metrics at 96.97%. This indicates that the NN is reliable and effective for FCC phase classification, similar to the SVM and KNN.

- IM Phases: The NN achieves the highest mean CV accuracy and test accuracy for IM phases at 86% and 87.88%, respectively. Precision and F1-scores are also high at around 87–88%, indicating that the NN is the most effective model for distinguishing IM phases among the four.

Overall, the results indicate that the NN and KNN consistently perform best for BCC and FCC phases followed closely by RF. The SVM also performs well but is slightly behind in term of the recall and F1-score for the BCC phase. IM phase classification has the lowest performance across models, which is likely due to the complex or overlapping features associated with IM phases. The NN's superior performance in this category suggests that it is particularly adept at handling complex, nonlinear patterns in the feature space. This finding highlights that both the KNN and NN are highly effective for BCC and FCC phase predictions, whereas the NN is the most robust for distinguishing IM phases, making it a promising model for complex HEA compositions where phase combinations may be intricate.

Table 4. Cross-validation performance metrics of ML algorithms for each phase prediction in HEAs.

Performance Metric	SVM			KNN			RF			NN		
	BCC	FCC	IM	BCC	FCC	IM	BCC	FCC	IM	BCC	FCC	IM
Mean CV Accuracy (%)	90.89	92.01	79.81	92.79	93.18	83.27	89.36	91.26	81.36	94.00	94.00	86.00
Test Accuracy (%)	92.42	96.97	85.24	98.48	96.97	86.36	95.45	93.94	81.81	95.45	96.97	87.88
Precision (%)	93.15	96.97	81.30	97.83	96.97	86.68	94.64	94.10	82.95	95.45	96.97	87.98
Recall (%)	90.14	96.97	82.15	98.48	96.97	85.34	95.50	93.94	79.98	94.49	96.97	87.12
F1 score (%)	91.38	96.97	82.15	98.48	96.97	85.81	95.04	93.93	81.39	94.94	96.97	87.46

To further assess the performance of the NN model that achieved the highest CV accuracy, additional evaluation methods, namely, the precision–recall (PR) and receiver operating characteristic (ROC) curve, were compiled in this study. These curves help analyze classifier performance and provide visual insights into the model's predictive abilities. The PR curve illustrated in Figure 8a is used to visualize the classifier's predictive accuracy. A larger area under the curve (AUC) indicates better predictive accuracy. For the

NN algorithm, the area for the BCC (PR AUC: 99%) and FCC (PR AUC: 100%) phases shows an excellent performance in predicting BCC and FCC phases. The area of the IM (PR AUC: 82%) phase has a reasonable performance but it is lower than the others. This suggests that the model has some difficulty in consistently identifying IM phase instances correctly compared to BCC and FCC phases. The discrepancy could be due to the inherent differences in the dataset's feature representation or the phase overlap in feature space. Similarly, the ROC curve shown in Figure 8b is another metric used to evaluate the performance of the ML model. The AUC of the ROC curve indicates the classifier's ability to distinguish between positive and negative classes. The diagonal line represents the random classifier baseline (ROC AUC: 50%), which serves as a reference for evaluating the model's predictive power. A model with an ROC curve significantly above this baseline demonstrates strong classification capability, effectively distinguishing between phase categories. The high AUC scores for BCC (ROC AUC: 99%) and FCC (ROC AUC: 100%) phases show that the model is very effective in identifying these two phases. However, the slightly lower AUC for the IM (ROC AUC: 89%) phase suggests that the model has more difficulty in distinguishing this phase, which may be due to fewer training samples for the IM phase in the dataset.

Figure 7. Cross-validation and test accuracy of different models.
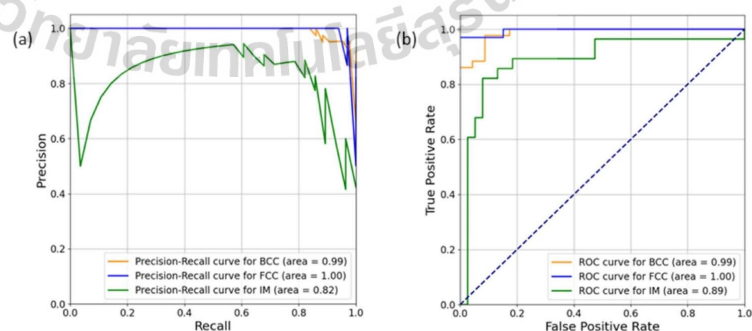

Figure 8. (**a**) Precision-recall (PR) curves for BCC, FCC, and IM phases illustrating the trade-off between precision and recall for each phase. (**b**) Receiver operation characteristic (ROC) curves for BCC, FCC, and IM phases, showing the trade-off between true positive rate and false positive rate.

134

3.3. Complex Combination Phase Prediction Results

The previous sections detailed each model's performance for individual phases, but in practical applications, accurate prediction of complex phase combinations in HEAs is crucial. Therefore, a strict evaluation criterion was applied, where only predictions with entirely correct combinations were scored as accurate. Partially or totally incorrect predictions were excluded from the accuracy calculation. This approach was chosen due to the multiphase nature of the task where accurate prediction of all phase components in each alloy is necessary for the prediction to be considered valid. As illustrated in Figure 9, the complex combination phase prediction accuracy across the models ranges from 75.75% to 84.85%, with notable variations in performance. This strict evaluation criterion reveals that the test accuracy for each model is slightly lower than the mean CV accuracy previously reported. This reduction is expected as the stringent criterion applied during testing requires the model to predict all phases in each combination accurately rather than evaluating them individually. Under this challenging evaluation, the NN and KNN achieved the highest prediction accuracy reaching 84.85%. This demonstrates the NN's and the KNN's robustness in handling complex phase combinations that may involve nonlinear interactions among features. In contrast, the SVM and RF achieved moderate prediction accuracies of 77.27% and 75.75%, respectively, indicating that they may be less reliable for complex combinations where all phases must be identified simultaneously. This difference could be attributed to the feature space complexity and the overlapping characteristics of certain phases.
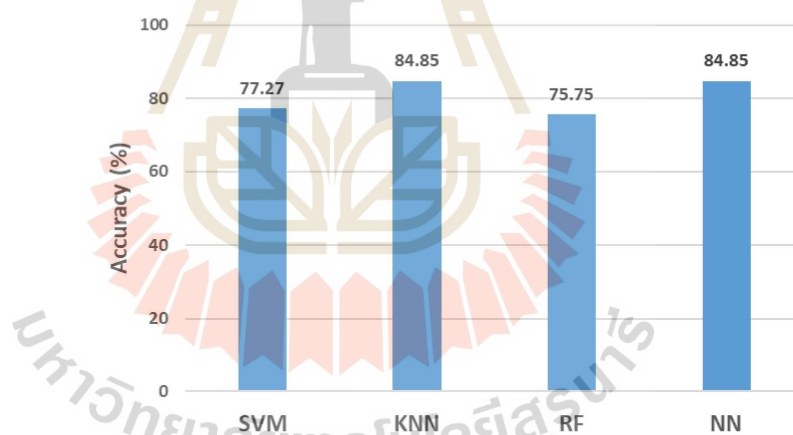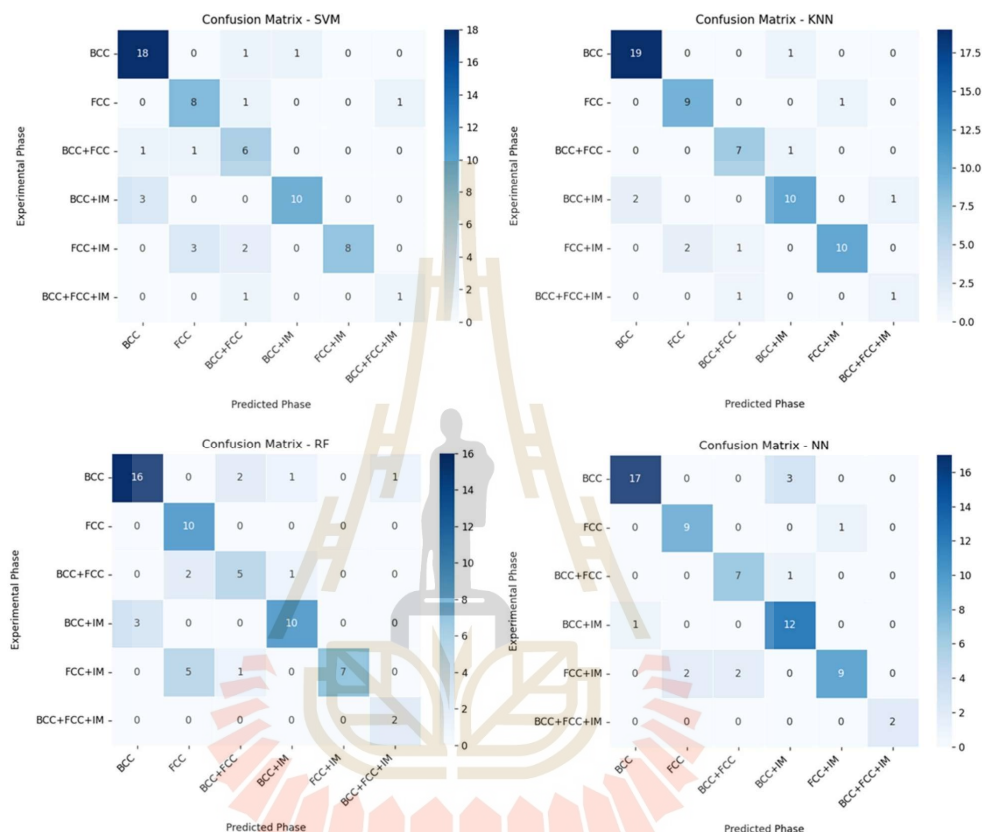

Figure 9. Complex combination phase prediction accuracy of different models.

The confusion matrices (Figure 10) illustrate the classification performance of all models in predicting HEA phase compositions, covering six categories: BCC, FCC, BCC + FCC, BCC + IM, FCC + IM and BCC + FCC + IM. The NN and KNN models demonstrated the highest accuracy, particularly for BCC (NN:17, KNN: 19 correct classifications) and FCC phases (NN/KNN: 9 correct classifications), with minimal misclassifications. SVM and RF showed moderate performance, struggling with multiphase categories due to overlapping feature distributions. BCC + FCC classification was most successful with KNN and NN (7 correct), whereas SVM and RF exhibited confusion in distinguishing binary-phase structures. The BCC + IM and FCC + IM phases were best predicted by NN, achieving 12 and 9 correct classifications, respectively, highlighting its superior ability to lean nonlinear interactions. Misclassifications were more frequent in multiphase categories, suggesting

that overlapping phase features require enhanced feature engineering and dataset balancing. Overall, NN emerged as the best-performing model, excelling in multiphase classification, while KNN performed best for single-phase predictions.

Figure 10. Confusion matrices for SVM, KNN, RF and NN models in complex combination phase prediction of HEAs.

HEAs are a relatively new class of materials, and their experimental data often exhibit inconsistencies due to variations in synthesis techniques, measurement conditions and phase identification methods. Differences in cooling rates, impurity levels and process routes, such as arc melting, induction melting, or mechanical alloying, can significantly influence the observed phase structures. Additionally, discrepancies occur in phase determination techniques, such as X-ray diffraction (XRD) and transmission electron microscopy (TEM). To further minimize inconsistencies, only HEAs synthesized via arc melting were considered in the dataset, ensuring consistency in processing conditions. Despite these efforts, limitations remain due to the relatively small dataset size and unavoidable variations in measurement conditions across different studies. The standard practice in arc melting involves five-time remelting to minimize element segregation and achieve a uniform composition. In contrast, induction melting, which provides better temperature control and mixing efficiency, generally requires only three remelting cycles to achieve similar compositional uniformity. Although arc melting is a conventional approach, induction melting is also an effective technique for HEA production, offering improved control over tempera-

ture and mixing. Future studies could explore how different melting techniques influence phase stability and incorporate these factors into ML models for more accurate prediction.

The dataset used in this study includes both equiatomic and non-equiatomic HEAs, allowing the ML model to learn phase formation trends across a wide range of compositions. Our approach considers variations in elemental ratios, making it applicable to HEAs with diverse compositions. This flexibility enhances the model's ability to generalize and predict phases for alloys beyond strict equiatomic formulations.

IM phases in HEAs can exhibit a variety of structures, including ordered B2, L12, Laves phases, sigma (σ) phases and other complex structures. In this study, IM phases were classified as a single category (IM) due to the limitations of the available dataset, which primarily reports experimental phase compositions without detailed crystallographic structure differentiation. The primary experimental data used for model training were obtained from XRD and TEM analyses, but they do not always specify the exact intermetallic substructures. While the ML model is capable of distinguishing between HEAs containing IM phases and those without, it does not explicitly differentiate between specific intermetallic structures due to the categorical nature of the dataset. The Boolean vector encoding method used in this study assigns a binary value to indicate the presence of an IM phase but does not provide subclassifications. The absence of detailed crystallographic information in the dataset prevents the model from making structure-specific predictions, which could be addressed in future work by incorporating datasets with precise phase labeling, such as those derived from high-resolution TEM or computational approaches like DFT.

The ML approach presented in this study is designed to predict phase compositions of HEAs based on their elemental composition. However, given the vast number of possible HEA compositions, it is impractical to predict phase structures for every conceivable combination of elements with absolute certainty. The predictive capability of the model is inherently limited by the composition range and phase information present in the training dataset. A key limitation of this approach is its reliance on available experimental data, which does not comprehensively cover the entire compositional space of HEAs. For instance, novel compositions that involve elements or atomic ratios outside the dataset distribution may yield unreliable predictions due to the lack of prior learning on such systems. Moreover, phase stability is influenced by thermodynamic and kinetic factors such as cooling rates, processing techniques and synthesis conditions, which are not explicitly considered in the ML model. While the model provides an efficient tool for rapid phase screening, it should not be solely relied upon for predicting the phases of entirely new, unexplored HEA systems without further experimental validation. To address this limitation, future work could incorporate more diverse datasets, including computational phase prediction methods such as CALPHAD and DFT, to enhance model generalization. Additionally, expanding the feature space to include thermodynamic and kinetic descriptors could improve the robustness of phase predictions for a broader range of HEA composition.

One of the key limitations of ML models in HEA phase prediction is the extrapolation challenge. Unlike physics-based models such as CALPHAD, which rely on thermodynamic principles to predict phase stability beyond known data points, ML models learn patterns from historical data and may struggle to make accurate predictions for unexplored compositions. This constraint is particularly important when dealing with HEAs, where minor compositional variables can lead to significant change in phase stability. To improve the generalizability of the model, several strategies can be considered. First, expanding the training dataset by incorporating more experimental data, especially on different synthesis methods (e.g., arc melting, induction melting, power metallurgy), can enhance the model's robustness. Second, transfer learning techniques can be applied, where a pre-trained model on one dataset is fine-tuned using a small dataset of newly explored compositions.

The computational cost associated with model training and prediction is one of the key challenges of applying ML to phase prediction of HEAs. The computational cost expense varies depending on the chosen algorithm, dataset size and complexity of the features. Among the four models that we employed in this study, KNN has relatively low computational cost during training but can become expensive when making predictions, as it requires distance calculation for all data points. SVM, especially with nonlinear kernels, can be computationally demanding during training, as it involve solving quadratic optimization problems, but it remains efficient for predictions. RF, an ensemble-based method, requires training multiple decision trees making training time longer but allowing for fast predictions. NN, particularly deep learning models, are the most computationally expensive due to the need for large datasets and iterative backpropagation processes, but they offer superior performance in capture complex, nonlinear relationships. When compared to traditional phase prediction methods such as CALPHAD and DFT, ML methods provide a faster and more scalable alternative. CALPHAD, based on thermodynamic modeling, requires significant prior knowledge and extensive thermodynamic databases, which may not always be available for HEAs. While CALPHAD predictions are grounded in physics, they often struggle with compositions beyond known datasets. DFT, on the other hand, provides highly accurate atomic-level phase stability predictions but is computationally intensive, limiting its application to small-scale studies.

To verify the predictive capability of the trained ML model, a five-fold CV approach was used to ensure generalizability and robustness. The dataset was split into 80% training and 20% testing sets, where the model was trained on the training set and evaluated on unseen data. Performance was assessed using accuracy, precision, recall and F1-score, with confusion matrices highlighting areas of strong and weak classification. Further validation using new experimental HEA compositions is necessary to confirm its predictive ability for unseen phase structures. Future work should incorporate additional datasets and experimental verification to improve real-world applicability. The prediction uncertainty arises from multiple factors, including dataset limitations, variations in experimental phase identification and model biases. Since our approach focuses on discrete classification rather than continuous property estimation, we do not explicitly compute confidence scores. However, model reliability can still be assessed through CV results, confusion matrix analysis and class-wise performance metrics. CV provides insight into model stability by evaluating accuracy across different data splits, reducing the risk of overfitting to specific compositions. The confusion matrices highlight the classification errors, show which phases are more frequently misclassified and indicate potential uncertainty in specific boundaries. Additionally, the lower predictive performance for certain multiphase combinations (e.g., BCC + FCC + IM) suggests that these phases are harder to distinguish due to their complex structural nature and possible experiment inconsistencies in reported phase data. To further improve uncertainty estimation in future work, techniques such as the Monte Carlo dropout in NNs, ensemble averaging across different ML models or bootstrapping methods could be explored. Expanding the dataset to include more diverse compositions and experimentally verified phase fractions would also enhance model generalization, thereby reducing prediction uncertainty.

Unlike previous studies that focused on binary or ternary phase classification (e.g., distinguishing between BCC, FCC and mixed phases), this study is the first to classify HEA phases into six distinct categories: BCC, FCC, BCC + FCC, BCC + IM, FCC + IM and BCC + FCC + IM. Existing works, such as those by Islam et al. [23], Huang et al. [24] and Dai et al. [25], primarily classified HEAs into fewer categories (BCC, FCC and IM), limiting their ability to predict multiphase combinations. Machaka et al. [27] applied RF, LDA and SVM to classify SS, but their study did not explicitly account for the IM phase. Our

approach introduces a Boolean vector encoding strategy, enabling multiphase classification by treating each phase type as an independent binary variable. This method enhances predictive accuracy, particularly for complex phase combinations like BCC + IM and BCC + FCC + IM, which were often overlooked in prior research.

While this study demonstrates the efficacy of ML in predicting complex phase combination in HEAs, certain limitations should be acknowledged. First, the dataset used in this study, although refined and diverse, remains relatively small for certain phase combinations such as BCC + FCC + IM. The limited sample size for these phases may have affected the model's ability to generalize effectively, particularly for underrepresented combinations. While Boolean vector encoding allowed for multiphase classification, the alternative explicit imbalance handling techniques, such as class weighting or synthetic oversampling were not implemented. This may have impacted the model's ability to generalize to rare phase combinations. Future studies should explore methods such as the synthetic minority over-sampling technique (SMOTE), weighted loss functions, or data augmentation to mitigate this issue and improve predictive performance. Second, although this study employed well-established algorithms like SVM, KNN, RF and NN, the exploration of other advanced ML models, such as ensemble approaches (e.g., gradient boosting, XGboost) or deep learning architectures tailored to materials data, may uncover addition insights or improve prediction accuracy for more complex phase structures. Future studies could explore these alternative approaches. Third, the current study employs binary classification to predict the presence or absence of individual phase components in HEAs. An alternative and potentially more informative approach would be to develop a regression-based model to estimate the exact percentage composition of each phase in alloy, allowing for precise quantification of phase fractions, which could provide better insights into material properties. However, several limitations prevent the use of regression modeling in this study. Firstly, the dataset used consists of categorical phase labels extracted from literature sources, rather than numerical phase fraction data. Experimental phase fraction measurements are rarely reported, making it difficult to accurately train a regression model. Secondly, the relatively small dataset size limits the ability to train a robust regression model without overfitting. For regression-based approaches to be feasible, a significantly larger dataset with labeled phases would be required. Despite these limitations, the binary classification approach used in this study still provides valuable insights into phase formation in HEAs. The model enables researchers to quickly determine the likelihood of phase formation, which can guide experimental alloy design and composition optimization. Future work should focus on expanding the dataset to include quantitative phase fraction data, enabling the development of more precise regression-based predictive models. Another limitation lies in the feature space, which is restricted to features derived from chemical composition. Incorporating additional features, such as processing parameters (e.g., cooling rate or synthesis method) or microstructural descriptors, could further enhance predictive capabilities by capturing more intricate dependencies.

The proposed ML model provides a valuable tool for materials scientists and engineers engaged in the design and development of HEAs. By leveraging predictive phase classification, the model can significantly reduce the reliance on time-consuming and costly experimental phase identification methods. Researchers can utilize the model for accelerated alloy design, allowing rapid screening of candidate compositions before synthesis. Additionally, it aids in guiding experimental efforts by prioritizing compositions with a higher likelihood of exhibiting desired phase structures, thereby minimizing trial-and-error approaches. Furthermore, its predictive capabilities support industrial applications by facilitating the development of HEAs with optimized mechanical, thermal and corrosion-resistant properties, making them suitable for industries.

4. Conclusions

This study presents a machine learning–based approach for predicting phase structures in HEAs. Four ML algorithms including SVM, KNN, RF and NN were evaluated to determine their efficacy in phase prediction. The feature importance analysis highlighted *VEC* and T_m as crucial factors in distinguishing the phase structures of HEAs. Among the algorithms studied, the NN and KNN showed superior performance, achieving a complex combination phase prediction accuracy of 84.85% on the test data. Although the prediction accuracy for BCC and FCC phases was high, the accuracy for IM phases was comparatively lower, suggesting potential limitations in capturing the complexities of IM phases. This result highlights the opportunity for further research focused on improving IM phase prediction accuracy, which could enhance the model's overall predictive capabilities for HEAs. The findings confirm the high effectiveness of machine learning in accurately predicting HEA phase structure and underscore its potential application in materials science for the design and development of HEAs with tailored properties.

Author Contributions: Conceptualization, J.T.; methodology, J.T. and S.K.; software, J.T.; data curation, J.T. and S.K.; validation, J.T.; writing—original draft preparation, J.T.; writing—review and editing, J.T. and S.K. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Yeh, J.-W.; Chen, S.-K.; Lin, S.-J.; Gan, J.-Y.; Chin, T.-S.; Shun, T.-T.; Tsau, C.-H.; Chang, S.-Y. Nanostructured High-Entropy Alloys with Multiple Principal Elements: Novel Alloy Design Concepts and Outcomes. *Adv. Eng. Mater.* **2004**, *6*, 299–303. [CrossRef]
2. Gao, M.; Qiao, J. High-Entropy Alloys (HEAs). *Metals* **2018**, *8*, 108. [CrossRef]
3. Shi, Y.; Yang, B.; Liaw, P. Corrosion-Resistant High-Entropy Alloys: A Review. *Metals* **2017**, *7*, 43. [CrossRef]
4. Li, T.; Wang, D.; Zhang, S.; Wang, J. Corrosion Behavior of High Entropy Alloys and Their Application in the Nuclear Industry—An Overview. *Metals* **2023**, *13*, 363. [CrossRef]
5. Liang, H.; Hou, J.; Liu, J.; Xu, H.; Li, Y.; Jiang, L.; Cao, Z. The Microstructures and Wear Resistance of CoCrFeNi2Mox High-Entropy Alloy Coatings. *Coatings* **2024**, *14*, 760. [CrossRef]
6. Firstov, S.A.; Gorban', V.F.; Krapivka, N.A.; Karpets, M.V.; Kostenko, A.D. Wear Resistance of High-Entropy Alloys. *Powder Met. Met. Ceram.* **2017**, *56*, 158–164. [CrossRef]
7. Hemphill, M.A.; Yuan, T.; Wang, G.Y.; Yeh, J.W.; Tsai, C.W.; Chuang, A.; Liaw, P.K. Fatigue Behavior of Al0.5CoCrCuFeNi High Entropy Alloys. *Acta Mater.* **2012**, *60*, 5723–5734. [CrossRef]
8. Gao, X.; Chen, R.; Liu, T.; Fang, H.; Qin, G.; Su, Y.; Guo, J. High-Entropy Alloys: A Review of Mechanical Properties and Deformation Mechanisms at Cryogenic Temperatures. *J. Mater. Sci.* **2022**, *57*, 6573–6606. [CrossRef]
9. Ujah, C.O.; Von Kallon, D.V. Characteristics of Phases and Processing Techniques of High Entropy Alloys. *Int. J. Lightweight Mater. Manuf.* **2024**, *7*, 809–824. [CrossRef]
10. Yang, X.; Zhang, Y. Prediction of High-Entropy Stabilized Solid-Solution in Multi-Component Alloys. *Mater. Chem. Phys.* **2012**, *132*, 233–238. [CrossRef]
11. Guo, S.; Hu, Q.; Ng, C.; Liu, C.T. More than Entropy in High-Entropy Alloys: Forming Solid Solutions or Amorphous Phase. *Intermetallics* **2013**, *41*, 96–103. [CrossRef]
12. 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.* **2011**, *109*, 103505. [CrossRef]
13. Liu, Y.; Yen, S.; Chu, S.; Lin, S.; Tsai, M.-H. Mechanical and Thermodynamic Data-Driven Design of Al-Co-Cr-Fe-Ni Multi-Principal Element Alloys. *Mater. Today Commun.* **2021**, *26*, 102096. [CrossRef]
14. Singh, P.; Smirnov, A.V.; Alam, A.; Johnson, D.D. First-Principles Prediction of Incipient Order in Arbitrary High-Entropy Alloys: Exemplified in Ti0.25CrFeNiAl. *Acta Mater.* **2020**, *189*, 248–254. [CrossRef]
15. Gao, M.C.; Zhang, C.; Gao, P.; Zhang, F.; Ouyang, L.Z.; Widom, M.; Hawk, J.A. Thermodynamics of Concentrated Solid Solution Alloys. *Curr. Opin. Solid. State Mater. Sci.* **2017**, *21*, 238–251. [CrossRef]

16. Jiang, C.; Uberuaga, B.P. Efficient *Ab. Initio* Modeling of Random Multicomponent Alloys. *Phys. Rev. Lett.* **2016**, *116*, 105501. [CrossRef] [PubMed]

17. Mulewicz, B.; Korpala, G.; Kusiak, J.; Prahl, U. Autonomous Interpretation of the Microstructure of Steels and Special Alloys. *MSF* **2019**, *949*, 24–31. [CrossRef]

18. 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.* **2020**, *185*, 152–158. [CrossRef]

19. Huang, X.; Wang, H.; Xue, W.; Xiang, S.; Huang, H.; Meng, L.; Ma, G.; Ullah, A.; Zhang, G. Study on Time-Temperature-Transformation Diagrams of Stainless Steel Using Machine-Learning Approach. *Comput. Mater. Sci.* **2020**, *171*, 109282. [CrossRef]

20. Elkatatny, S.; Abd-Elaziem, W.; Sebaey, T.A.; Darwish, M.A.; Hamada, A. Machine-Learning Synergy in High-Entropy Alloys: A Review. *J. Mater. Res. Technol.* **2024**, *33*, 3976–3997. [CrossRef]

21. Qiao, L.; Liu, Y.; Zhu, J. A Focused Review on Machine Learning Aided High-Throughput Methods in High Entropy Alloy. *J. Alloys Compd.* **2021**, *877*, 160295. [CrossRef]

22. Jiang, D.; Xie, L.; Wang, L. Current Application Status of Multi-Scale Simulation and Machine Learning in Research on High-Entropy Alloys. *J. Mater. Res. Technol.* **2023**, *26*, 1341–1374. [CrossRef]

23. Islam, N.; Huang, W.; Zhuang, H.L. Machine Learning for Phase Selection in Multi-Principal Element Alloys. *Comput. Mater. Sci.* **2018**, *150*, 230–235. [CrossRef]

24. Huang, W.; Martin, P.; Zhuang, H.L. Machine-Learning Phase Prediction of High-Entropy Alloys. *Acta Mater.* **2019**, *169*, 225–236. [CrossRef]

25. Dai, D.; Xu, T.; Wei, X.; Ding, G.; Xu, Y.; Zhang, J.; Zhang, H. Using Machine Learning and Feature Engineering to Characterize Limited Material Datasets of High-Entropy Alloys. *Comput. Mater. Sci.* **2020**, *175*, 109618. [CrossRef]

26. Zhang, L.; Chen, H.; Tao, X.; Cai, H.; Liu, J.; Ouyang, Y.; Peng, Q.; Du, Y. Machine Learning Reveals the Importance of the Formation Enthalpy and Atom-Size Difference in Forming Phases of High Entropy Alloys. *Mater. Des.* **2020**, *193*, 108835. [CrossRef]

27. Machaka, R. Machine Learning-Based Prediction of Phases in High-Entropy Alloys. *Comput. Mater. Sci.* **2021**, *188*, 110244. [CrossRef]

28. Syarif, J.; Elbeltagy, M.B.; Nassif, A.B. A Machine Learning Framework for Discovering High Entropy Alloys Phase Formation Drivers. *Heliyon* **2023**, *9*, e12859. [CrossRef] [PubMed]

29. Ghouchan Nezhad Noor Nia, R.; Jalali, M.; Houshmand, M. A Graph-Based k-Nearest Neighbor (KNN) Approach for Predicting Phases in High-Entropy Alloys. *Appl. Sci.* **2022**, *12*, 8021. [CrossRef]

30. Gao, J.; Wang, Y.; Hou, J.; You, J.; Qiu, K.; Zhang, S.; Wang, J. Phase Prediction and Visualized Design Process of High Entropy Alloys via Machine Learned Methodology. *Metals* **2023**, *13*, 283. [CrossRef]

31. He, Z.; Zhang, H.; Cheng, H.; Ge, M.; Si, T.; Che, L.; Zheng, K.; Zeng, L.; Wang, Q. Machine Learning Guided BCC or FCC Phase Prediction in High Entropy Alloys. *J. Mater. Res. Technol.* **2024**, *29*, 3477–3486. [CrossRef]

32. Liu, Y.; Zhao, T.; Ju, W.; Shi, S. Materials Discovery and Design Using Machine Learning. *J. Mater.* **2017**, *3*, 159–177. [CrossRef]

33. Gorsse, S.; Nguyen, M.H.; Senkov, O.N.; Miracle, D.B. Database on the Mechanical Properties of High Entropy Alloys and Complex Concentrated Alloys. *Data Brief.* **2018**, *21*, 2664–2678. [CrossRef] [PubMed]

34. Jiang, S.; Lin, Z.; Xu, H.; Sun, Y. Studies on the Microstructure and Properties of AlxCoCrFeNiTi1-x High Entropy Alloys. *J. Alloys Compd.* **2018**, *741*, 826–833. [CrossRef]

35. Song, R.; Ye, F.; Yang, C.; Wu, S. Effect of Alloying Elements on Microstructure, Mechanical and Damping Properties of Cr-Mn-Fe-V-Cu High-Entropy Alloys. *J. Mater. Sci. Technol.* **2018**, *34*, 2014–2021. [CrossRef]

36. Zhao, J.; Gao, X.; Zhang, J.; Lu, Z.; Guo, N.; Ding, J.; Feng, L.; Zhu, G.; Yin, F. Phase Formation Mechanism of Triple-Phase Eutectic AlCrFe2Ni2(MoNb)x (x = 0.2, 0.5) High-Entropy Alloys. *Mater. Charact.* **2024**, *217*, 114449. [CrossRef]

37. Olorundaisi, E.; Babalola, B.J.; Anamu, U.S.; Teffo, M.L.; Kibambe, N.M.; Ogunmefun, A.O.; Odetola, P.; Olubambi, P.A. Thermo-Mechanical and Phase Prediction of Ni25Al25Co14Fe14Ti9Mn8Cr5 High Entropy Alloys System Using THERMO-CALC. *Manuf. Lett.* **2024**, *41*, 160–169. [CrossRef]

38. Tirunilai, A.S.; Somsen, C.; Laplanche, G. Excellent Strength-Ductility Combination in the Absence of Twinning in a Novel Single-Phase VMnFeCoNi High-Entropy Alloy. *Scr. Mater.* **2025**, *256*, 116430. [CrossRef]

39. Záděra, A.; Sopoušek, J.; Buršík, J.; Čupera, J.; Brož, P.; Jan, V. Influence of Substitution of Cr by Cu on Phase Equilibria and Microstructures in the Fe–Ni–Co–Cr High-Entropy Alloys. *Intermetallics* **2024**, *174*, 108455. [CrossRef]

40. Xu, C.; Chen, D.; Yang, X.; Wang, S.; Fang, H.; Chen, R. Enhancing Mechanical Performance of Ti2ZrNbHfVAl Refractory High-Entropy Alloys through Laves Phase. *Mater. Sci. Eng. A* **2024**, *918*, 147438. [CrossRef]

41. Gong, J.; Lu, W.; Li, Y.; Liang, S.; Wang, Y.; Chen, Z. A Single-Phase Nb25Ti35V5Zr35 Refractory High-Entropy Alloy with Excellent Strength-Ductility Synergy. *J. Alloys Compd.* **2024**, *1006*, 176290. [CrossRef]

42. Wei, L.; Liu, B.; Han, X.; Zhang, C.; Wilde, G.; Ye, F. Effect of Al–Zr and Si–Zr Atomic Pairs on Phases, Microstructure and Mechanical Properties of Si-Alloyed (Ti28Zr40Al20Nb12)100-Si (=1, 3, 5, 10) High Entropy Alloys. *J. Mater. Res. Technol.* **2024**, *32*, 2563–2577. [CrossRef]

43. Zhu, C.; Li, X.; Dilixiati, N. Phase Evolution, Mechanical Properties, and Corrosion Resistance of Ti2NbVAl0.3Zrx Lightweight Refractory High-Entropy Alloys. *Intermetallics* **2024**, *173*, 108433. [CrossRef]

44. Kim, Y.S.; Ozasa, R.; Sato, K.; Gokcekaya, O.; Nakano, T. Design and Development of a Novel Non-Equiatomic Ti-Nb-Mo-Ta-W Refractory High Entropy Alloy with a Single-Phase Body-Centered Cubic Structure. *Scr. Mater.* **2024**, *252*, 116260. [CrossRef]

45. Saboktakin Rizi, M.; Ebrahimian, M.; Minouei, H.; Shim, S.H.; Pouraliakbar, H.; Fallah, V.; Park, N.; Hong, S.I. Enhancing Mechanical Properties in Ti-Containing FeMn40Co10Cr10C0.5 High-Entropy Alloy through Chi (χ) Phase Dissolution and Precipitation Hardening. *Mater. Lett.* **2024**, *377*, 137516. [CrossRef]

46. Liu, X.; Feng, S.; Xu, H.; Liu, C.; An, X.; Chu, Z.; Wei, W.; Wang, D.; Lu, Y.; Jiang, Z.; et al. A Novel Cast Co68Al18.2Fe6.5V4.75Cr2.55 Dual-Phase Medium Entropy Alloy with Superior High-Temperature Performance. *Intermetallics* **2024**, *169*, 108301. [CrossRef]

47. Wang, H.; Chen, W.; Liu, S.; Chu, C.; Huang, L.; Duan, J.; Tian, Z.; Fu, Z. Exceptional Combinations of Tensile Properties and Corrosion Resistance in a Single-Phase Ti1.6ZrNbMo0.35 Refractory High-Entropy Alloy. *Intermetallics* **2024**, *171*, 108349. [CrossRef]

48. Wagner, C.; George, E.P.; Laplanche, G. Effects of Grain Size and Stacking Fault Energy on Twinning Stresses of Single-Phase Cr Mn20Fe20Co20Ni40- High-Entropy Alloys. *Acta Mater.* **2025**, *282*, 120470. [CrossRef]

49. Liu, X.; Liu, H.; Wu, Y.; Li, M.; Xing, C.; He, Y. Tailoring Phase Transformation and Precipitation Features in a Al21Co19.5Fe9.5Ni50 Eutectic High-Entropy Alloy to Achieve Different Strength-Ductility Combinations. *J. Mater. Sci. Technol.* **2024**, *195*, 111–125. [CrossRef]

50. Liang, J.; Li, G.; Ding, X.; Li, Y.; Wen, Z.; Zhang, T.; Qu, Y. The Synergistic Effect of Ni and C14 Laves Phase on the Hydrogen Storage Properties of TiVZrNbNi High Entropy Hydrogen Storage Alloy. *Intermetallics* **2024**, *164*, 108102. [CrossRef]

51. Sun, Y.; Wang, Z.; Zhao, X.; Liu, Z.; Cao, F. Effects of Sc Addition on Microstructure, Phase Evolution and Mechanical Properties of Al0.2CoCrFeNi High-Entropy Alloys. *Trans. Nonferrous Met. Soc. China* **2023**, *33*, 3756–3769. [CrossRef]

52. Yao, X.; Wang, W.; Qi, X.; Lv, Y.; Yang, W.; Li, T.; Chen, J. Effects of Heat Treatment Cooling Methods on Precipitated Phase and Mechanical Properties of CoCrFeMnNi–Mo5C0.5 High Entropy Alloy. *J. Mater. Res. Technol.* **2024**, *29*, 3566–3574. [CrossRef]

53. Yu, Z.; Xing, W.; Liu, C.; Yang, K.; Shao, H.; Zhao, H. Construction of Multiscale Secondary Phase in Al0.25FeCoNiV High-Entropy Alloy and in-Situ EBSD Investigation. *J. Mater. Res. Technol.* **2024**, *30*, 7607–7620. [CrossRef]

54. Zhao, Q.; Luo, H.; Yang, Z.; Pan, Z.; Wang, Z.; Islamgaliev, R.K.; Li, X. Hydrogen Induced Cracking Behavior of the Dual-Phase Co30Cr10Fe10Al18Ni30Mo2 Eutectic High Entropy Alloy. *Int. J. Hydrog. Energy* **2024**, *50*, 134–147. [CrossRef]

55. Shafiei, A.; Khani Moghanaki, S.; Amirjan, M. Effect of Heat Treatment on the Microstructure and Mechanical Properties of a Dual Phase Al14Co41Cr15Fe10Ni20 High Entropy Alloy. *J. Mater. Res. Technol.* **2023**, *26*, 2419–2431. [CrossRef]

56. Vaghari, M.; Dehghani, K. Computational and Experimental Investigation of a New Non Equiatomic FCC Single-Phase Cr15Cu5Fe20Mn25Ni35 High-Entropy Alloy. *Phys. B Condens. Matter* **2023**, *671*, 415413. [CrossRef]

57. Tamuly, S.; Dixit, S.; Kombaiah, B.; Parameswaran, V.; Khanikar, P. High Strain Rate Deformation Behavior of Al0.65CoCrFe2Ni Dual-Phase High Entropy Alloy. *Intermetallics* **2023**, *161*, 107983. [CrossRef]

58. Han, P.; Wang, J.; Li, H. Ultrahigh Strength and Ductility Combination in Al40Cr15Fe15Co15Ni15 Triple-Phase High Entropy Alloy. *Intermetallics* **2024**, *164*, 108118. [CrossRef]

59. Li, J.; Zhou, G.; Han, J.; Peng, Y.; Zhang, H.; Zhang, S.; Chen, L.; Cao, X. Dynamic Recrystallization Behavior of Single-Phase BCC Structure AlFeCoNiMo0.2 High-Entropy Alloy. *J. Mater. Res. Technol.* **2023**, *23*, 4376–4384. [CrossRef]

60. Wang, H.; Chen, W.; Chu, C.; Fu, Z.; Jiang, Z.; Yang, X.; Lavernia, E.J. Microstructural Evolution and Mechanical Behavior of Novel Ti1.6ZrNbAl Lightweight Refractory High-Entropy Alloys Containing BCC/B2 Phases. *Mater. Sci. Eng. A* **2023**, *885*, 145661. [CrossRef]

61. Xu, F.; Gao, X.; Cui, H.; Song, Q.; Chen, R. Lightweight and High Hardness (AlNbTiVCr)100-Ni High Entropy Alloys Reinforced by Laves Phase. *Vacuum* **2023**, *213*, 112115. [CrossRef]

62. Chen, B.; Li, X.; Niu, Y.; Yang, R.; Chen, W.; Yusupu, B.; Jia, L. A Dual-Phase CrFeNbTiMo Refractory High Entropy Alloy with Excellent Hardness and Strength. *Mater. Lett.* **2023**, *337*, 133958. [CrossRef]

63. Zhou, J.; Liao, H.; Chen, H.; Feng, D.; Zhu, W. Realizing Strength-Ductility Combination of Fe3.5Ni3.5Cr2MnAl0.7 High-Entropy Alloy via Coherent Dual-Phase Structure. *Vacuum* **2023**, *215*, 112297. [CrossRef]

64. Ren, H.; Chen, R.R.; Gao, X.F.; Liu, T.; Qin, G.; Wu, S.P.; Guo, J.J. Development of Wear-Resistant Dual-Phase High-Entropy Alloys Enhanced by C15 Laves Phase. *Mater. Charact.* **2023**, *200*, 112879. [CrossRef]

65. Wang, S.; Yao, X. Multiclass Imbalance Problems: Analysis and Potential Solutions. *IEEE Trans. Syst. Man. Cybern. B* **2012**, *42*, 1119–1130. [CrossRef]

66. 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.* **2008**, *10*, 534–538. [CrossRef]

67. Zhang, Y.; Wen, C.; Wang, C.; Antonov, S.; Xue, D.; Bai, Y.; Su, Y. Phase Prediction in High Entropy Alloys with a Rational Selection of Materials Descriptors and Machine Learning Models. *Acta Mater.* **2020**, *185*, 528–539. [CrossRef]
68. Hastie, T.; Tibshirani, R.; Friedman, J.H. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed.; Springer series in statistics. Springer: New York, NY, USA, 2009; ISBN 978-0-387-84857-0.
69. Zhang, W.; Li, P.; Wang, L.; Wan, F.; Wu, J.; Yong, L. Explaining of Prediction Accuracy on Phase Selection of Amorphous Alloys and High Entropy Alloys Using Support Vector Machines in Machine Learning. *Mater. Today Commun.* **2023**, *35*, 105694. [CrossRef]
70. Zhang, Z. Introduction to Machine Learning: K-Nearest Neighbors. *Ann. Transl. Med.* **2016**, *4*, 218. [CrossRef]
71. Dewangan, S.K.; Nagarjuna, C.; Jain, R.; Kumawat, R.L.; Kumar, V.; Sharma, A.; Ahn, B. Review on Applications of Artificial Neural Networks to Develop High Entropy Alloys: A State-of-the-Art Technique. *Mater. Today Commun.* **2023**, *37*, 107298. [CrossRef]
72. Arlot, S.; Celisse, A. A Survey of Cross-Validation Procedures for Model Selection. *Statist. Surv.* **2010**, *4*, 40–79. [CrossRef]
73. Armah, G.K.; Luo, G.; Qin, K. A Deep Analysis of the Precision Formula for Imbalanced Class Distribution. *Int. J. Mach. Learn. Cybern.* **2014**, *4*, 417–422. [CrossRef]
74. Zhang, Y.; Zuo, T.T.; Tang, Z.; Gao, M.C.; Dahmen, K.A.; Liaw, P.K.; Lu, Z.P. Microstructures and Properties of High-Entropy Alloys. *Prog. Mater. Sci.* **2014**, *61*, 1–93. [CrossRef]
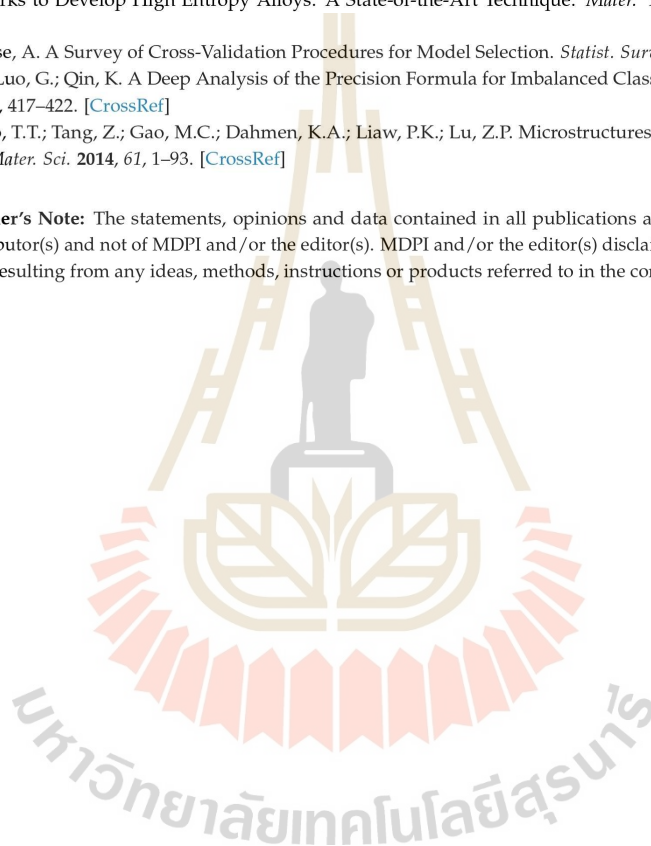
Mechanical Properties Prediction of Austenitic Stainless Steel using Machine Learning

Jirapracha Thampiriyanon[1], Sakhob Khumkoa[2*]

[1,2] School of Metallurgical Engineering, Institute of Engineering, Suranaree University of Technology, Thailand

*Corresponding author, e-mail: sakhob@sut.ac.th

Abstract

The application of machine learning, a branch of artificial intelligence, in material engineering is increasing exponentially. This study proposes the machine learning method for predicting yield strength and ultimate tensile strength of austenitic stainless steel from their chemical composition, solution treatment conditions including temperature during tensile testing. 1,037 samples from the experiments database are used for dataset. DT, SVM, GPR and NN model are selected for training with a dataset to compare their prediction performance. Prediction accuracy of each model is considered from coefficient of determination (R-squared). The results show that GPR model obtained the highest R-squared value for yield strength and ultimate tensile strength 0.96 and 0.95 respectively. This study demonstrates that machine learning method is one of effective tools for predicting mechanical properties of austenitic stainless steel.

Keywords: Machine learning, Tensile properties, Austenitic stainless steel

1. Introduction

Austenitic stainless steel is the most widely used in various industries because of its excellent properties including strong corrosion resistance, high creep resistance, good weldability, easily fabricated and formed [1]. For this reason, it is commonly used in construction of pressure vessel, pipeline, storage tank, etc. The property of austenitic stainless steel is important for their application. Besides corrosion resistance. Tensile properties have attracted much from researchers. There are many studies using theoretical computation and experimental methods that have been used for acquiring tensile properties of austenitic stainless steel [2-4].

Different from methods mentioned above, machine learning is the science of getting a computer to learn to recognize the inner data pattern and construct the model for prediction without being explicitly programmed. It is one of the efficient methods widely used in material engineering [5-7].

In the recent study, Desu et al. [8] applied artificial neural network that is a part of machine learning to predict mechanical properties of 304L and 316L austenitic stainless steels at elevated temperatures from their chemical composition. Wang et al. [9] also used artificial neural network for predicting tensile properties of more general grade of austenitic stainless steel than Desu's study. However, to get better prediction accuracy, alternative machine learning algorithms shall be used for further study.

In this study, machine learning is proposed for predicting yield strength and tensile strength from chemical composition, solution treatment and testing temperature by using various models including DT, SVM, GPR and NN to compare prediction accuracy of each machine learning algorithm.

2. Experimental Procedures

2.1. Dataset

This study used an experimental dataset summarized by Material Algorithm Project (MAP) of University of Cambridge [10]. The given original dataset contains 2,180 samples. but we removed a portion of them due to incomplete information. Only 1,037 samples are retained as final dataset. The chemical compositions, solution treatment condition and test temperature are required as input information. Yield strength and ultimate tensile strength are required as output information. We

performed a statistical analysis of these data for variable range, mean and standard deviation values of each sample as shown in Table 1.

Table 1. Variable range, mean and standard deviation in dataset for construct predictive model for predicting yield strength and ultimate tensile strength.

| Variables	Range	Mean	Standard deviation
C (wt%)	15.90 - 21.06	17.9	1.02
Ni (wt%)	8.40 – 34.45	13.29	6.32
Mo (wt%)	0 – 2.72	0.71	1.6
Mn (wt%)	0.69 – 1.75	1.42	0.28
Si (wt%)	0.29 – 1.15	0.54	0.14
Nb (wt%)	0 – 0.90	0.08	0.22
Ti (wt%)	0 – 0.56	0.24	0.22
V (wt%)	0 - 0.057	0	0.01
Cu (wt%)	0 – 0.35	0.06	0.1
N (wt%)	0 – 0.08	0.01	0.01
C (wt%)	0.012 – 0.12	0.06	0.02
B (wt%)	0 – 0.004	0	0
P (wt%)	0 – 0.04	0.02	0.01
S (wt%)	0 – 0.05	0.01	0.01
Co (wt%)	0 – 0.54	0.05	0.11
Al (wt%)	0 – 0.52	0.05	0.13
Solution treatment temperature (K)	1293 - 1473	1364.41	39.33
Solution treatment time (s)	0 - 7200	1793.4	1928.53
Type of quenching	Water quenched or Air quenched	NA	NA
Testing temperature (K)	293 – 1273	678.62	242.68
Yield strength (MPa)	35 – 341.27	154.68	45.01
Ultimate tensile strength (MPa)	47 - 653.13	425.86	102.01

We randomly divided the final dataset into training set and testing set. 80% of dataset are selected as training set while the remaining 20% dataset are testing set. The training set is used to train the model. The testing set is unseen data that the model never learned before and used to evaluate the final prediction performance of the model. We also applied 5-fold cross-validation method for training set as shown in Figure 1.
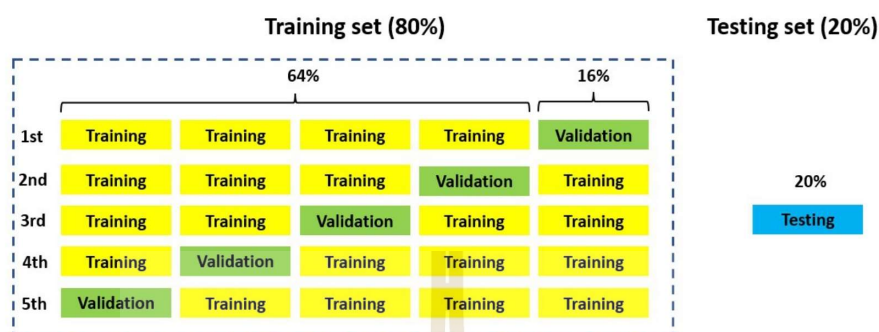
Figure 1. Dataset into training set (80%) and testing set (20%). 5-fold cross-validation is also applied in the training set.

2.2 Machine learning

In this work, we selected several machine learning models commonly used in regression tasks including DT (Decision Tree), SVM (Support Vector Machine), GPR (Gaussian Process Regression) and NN (Neural Network). All models are implemented by Regression Learner App in Matlab R2022a. To construct the predictive model, the dataset has been randomly divided into training set and testing set as mentioned above. Predictive model is the model that has been trained by training data and used for making prediction. The testing set is unseen data that the model never learned before and used to evaluate the final prediction performance of the model. This process is shown in Figure 2.
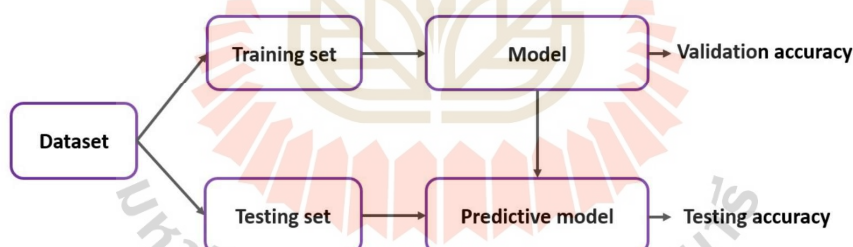
Figure 2. Flow diagram for modeling the predictive model for tensile properties prediction.

3. Results and Discussion

To compare the prediction performance of the model. The plots of data point distribution between actual experimental and predicted value of yield strength and ultimate tensile strength by DT, SVM, GPR and NN model are shown in Figure 3 and 4 respectively. By visually comparing, the data points are distributed more evenly near the asymptote with GPR Model both yield strength and ultimate tensile strength prediction. To quantify the prediction performance, three statistical parameters are compared including coefficient of determination (R-squared), root mean square error (RMSE) and maximum absolute error (MAE). R-squared indicates the proportion of variation in the predicted values and experimental values and can provide the strength of correlation in the linear regression model. These statistical parameters are calculated by using the following equation.

$$R\text{-squared}=\left(\frac{\sum_{i=1}^{n}\left(f(x_i)-\overline{f(x)}\right)\left(y_i-\overline{y}\right)}{\sqrt{\sum_{i=1}^{n}\left(f(x_i)-\overline{f(x)}\right)^2}\sqrt{\sum_{i=1}^{n}\left(y_i-\overline{y}\right)^2}}\right)^2 \tag{1}$$

$$RMSE=\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(f(x_i)-y_i\right)^2} \tag{2}$$

$$MAE=\frac{1}{n}\sum_{i=1}^{n}\left|\frac{f(x_i)-y_i}{y_i}\right| \tag{3}$$

Where n is the total number of samples, $f(x_i)$.

$f(x_i)$ is predicted value of i.

$\overline{f(x)}$ is average predicted value.

y_i is the experimental value of i.

$\overline{y}$ is the average experimental value.

The results of prediction are quantified by R-squared, RMSE, MAE of each model both training set and testing set shown in Table 2. For yield strength prediction, GPR model obtained the highest accuracy with 0.95 R-squared, 9.63 RMSE and 7.21 MAE. The prediction accuracy of SVM model is slightly less than GPR model while DT and NN model have poor accuracy. For ultimate tensile strength prediction, GPR model also obtained the highest accuracy with 0.95 R-squared, 22.63 RMSE and 7.21 MAE. DT model obtained 0.95 R-squared as same as GPR but amount of RMSE and MAE higher than GPR model. This result indicates that GPR obtained the highest prediction accuracy for both yield strength and ultimate tensile strength prediction in this study.

Table 1. R-squared, RMSE, MAE of yield strength and ultimate tensile strength predicted DT, SVM, GPR and NN model for training set and testing set.

Model	Tensile Properties	Training			Testing		
		R-squared	RMSE	MAE	R-squared	RMSE	MAE
DT	Yield Strength	0.75	22.45	16	0.74	23.53	17.1
	Ultimate Tensile Strength	0.91	30.69	18.3	0.95	22.91	16.1
SVM	Yield Strength	0.89	14.85	8.14	0.93	11.84	8.14
	Ultimate Tensile Strength	0.79	46.5	30.6	0.89	34.29	23.8
GPR	Yield Strength	0.93	12.16	8.89	0.96	9.63	7.21
	Ultimate Tensile Strength	0.92	29.03	16.3	0.95	22.63	13.6
NN	Yield Strength	0.73	23.42	17.3	0.74	23.33	17.8
	Ultimate Tensile Strength	0.74	23.42	17.3	0.93	28.18	15.6
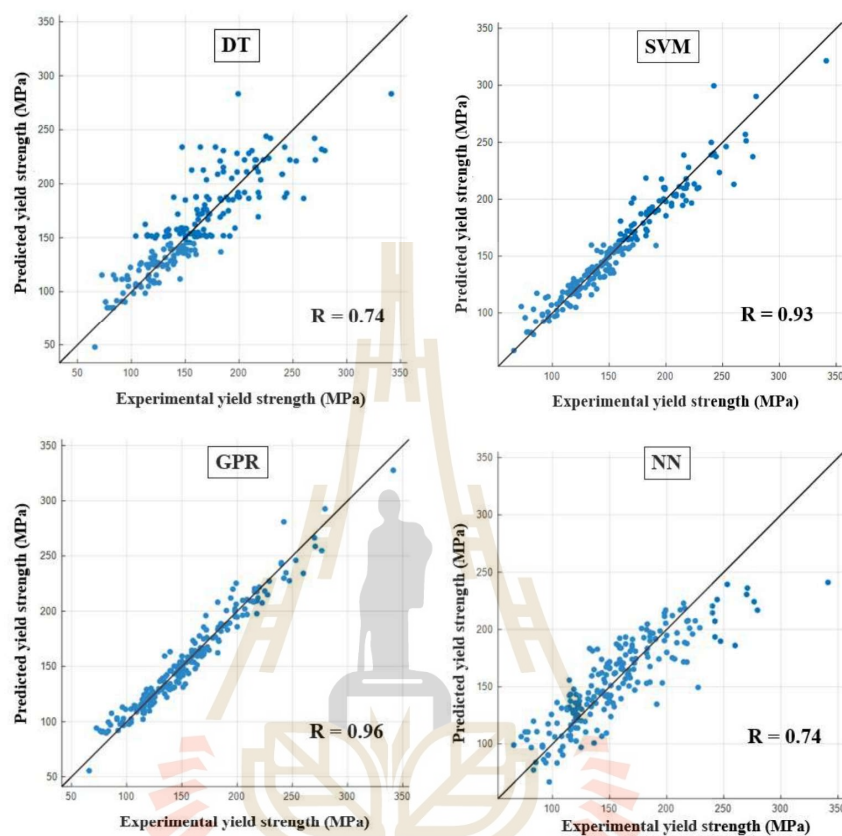
Figure 3. Comparison between predicted and experimental values of yield strength by DT, SVM, GPR and NN.
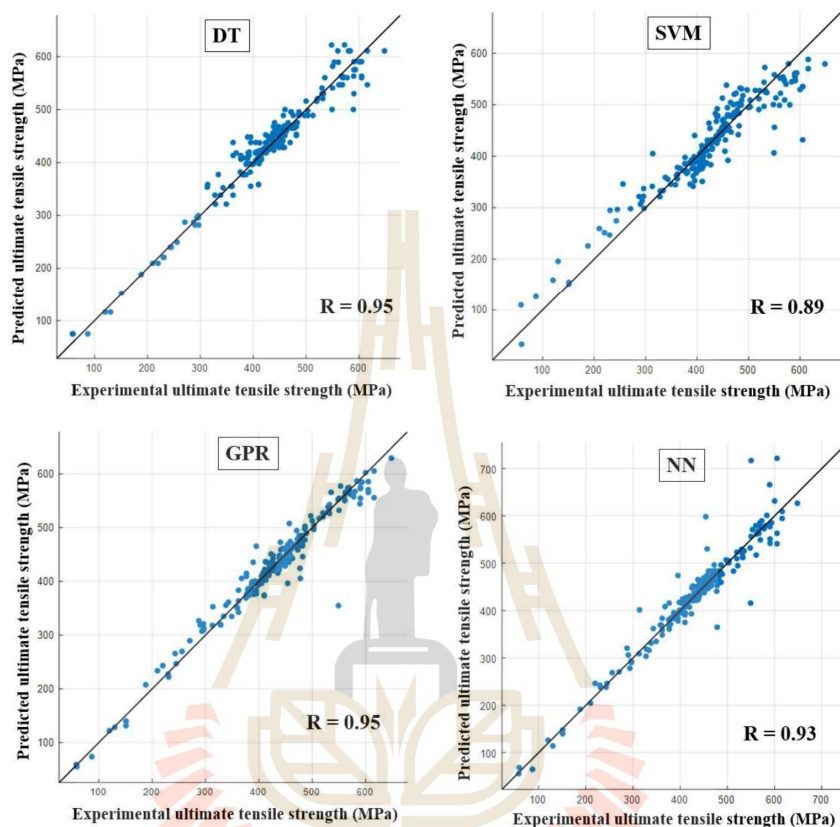
Figure 4. Comparison between predicted and experimental values of ultimate tensile strength by DT, SVM, GPR and NN.

4. Conclusion

In conclusion, machine learning is proposed for predicting the yield strength and ultimate tensile strength from their chemical composition, heat treatment condition and testing temperature. The results demonstrate that machine learning performance which is evaluated by correlation coefficient is highly predictive accuracy and can be one of the effective tools for predicting tensile properties. To improving prediction accuracy of machine learning, Presence of larger dataset will be more helpful samples and result in better prediction accuracy. Detecting and removing the outlier samples that are outside the range of what is expected and unlike the other data that will lead misunderstanding of model during training. Collecting more variables of manufacturing process that are highly correlated to material properties is helpful for machine learning to recognize the pattern in data and results in better prediction.

Acknowledgements

This work is a part of a collaborative research project by Innovative Processing and Recycling of Metal Research Center (IPRMRC) Suranaree University of Technology.

References

[1] S. Singh, R. Praneeth, V. Sankalp, S. Sashank and R. Karthikeyan, Welding, mechanical properties and microstructure of different grades of austenitic stainless steels: a review, Materials today: Proceedings, vol.62, pp.3675-3680, 2022.

[2] S, Park, K. Kim, J. Kang and S. Kim, Effects of carbon and nitrogen on Hall-Petch relationship in austenitic stainless steel, Journal of Materials Research and Technology, vol. 19, pp. 2960-2964, 2022.

[3] H. Wu, B. Yang, M. Zhang, S. Wang and Y. Shi, Effect of Solution Temperature on the microstructure and mechanical properties of wrought 316LN stainless steel, Advanced Materials Research, vol.915-916, pp.576-582, 2014.

[4] D. Morrall, J. Gao, Z. Zhang, K. Yabuuchi, A. Kimura, T. Ishizaki and Y. Maruno, Tensile properties of mechanically alloyed Zr added austenitic stainless steel, Nuclear Materials and Energy, vol. 15, pp.92-96, 2018.

[5] X. Tian, S. Song, F. Chen, X. Qi, Y. Wang, Q. Zhang, Machine learning-guided property prediction of energetic materials: Recent advances, challenges, and perspectives, Energetic Materials Frontiers, vol.3, pp.177-186, 2022.

[6] J. Schmidt, M. Marques, S. Botti and M. Marques, Recent advances and application of machine learning in solid state materials science, npj Computational Materials, vol.5, pp.83, 2019.

[7] C. Chan, M. Sun and B. Huang, Application of machine learning for advanced material prediction and design, EcoMat, vol.4, 2022.

[8] R. Desu, H. Krishnamurthy, A. Balu, A. Gupta and S. Singh, Mechanical properties of Austenitic Stainless Steel 304L and 316L at elevated temperatures, Journal of Materials Research and Technology, vol. 5, pp.13-20, 2016

[9] Y. Wang, X. Wu, X. Li, Z. Xie, R. Liu, W. Liu, Y. Zhang, Y. Xu and C. Liu, Prediction and Analysis of Tensile Properties of Austenitic Stainless Steel Using Artificial Neural Network, Metals, vol. 10, pp.234, 2020.

[10] Materials Algorithms Project. Available online: http://www.phase-trans.msm.cam.ac.uk/map (accessed on 1 Aug 2022)

A Feasibility Study on Metallurgical Slag Classification by Microstructure Recognition

Jirapracha Thampiriyanon[1,a], Kitti Laungsakulthai[1,b], Piamsak Laokhen[1,c], Siam Thongnak[2,d] and Sakhob Khumkoa[1,e]

[1]School of Metallurgical Engineering, Institute of Engineering, Suranaree University of Technology, Thailand

[2]School of Design Technology, Institute of Engineering, Suranaree University of Technology, Thailand

[a]jirapracha@gmail.com, [b]leo8055@hotmail.com, [c]m6111496@g.sut.ac.th, [d]siam@sut.ac.th, [e]sakhob@sut.ac.th

Abstract. Property of metallurgical slag generated in smelting or refining process of ferrous production can be determined by its microstructure which depends on chemical composition and production process. This study proposes a deep learning method which is a subfield of artificial intelligence for autonomous slag classification by microstructure recognition. This present work focus on the implementation of a convolutional neural network (CNN) to classify four types of slags that the variance microstructure resulted from the difference of their formation condition. Both secondary electron (SE) and backscattered electron (BSE) image type captured by scanning electron microscope (SEM) are used as dataset. ResNet50, InceptionV3 and DenseNet201 network architectures are selected in this study to evaluate their classification performance. In addition, data augmentation manipulated by the software is randomly flipped both horizontally and vertically to avoid overfitting from a limited number of training images. The results showed that the best approach to classification accuracy is reached 98.89% by CNN. Therefore, it can be concluded that CNN is excellent potential method for autonomous slag microstructure classification systems.

Introduction

Slag formation during smelting or refining process of ferrous production, e.g. steel and ferroalloy, play significantly roles to cover molten metal from oxidation, trap impurities and inclusions, involve in desulfurization process and improve quality and properties of steel or ferroalloy [1]. The slag microstructure varies depend on the chemical composition, solidification cooling rate, type and amounts of oxides contained during smelting or refining process [2, 3]. Microstructural images are key components for materials analytic approaches. Generally, the slag microstructure classification is mostly performed manually by experts or metallographers due to the complexity of microstructural features. However, the individual background, education and experience may cause potential error in microstructure interpretation from bias and uncertainty. Thus, the development of alternative way for microstructural analysis to get more accurate and precise is required.

A systematic way known as machine learning is a part of artificial intelligence by creating model that can effectively learn from dataset. Machine learning technique, particularly deep learning was successfully applied in various field such as face, voice recognition, text translation, medical diagnostic, image classification and beginning in roadmap of material science research and recovery [4,5]. Convolutional neural network (CNN) is one of the most popular algorithms for image classification. CNN is composed of an input layer, many hidden layers and output layer. Convolution layer puts the input images through a set of convolutional filters which extract feature from the image [6]. Pooling layer simplifies the output by performing nonlinear sampling to reduce the number of parameters need to learn. Rectified linear unit (RELU) maps negative values to zero and maintains positive values. Classification layers is a fully connected layer (FC) that outputs are the number of

classes that network will be able to predict. SoftMax layer is the final layer to provide the classification output. CNN developed by expert such as Xception [7], ResNet50 [8], InceptionV3 [9] and DenseNet201 [10], availability access to massive set of image data and increasing of computing power by graphics processing unit (GPU) help to enable deep learning techniques to obtain higher degree of accuracy.

Prior this study, only a few studies applied CNN to classify types of material based on their microstructural image. One of the recent studies classify 8 different microstructure types of steel (bainite with carbide, granular bainite, columnar bainite, carbide free bainite, ferrite with grain boundaries, ferrite with grain boundaries and small island of pearlite, martensite, pearlite) result from variation in material grade and heat treatment condition by using CNN based microstructural image from light optical microscopy. 1,000 images are selected for training network and the remaining 20,283 images are used for testing. The results confirm that the classification accuracy reached 96.47%(only 715 misclassifications out of 20,283 test images) [11]. Another recent study proposes CNN to classify 4 different microstructure types (martensitic, tempered martensitic, bainitic and pearlitic microstructure). The results shown that their system achieves 93.94% classification accuracy when dataset images is capture by scanning electron microscopy (SEM) [12]. However, there is no existing study of application of deep learning method or CNN in slag microstructural analysis.

Dataset

In this study, four metallurgical slag samples generated from different smelting conditions in recycle process were investigated; (1) Ferrochromium alloy slag discharged from smelting reduction of scale removed from stainless-steel substrate by shot blasting process, (2) Ferromanganese alloy slag discharged from smelting reduction of manganese oxide powder in spent batteries and iron chip from machining process, (3) Ferronickel alloy slag discharged from smelting reduction of spent nickel catalysts and mill scale, and (4) Ferrosilicon alloy slag discharged from smelting reduction of silicon crystal and mill scale. Both secondary electron (SE) and backscattered image (BSE) image type captured at 1000x magnification by scanning electron microscope (SEM) from flat-polished slag samples are used as dataset. The dataset images have been randomly divided into training dataset and validation dataset (10% for training and 90% for validation). The number of training images and validation images of each slag and image type are shown in Table 1 and sample slag microstructural image are shown in Fig. 1 to 4.

Table 1. Dataset information.

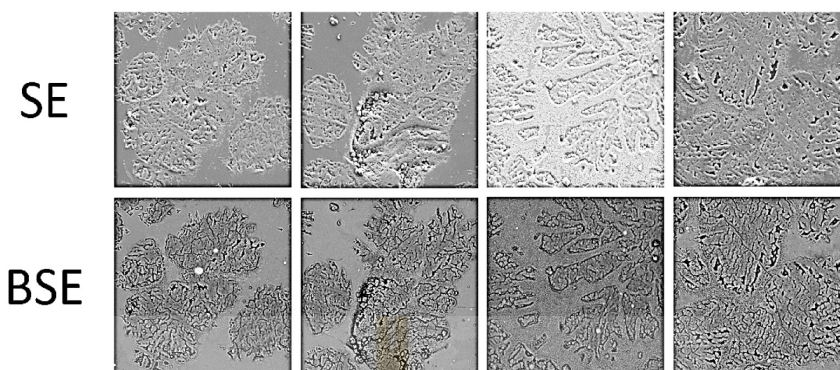| Image Type	Slag Type	Number of Training image	Validation image
SE	Ferrochromium alloy	10	90
	Ferromanganese alloy	10	90
	Ferronickel alloy	10	90
	Ferrosilicon alloy	10	90
Total		40	360
BSE	Ferrochromium alloy	10	90
	Ferromanganese alloy	10	90
	Ferronickel alloy	10	90
	Ferrosilicon alloy	10	90
Total		40	360

Fig. 1. Example of microstructural image of Ferrochromium alloy slag: secondary electron (SE) images (top row) and backscattered (BSE) images (bottom row), 1000x magnification.
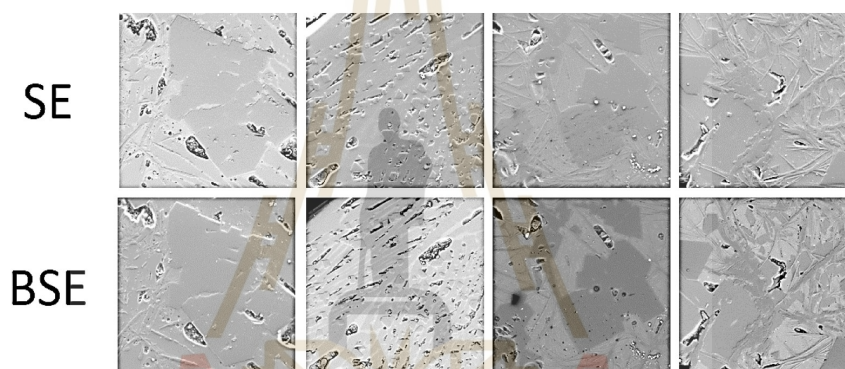

Fig. 2. Example of microstructural image of Ferromanganese alloy: secondary electron (SE) images (top row) and backscattered (BSE) images (bottom row), 1000x magnification.
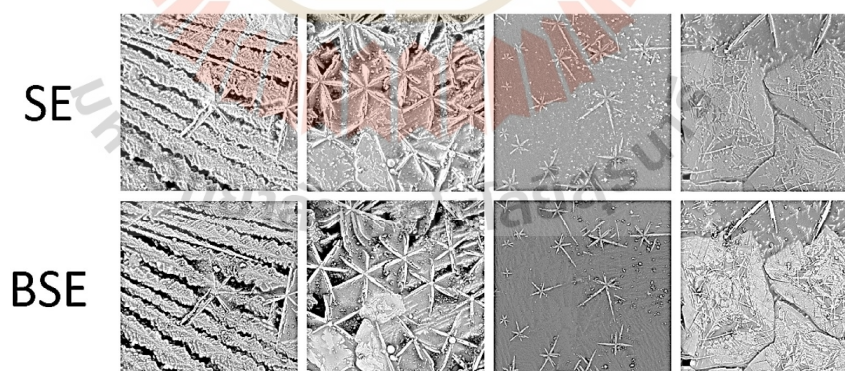

Fig. 3. Example of microstructural image of Ferronickel alloy slag: secondary electron (SE) images (top row) and backscattered (BSE) images (bottom row), 1000x magnification.
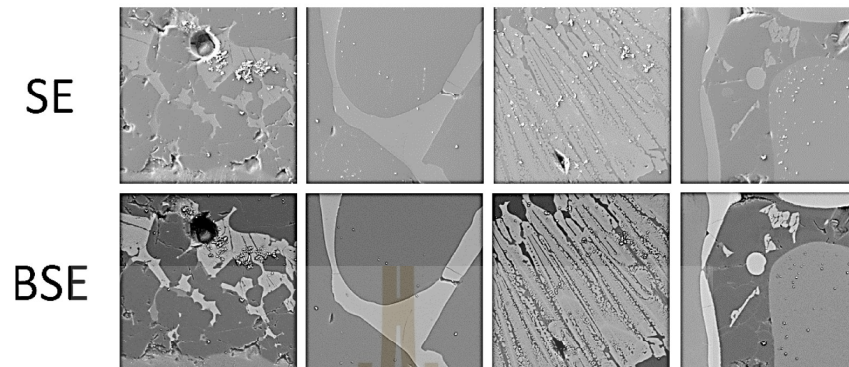
Fig. 4. Example of microstructural image of Ferrosilicon alloy slag: secondary electron (SE) images (top row) and backscattered (BSE) images (bottom row), 1000x magnification.

Methodology

CNN application for autonomous slag classification is trained by training dataset. Data augmentation is a technique that enables the diversity of data applied to training dataset images in order to prevent network from overfitting and memorizing the exact details from very limited number of datasets [6, 13,14] by randomly rotation (horizontally or vertically flipped). When network receives augmented microstructural image as input, network will compute and return output as predict type of slag. Weight inside the network is adjusted to reduce discrepancies between predict and actual type of slag. The way to update weight of network depend on selected optimization algorithm. Network is repeatedly trained until the error is reduced to the satisfactory level or reach to maximum setting loop as illustrated in Fig. 5. After network has been already trained by training dataset, the classification performance of network will be evaluated by prediction of validation dataset.
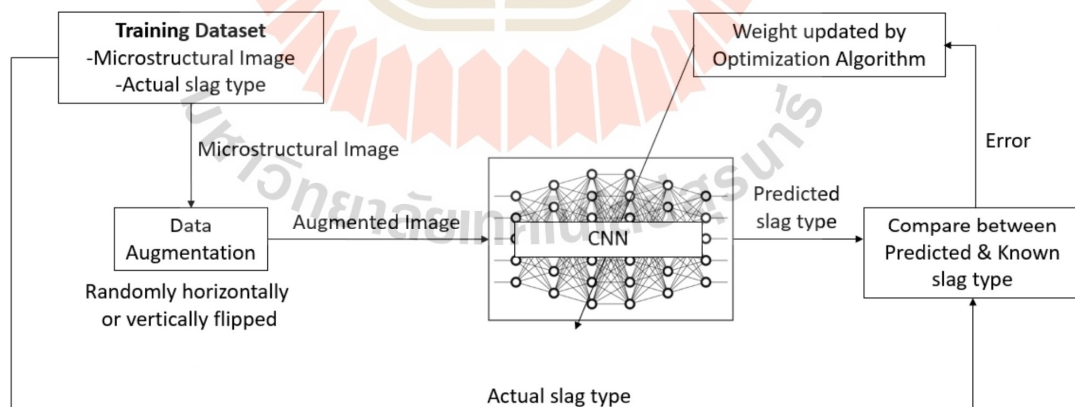

Fig. 5. CNN for slag microstructure classification.

In order to obtain best possible approach, three CNN, well-known network architectures are selected and using for this experiment to find out their classification performance. ResNet50 (Deep Residual Network) is developed by Microsoft team, with 224 x 224 input size [8]. InceptionV3 (Inception) is developed by Google with 299 x 299 input size [9]. DenseNet201(Densely Connected Convolutional Networks) is developed by Cornell University, Tsinghua University and Facebook AI Research (FAIR) with layers 224x224 input size [10]. Stochastic Gradient Descent with Momentum (SGDM) are used during training as algorithm optimizer for weights and biases updating.

All experiments of this study are performed by using Matlab R2019b and a Windows 10 Pro computer with an Intel® Core™ i5-8265U CPU 1.6 GHz and 0.001 initial learning rate, mini batch size of 10 and SGDM algorithm optimizer are the training condition applied for all network.

Results and Discussion

The classification accuracy in this study calculated by fraction of number of correct predictions and total number of predictions. The higher accuracy, the more potential of the proposed deep learning method for improving an autonomous classification system.

The classification performance of networks are trained from scratch with just 40 images per image type (SE and BSE) and validated with 360 images per image type shown in Table 2. The classification accuracy ranges from 88.06 to 98.89.89% and an overall average accuracy is 94.28%. DenseNet201 network architecture with data augmentation for BSE images classification outperformed all other networks with the highest accuracy, 98.89% and its training progress is shown in Fig. 6 classification performance is summarized as confusion matrix shown in Fig. 7.

From the result we can confirm that data augmentation can help network get a small increase of their accuracy by enabling diversity of training data and the classification accuracy obtained from BSE images higher than SE images.

Table 2. Classification performance by combination between network and optimization algorithm.

| Network architecture	Training Data Type	Data Augmentation	Epoch	Elapse Training time (second)	Classification Accuracy [%]
RestNet50	SE image	-	6	724	93.06
RestNet50	SE image	Randomly flipped	6	728	96.39
RestNet50	BSE image	-	6	718	96.39
RestNet50	BSE image	Randomly flipped	6	741	96.67
InceptionV3	SE image	-	12	1,842	88.06
InceptionV3	SE image	Randomly flipped	12	1,833	89.72
InceptionV3	BSE image	-	12	1,892	88.61
InceptionV3	BSE image	Randomly flipped	12	1,646	92.50
DenseNet201	SE image	-	12	3,321	96.39
DenseNet201	SE image	Randomly flipped	12	2,880	96.39
DenseNet201	BSE image	-	12	1,570	98.33
DenseNet201	BSE image	Randomly flipped	12	2,973	98.89
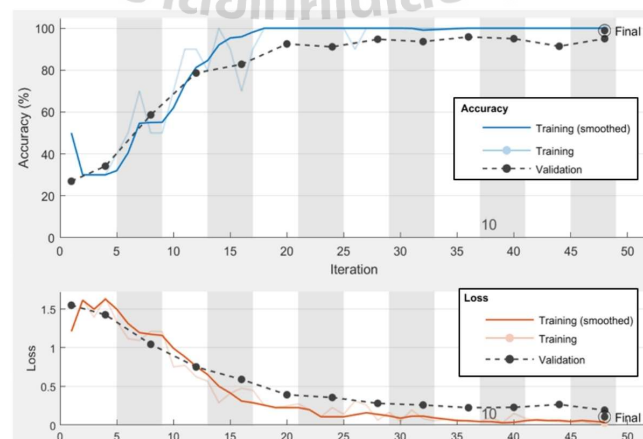
Fig. 6. Training progress of the best approach, DenseNet201 with data augmentation, shows the plot of accuracy curve of and loss curve against iteration.

Fig. 7. Confusion matrix of the best approach, DenseNet201 with data augmentation, shows the number of samples for each class predicted by system. Overall accuracy is 98.89% (4 misclassifications out of 360).

The classification accuracy of CNN be improved in many ways. Increasing model capacity by both increasing number of nodes and layers in the network architecture but this method will be trade-off between accuracy and speed. Adding new training images similarly to misclassified images and retraining network can help network get more learning and improve its weak point. Lastly fine tuning of training parameter is a manually task of trial and error and see which one gives the best results.

Classification of real-world is not an easy task. The challenge factors such as inhomogeneity in slag, may cause they lose key feature to be identified and increasing number of microstructure class also increase chance of confusion in prediction result to decrease network performance. This is required network to generalize inherent pattern when trained with relatively small dataset.

Conclusion

An autonomous classification system based on microstructural image by deep learning is proposed to classfy type of slag. The system intimates the way neurons response in the human brain by visual perception. This study demonstrates that the slag classification accuracy can be achieve 98.89% with very limited training dataset. We conclude that deep learning method is feasible to be one of the alternative methods for an automatic classification system.

Acknowledgment

References

[1] D.R. Gaskell, in: Method for Phase Diagram Determination, Elsevier Science (2007), pp. 442-458

[2] S.K. Park, I.C. Kwon, S.J. Lee, I.K. Huh and N.C. Cho. Microstructural analysis of slags using Raman micro spectroscope, Journal of Conservation Science, Vol. 35, no. 2 (2019), pp. 145-152.

[3] I. Joncsky. Microstructures of Metallurgical slags, Archives of Metallurgy and Materials, Vol. 61 (2016), pp. 61-66.

[4] T. Mueller, A. G. Kusne and R. Ramprasad. Machine Learning in materials science: Recent progress and emerging applications, Reviews in Computational Chemistry, Vol. 29 (2016).

[5] A. Chowdhury, E. Kautz, B. Yener and D. Lewis. Image driven machine learning methods for microstructure recognition, Computational Materials Science, Vol. 123 (2016), pp. 176-187.

[6] P. Kim. Deep Learning: With Machine Learning, Neural Networks and Artificial Intelligence, A press (2017)

[7] F. Chollet "Xception. Deep Learning with Depthwise Separable Convolutions, 2017 IEEE Conference on Computer Vison and Pattern Recognition (CVPR), Honolulu, HI (2017), pp. 1800-1807.

[8] K. He, X. Zhang, S. Ren and J. Sun. Deep Residual Learning for Image Recognition, 2016 IEEE Conference on Computer Vison and Pattern Recognition (CVPR), Las Vegas, NV (2016), pp. 770-778.

[9] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens and Z. Wojna. Rethinking the Inception Architecture for Computer Vision, 2016 IEEE Conference on Computer Vison and Pattern Recognition (CVPR), Las Vegas, NV (2016), pp. 2818-2826.

[10] G. Huang, Z. Liu, L. v. d. Maaten and K.Q. Weinberger. Densely Connected Convolutional Network, 2017 IEEE Conference on Computer Vison and Pattern Recognition (CVPR), Honolulu, HI (2017), pp. 2261-2269.

[11] M. Bartlomiej, K. Grzegorz, K. Jan and P. Ulrich. Autonomous Interpretation of the Microstructure of Steels and Special Alloys, Materials Science Forum, Vol. 949 (2019), pp. 24-31.

[12] S.M. Azimi, D. Britz, M. Engstler, M. Fritz and F. Mücklich. Advanced Steel Microstructural Classification by Deep Learning Method, Scientific Reports, Vol. 8 (2018).

[13] C. Vrancken, P. Longhurst and S. Wagland. Deep learning in material recovery: Development of method to create training database, Expert Systems with Applications, Vol. 125 (2019), pp. 268-280.

[14] C. Shorten and T.M. Khoshgoftaar. A survey on Image Data Augmentation for Deep Learning, Journal of Big Data, Vol. 6 no.60 (2019).

BIOGRAPHY

Jirapracha Thampiriyanon was born in Nakhon Ratchasima, Thailand. He holds a Master of Engineering (M.Eng.) in Welding Engineering from King Mongkut's University of Technology Thonburi (KMUTT) in 2015 and a Bachelor of Engineering (B.Eng.) in Electrical Engineering from Kasetsart University in 2004. In additional to his academic pursuits, he holds several professional certifications in welding inspection, non-destructive testing, and materials integrity assessment, including: AWS Senior Certified Welding Inspector, International Welding Engineer diploma, ASNT NDT Level III, NACE Coating Inspector Level 2. Prior to his Ph.D. studies, he gained significant industry experience for 15 years. He served as Quality Assurance Engineer at Thai Herrick Co., Ltd. (2010 – 2018), where he developed quality management documentation and ensured adherence to quality objectives. Prior to that, he worked as a Production Engineer at Eurotech Engineering International Co., Ltd. (2008 - 2010), overseeing steel fabrication and welding process, and as a Pipeline Protection Engineer at Thai Petroleum Pipeline Co., Ltd. (2004 – 2008) focusing on pipeline integrity corrosion control.