

Research article

A machine learning framework for discovering high entropy alloys phase formation drivers

Junaidi Syarif^{a,b}, Mahmoud B. Elbeltagy^{a,*,1}, Ali Bou Nassif^c^a Department of Mechanical and Nuclear Engineering, University of Sharjah, United Arab Emirates^b Nuclear Energy System Simulation and Safety Research Group, University of Sharjah, United Arab Emirates^c Department of Computer Engineering, University of Sharjah, United Arab Emirates

ARTICLE INFO

Keywords:High entropy alloy
Artificial neural networks
Self organizing maps
Genetic algorithms
Chemical composition
Optimization
Primitive prediction
Phase formation
Data generation

ABSTRACT

In the past years, high entropy alloys (HEAs) witnessed great interest because of their superior properties. Phase prediction using machine learning (ML) methods was one of the main research themes in HEAs in the past three years. Although various ML-based phase prediction works exhibited high accuracy, only a few studied the variables that drive the phase formation in HEAs. Those (the previously mentioned work) did that by incorporating domain knowledge in the feature engineering part of the ML framework. In this work, we tackle this problem from a different direction by predicting the phase of HEAs, based only on the concentration of the alloy constituent elements. Then, pruned tree models and linear correlation are used to develop simple primitive prediction rules that are used with self-organizing maps (SOMs) and constructed Euclidean spaces to formulate the problem of discovering the phase formation drivers as an optimization problem. In addition, genetic algorithm (GA) optimization results reveal that the phase formation is affected by the electron affinity, molar volume, and resistivity of the constituent elements. Moreover, one of the primitive prediction rules reveals that the FCC phase formation in the AlCoCrFeNiTiCu family of high entropy alloys can be predicted with 87% accuracy by only knowing the concentration of Al and Cu.

1. Introduction

High entropy alloys (HEAs) are multi-component alloys that are characterized by having high configurational entropy [1,2]. Furthermore, these alloys reveal superior and interesting properties such as their thermal conductivity behavior versus temperature [3], high wear resistance [4], high yield strength and plasticity [5], corrosion resistance [6], superior elevated-temperature strength [7], and high hardness [8]. However, designing and manufacturing HEAs may be time- and cost-consuming because of their vast design space [9]. The constituent phase is one of the critical considerations for designing HEAs as it significantly impacts the alloy's mechanical properties. Different methods have been implemented for HEAs phase prediction such as empirical rules [10], calculation of phase diagram (CALPHAD) [11,12], and density functional theory [13]. However, those methods are suffering from low accuracy [10], absence of complete datasets and thermodynamic models [9], and high computational cost [14], respectively. On the contrary,

* Corresponding author.

E-mail address: m.bakr9911@gmail.com (M.B. Elbeltagy).¹ This author has equal contribution with the first author.<https://doi.org/10.1016/j.heliyon.2023.e12859>

Received 19 June 2022; Received in revised form 6 October 2022; Accepted 4 January 2023

Available online 13 January 2023

2405-8440/© 2023 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Table 1
A snippet of the training dataset.

Al	Co	Fe	Ni	Ti	Cu	Cr	FCC	BCC	IM
0.0963	0.3073	0.3045	0.2919	0	0	0	1	0	0
0.1742	0.2749	0.2781	0.2728	0	0	0	1	1	0
0.2287	0.2585	0.2613	0.2515	0	0	0	1	1	0
0.2842	0.2392	0.2438	0.2328	0	0	0	1	0	0

machine learning (ML) models showed excellent accuracy in phase prediction while being computationally inexpensive [15–17]. Similar to the CALPHAD method, ML models require the availability of a dataset. Fortunately, the availability of HEA phase data based on the experimental study has been growing because of the increasing research interest in HEA. Thus, ML seems to be the way to go for HEAs phase prediction yet they tend to lack full physical interpretability when complex models are used. In addition to that, normalizing the data is very common in ML practice, which further reduces or even eliminates the model’s physical meaning. Previous researchers succeeded in capturing some quantitative rules/guides of which factors affect phase formation in HEAs using physical quantities (features) that influence phase formation in ordinary alloys. Those quantities were used as predictor variables in the ML framework. In addition, they were able to reveal the feature’s relative importance by using feature selection/ranking algorithms. For example, the authors in [18] used valence electron concentration (VEC), the difference in the Pauling negativities δ_χ , the atomic size difference δ , the mixing enthalpy δH_{mix} , and the mixing entropy δS_{mix} to predict the phase formation in HEAs. They concluded that the valence electron concentration played the most dominant role in determining the ensuing phases. In addition to the previous features, the authors in [15] added five more features, such as the mean melting temperature and the mean bulk modulus. In [17], the authors considered nine thermodynamic properties. The results indicated that elastic energy and atom-size difference contributed significantly to the formation of different phases. Meanwhile, a new set of features was used and introduced in [19]. These features were based on analyzing binary phase diagrams to predict HEA phases. According to the authors, “The hypothesis is that the HEA structural stability is encoded within the binary phase diagram”. While the previous work [15,17–19] used thermodynamic properties, the work in [16] used the constituent elements’ physical and chemical properties in the form of a concentration-weighted sum. The Results in [16] indicated that molar volume, bulk modulus, and melting temperature are the most influential features in HEAs’ phase formation. A common theme of the previous work is the direction of the workflow. The problem starts with defining the descriptor variables, then training the model, and performing some form of feature ranking. Although this has yielded high prediction accuracy, approaching the problem from another direction might give another perspective or at least enforce the previous findings. Thus, this was pursued in this work using the chemical concentration -instead of physical variables- to infer which variables influence the phase formation via a novel algorithm. First, a phase prediction model that uses chemical concentration as input was trained, which is used to increase the size of the dataset by generating hypothetical samples. Then, from the extended dataset, simple/primitive prediction rules were developed using the Pearson correlation coefficient and pruned tree models. Finally, the variables/properties that satisfy the developed rules were searched using self-organizing maps (SOMs) along with genetic algorithm (GA) optimization. We believe that these variables affect HEAs’ phase formation the most. The analysis was performed on AlCoCrFeNiTiCu family of HEAs for two reasons. First, the availability of phase experimental data on this system. Second, these elements are frequently used in engineering alloys.

2. Data generation

2.1. Phase prediction model

2.1.1. Dataset

The phase dataset contained a total of 225 experimental HEA samples with seven features. Those features were the weight percentage of each of the seven constituent chemical elements. The dataset was collected by referring to the individual experimental studies on HEAs’ phase formation. Herein, the phase -target variable- was represented by a Boolean vector with three entries corresponding to $\{FCC, BCC, IM\}$ phases, as shown in Table 1. The logical values in the target variable represent the set of phases present in the corresponding HEA sample. For example, $[1, 1, 0]$ translates to $\{FCC, BCC\}$, whereas $[0, 1, 0]$ translates to $\{BCC\}$. Figs. 1 and 2 show the anatomy of the phase dataset.

2.1.2. Prediction model

A feedforward artificial neural network (FFANN) was used for phase prediction as a result of the previous work of [15,18,20], which showed the superiority of FFANN over other machine learning models in phase prediction. The FFANN was implemented using MATLAB and MATLAB Deep Learning Toolbox. The network had 25, 5, and 3 neurons in the first, second, and third hidden layers, respectively. The network architecture was not optimum, however, it yielded satisfactory results. Neurons used the hyperbolic tangent activation function in the first and second hidden layers, whereas the logistic sigmoid activation function was used in the last hidden layer. The logistic sigmoid function was used in the last layer because its output may be interpreted as the probability of the existence of the corresponding phase. Levenberg-Marquardt backpropagation optimization algorithm and the mean squared error (MSE) cost function were used for training. Furthermore, four-folds cross-validation was used to obtain a better estimation of the performance of the network. The network prediction results are shown in Table 2.

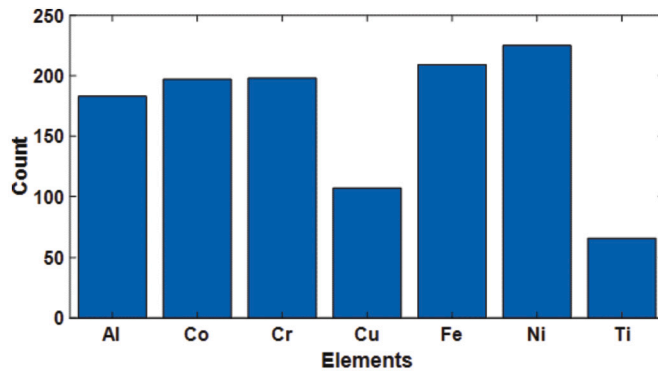


Fig. 1. Histogram of the chemical elements occurrence in the phase dataset.

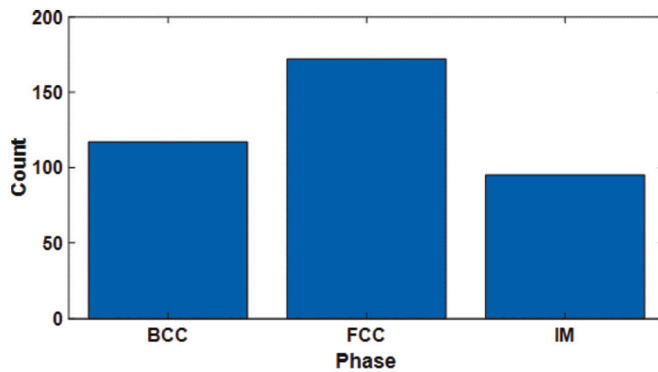


Fig. 2. Histogram of the phase occurrence in the phase dataset.

2.2. Chemical composition generation

Algorithm 1 was used to generate 8,000 hypothetical HEAs. In Algorithms 1 and 2 ‘distribution’ is a function that generates a probability distribution of type ‘distribution_type’ for a given sample space. In addition, ‘sample’ is a function that draws ‘num_of_samples’ samples from a given distribution. In the samples generated by Algorithm 1, the number of chemical elements is constrained to be equal to or greater than four. Moreover, each chemical element concentration is at least 10%. Thus, these two rules preserve the definition of high entropy alloys.

Algorithm 1: $HEA_samples = generate_HEA_samples(l)$.

Input : The number of samples, l
Output : An l by seven matrix representing the chemical composition of an l HEAs, $HEA_samples$

```

1  $HEA\_samples = \text{zeros}(l,7)$ 
2 for  $j = 1 \rightarrow l$  do
3    $n\_distribution = \text{distribution}(\{4,5,6,7\}, 'distribution\_type', \text{uniform})$ 
4    $n = \text{samples}(n\_distribution, 'num\_of\_samples', 1)$ 
5    $i\_distribution = \text{distribution}(\{Al, Co, Fe, Ni, Ti, Cu, Cr\}, 'distribution\_type', \text{uniform})$ 
6    $i = \text{sample}(i\_distribution, 'num\_of\_samples', n)$ 
7    $rand\_comp = \text{gen\_rand\_comp}(0.1, 0.05, n)$  Check Algorithm 2
8    $HEA\_samples[j, i] = rand\_comp$ 
9 end
```

2.3. Phase data generation

The phase data was generated by passing the generated chemical composition through the trained neural network model developed in the previous section.

Table 2

A comparison of the performance measures between our model and other models from the literature. The datasets used in each work are different in terms of the input features and the number of samples (refer to the references).

Performance Measure	FCC	BCC	IM	Model
This work				
Accuracy	-	-	-	92.74
Recall%	97.17	94.03	83.59	-
Precision%	97.72	97.37	84.67	-
F1 score	97.44	95.62	83.82	-
Bakr et al. [21]				
Accuracy%	-	-	-	93.42
Recall%	95.24	94.9	84.98	-
Precision%	96.30	96.03	86.41	-
F1 score	95.76	95.41	85.54	-
Alshibany et al. [20]				
Accuracy%	-	-	-	90
F1 score	89	96	82 ^a	-
Lee et al. [15]				
Accuracy%	-	-	-	93
Pei et al. [16]				
Accuracy%	-	-	-	93.17
Zhang et al. [17]				
Accuracy%	-	-	-	97.43
Dai et al. [22]				
Accuracy%	-	-	-	89
Klimenko et al. [23]				
Accuracy%	-	-	-	91
Han et al. [24]				
Accuracy%	-	-	-	85

Algorithm 2: $gen_rand_comp(C_{min}, \delta, n)$.

Input : Minimum element concentration C_{min} , concentration resolution δ , number of elements n
Output : High entropy alloy random sample composition $rand_comp$

```

1 remained_composition = 1
2 for i = 1 → n do
3   up_lim = remained_comp - (n - i) * C_min
4   s_distribution = distribution({C_min : δ : up_lim}, 'distribution_type', uniform)
5   s = sample(s_distribution, 'num_of_samples', 1)
6   if i == n then
7     | s = remained_composition
8   end
9   rand_comp[i] = s
10 end
11 return rand_comp

```

Table 3

The performance of a Neural Network model trained on the generated data. The model was evaluated on the original experimental set.

Performance Measure	FCC	BCC	IM	Model
Accuracy %	-	-	-	94.55
Recall	97.79	95.90	82.95	-
Precision	98.14	97.40	93.19	-
F1 score	97.96	96.64	89.62	-

2.4. Validating the generated data

The generated samples' validity was ensured by using them to train a new neural network model and then evaluating the model accuracy on the original experimental dataset (225 samples). The data generation and validation process is illustrated in Fig. 3 and the results shown in Table 3 were obtained:

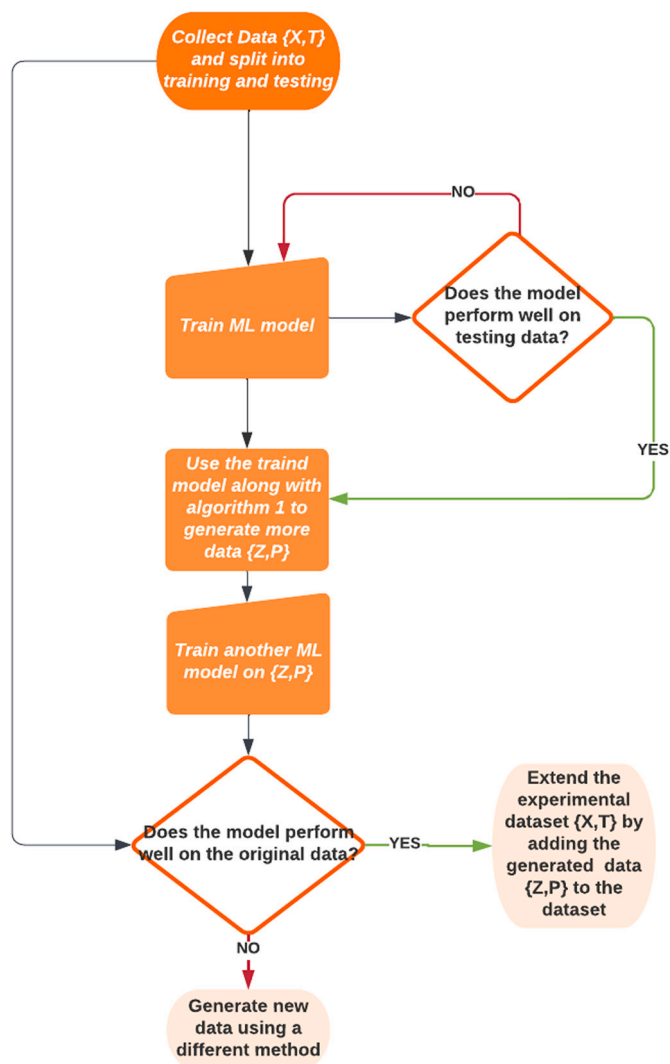


Fig. 3. Illustration showing the steps of the hypothetical data generation and validation.

3. Developing primitive prediction rules

3.1. Methods

The chemical composition effect on each phase has been studied using the generated dataset to develop relations/rules governing the phase formation in $\{Al, Co, Fe, Ni, Ti, Cu, Cr\}$ family of HEAs. For that the Pearson correlation coefficient and classification tree models were used. Tree models have the advantage of interpretability and can capture the nonlinear relations between features and target values missed by using the Pearson correlation coefficient. The maximum number of splits was limited to 20 to maintain the interpretability of the tree models. Meanwhile, in Section 3.2, the original experimental phase dataset (225 samples) was used to evaluate the accuracy of the developed rules.

3.2. Developed primitive rules

The obtained Pearson correlation coefficient is shown in Table 4. Results revealed that FCC formation in AlCoCrFeNiTiCu is primarily affected by the concentration of Al and Cu. In addition, the Pearson correlation coefficient showed that Al has a strong negative correlation and Cu has a strong positive correlation with the formation of the FCC phase. Tree models also yielded the same conclusion. The highest accuracy was associated with using a tree model with only three splits, as shown in Fig. 4. Subsequently, the presence of a particular phase is indicated by a logical value of 1 and vice versa.

For BCC, the Pearson correlation coefficient revealed a high positive correlation between Al, Cr, Fe, and BCC formation. A high negative correlation also existed between Cu, Ni, and BCC formation. On the one hand, for the IM phase, Ti, and Fe exhibited

Table 4
Pearson coefficient of correlation between the chemical elements and {FCC, BCC, IM} phases.

FCC	Element	Al	Ti	Cr	Fe	Co	Ni	Cu
	Correlation	-0.4361	-0.1829	-0.0547	0.0909	0.1059	0.1263	0.3487
BCC	Element	Cu	Ni	Co	Ti	Fe	Cr	Al
	Correlation	-0.3941	-0.3725	-0.1984	-0.0394	0.2851	0.3052	0.3926
FCC	Element	Cu	Co	Cr	Al	Ni	Fe	Ti
	Correlation	-0.2328	-0.2116	-0.1675	0.0115	0.0759	0.2497	0.2666

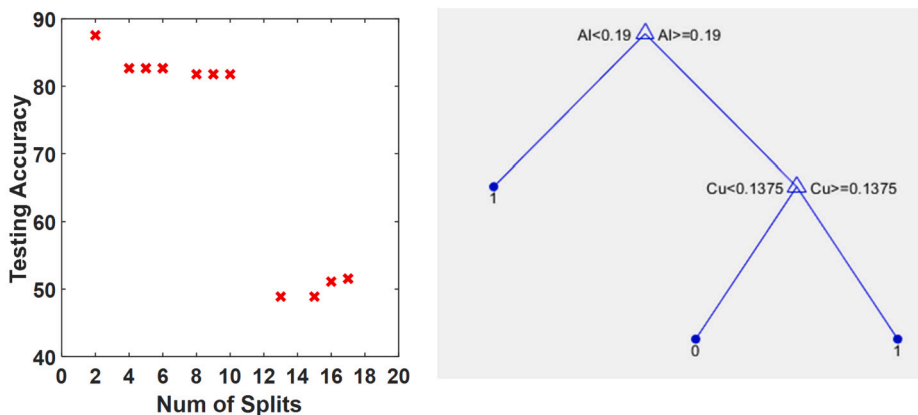


Fig. 4. Left: Accuracy of the FCC tree model VS the number of the tree splits. Right: FCC tree model structure.

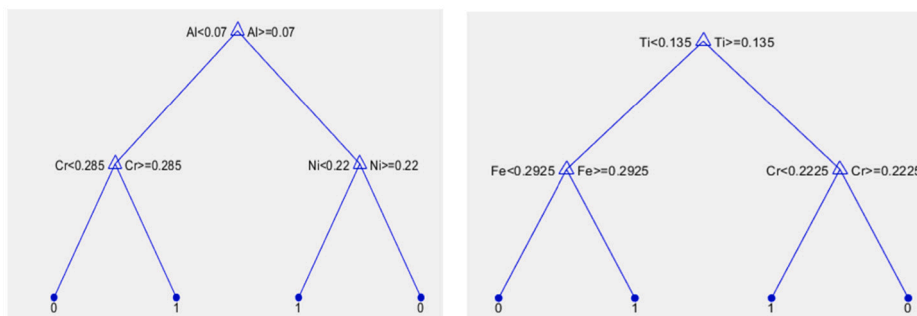


Fig. 5. Tree prediction model of BCC formation (Left) and tree prediction model of IM formation (Right).

Table 5
Elements classification into Stabilizing and Destabilizing groups for each phase.

Phase	Stabilizing set (<i>S</i>)	Destabilizing set (<i>D</i>)
FCC	Cu	Al
BCC	Al, Cr, Fe	Cu, Ni
IM	Ti, Fe	Cu, Co

a high positive correlation, while Cu and Co had a negative correlation. The trained tree models confirmed the previous results, as shown in Fig. 5.

4. Discovering the phase formation drivers

4.1. Problem reformulation

From the results, the chemical elements {AlCoCrFeNiTiCu} can be grouped into a set containing elements stabilizing the formation of a specific phase (*S*), and a set of elements destabilizing the formation of that phase (*D*), as depicted in Table 5.

This categorization is based on the Pearson correlation coefficient and the division learned by the tree models. Then, the question transforms from “what affects the phase formation in a certain alloy?” to “what are the elemental properties that differentiate the given elements into their perspective groups -stabilizing/destabilizing- for each phase?”.

4.2. Elements grouping

4.2.1. Self organizing map

A self-organizing map (SOM) is a type of neural network that uses unsupervised competitive learning and can reveal topological structure within the data. The presence of a topological entity gives rise to the notion of distance. Thus, the answer to our question could be obtained by finding the set of elemental properties e_{opt} that maximizes the distance between elements from different groups, that is $(S - D)$, and simultaneously minimizes the distance between elements from the same group $(S - S)$ OR $(D - D)$ for all phases in a SOM. Let (S) contain n elements, let (D) contain m elements, let χ be a 2D self-organizing map, and let d_i be the position of the i th element in χ . Herein, we define the average distance between the elements in (S) and the elements in (D) , α to be:

$$\alpha = \frac{1}{n * m} \sum_{i=1}^n \sum_{j=1}^m ||d_i - d_j|| \tag{1}$$

Also, we define the average distance between the elements in (S) β_S , and the average distance between the elements in (D) β_D to be -respectively-:

$$\beta_S = \frac{1}{C(n * 2)} \sum_{i=1}^n \sum_{j=1|j \neq i}^n ||d_i - d_j|| \tag{2}$$

$$\beta_D = \frac{1}{C(m * 2)} \sum_{i=1}^m \sum_{j=1|j \neq i}^m ||d_i - d_j|| \tag{3}$$

Where $C(x, 2)$ is the number of possible combinations of a set of x elements taken two at a time and defined as:

$$C(x, k) = \frac{x!}{k!(x - k)!} \tag{4}$$

Now the question can be formulated as an optimization problem to find the solution e_{opt} such that

$$e_{opt} = \min_e(Cost_{tot} + penalty) \tag{5}$$

where

$$\begin{cases} Cost_{tot} = Cost_{FCC} + Cost_{BCC} + Cost_{IM} & \text{(a)} \\ Cost_{phase-i} = (\beta_S + \beta_D - \alpha) ; \text{ phase} - i \in \{FCC, BCC, IM\} & \text{(b)} \end{cases} \tag{6}$$

And $penalty$ is a term that was added to prevent the optimization algorithm from using correlated variables by increasing the cost as the number of properties ($num_properties$) used in the solution is increased

$$penalty = w * num_properties \tag{7}$$

Based on our methods and reasoning, e_{opt} is the set of elemental properties that affect the phase formation in $\{AlCoCrFeNiTiCu\}$ based HEAs.

4.2.2. Euclidean distance

Instead of using SOMs, we can construct a Euclidean space where the space coordinates are the set of the elemental properties. Then we look for the Euclidean space in which the correct chemical elements grouping -Table 5- occurs. This space is constructed by e_{opt} Equation (5), which is the set of properties that minimizes the cost described by Equation (6). Note that to find e_{opt} we use the same method described earlier (Equations (1)-(6)). The only difference is that the distances now are measured in a Euclidean space instead of a Self Organizing Map. This means d from Equations (1), (2), (3) is an l dimensional vector -instead of 2 dimensional vector- that describes the position of a certain chemical element in the Euclidean space, where l represents the space dimension.

4.3. Elemental properties

An element is characterized by its properties. For example, Al is an element with an atomic number equal to 13, FCC crystal structure, melting point = 660.3 °C, and so on. This work included a total of 23 properties, i.e. the size(e) = $[1 \times 23]$. However, the proposed framework can be used with more and properties. The data were obtained from [25] and are provided in the supplementary material.

4.4. Genetic algorithm optimization

Genetic Algorithm (GA) is an evolutionary algorithm based on the concept of natural selection. In GA, a heuristic is reached by the iterative process of elimination -based on an individual’s fitness-, crossover and procreation, and mutation. At each step, the

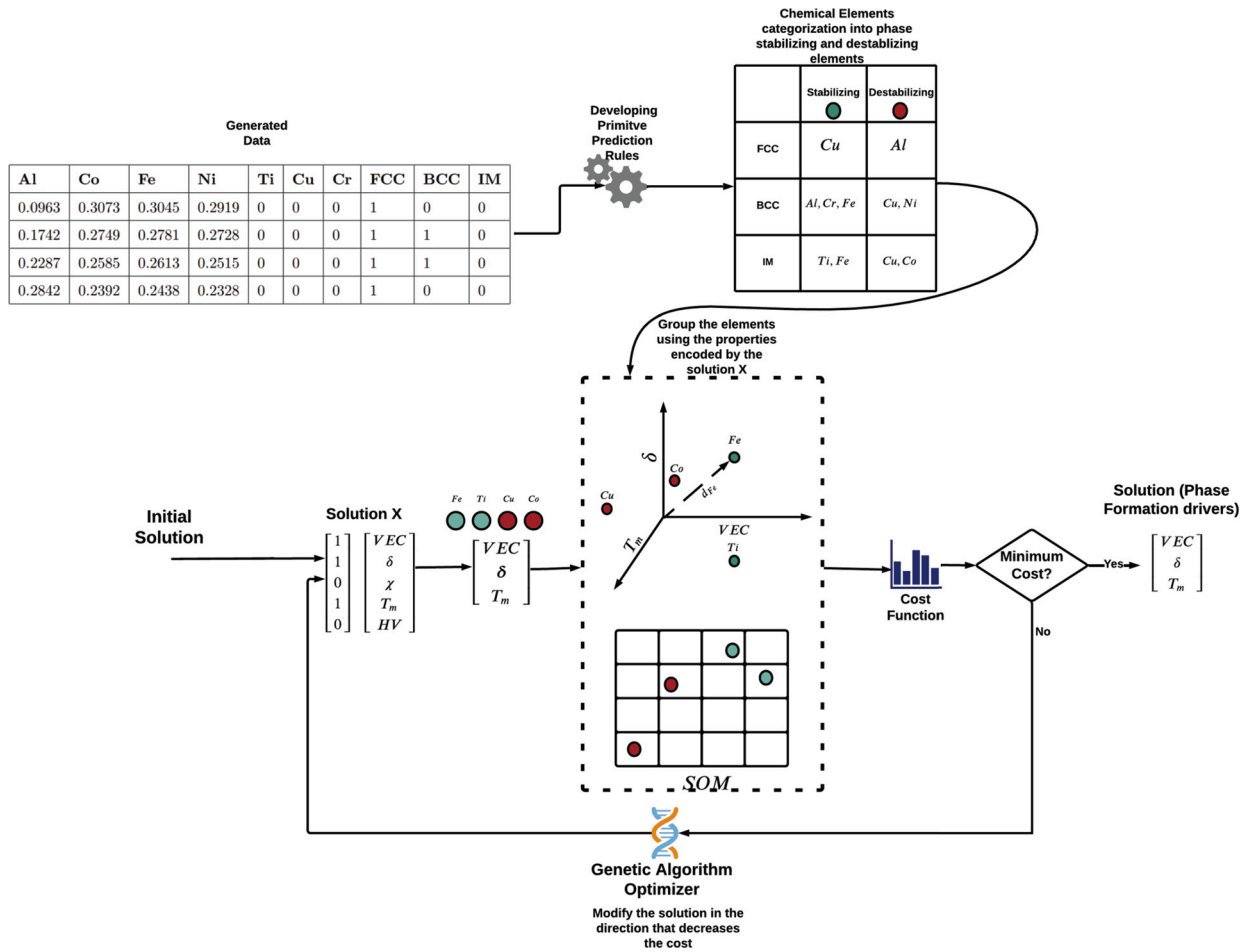


Fig. 6. Illustrative figure showing how the proposed framework works. In the figure, the GA optimizer keeps searching for the set of elemental properties that results in the chemical elements grouping -Table 5- developed in Section 3. Note that only five properties are used in the figures because of the space, the actual number of properties is 23. Also, the figure shows the grouping associated with one phase. The solution in the figure was arbitrarily chosen to clarify the framework. The actual solution is given in Section 4.6.

algorithm evaluates a population of solutions rather than a single one. Each solution (individual) is represented by a chromosome containing k genes. Consequently, the form of the chromosome is problem-dependent. In our case, the chromosome should encode the identity of the elemental properties present in the sets of elemental properties. This was done using a $[1 \times 23]$ binary vector, where 1 indicates the presence of the corresponding property. For example, if the $[VEC, T_m, \delta, HV]$ are the elemental properties, then the set $\{VEC, \delta\}$ is represented by $e = [1, 0, 1, 0]$.

4.5. All together

The workflow illustrated in Fig. 6 started with the GA initializing a random population of sets of elemental properties (individuals/chromosomes). Each individual represented a possible solution. Then, for each individual, a SOM was trained (through unsupervised learning) to group the elements based on the properties encoded in the individual's genome. Afterwards, the fitness of a given individual was evaluated using Equation (6). A crossover was performed between the fittest - lowest cost - population members to produce the next generation, and the process was repeated until the optimum solution was reached. Occasionally, a mutation occurred in the offspring generation with a probability of $P_{mutation}$. Fig. 6 shows how the proposed framework components interact with each other. The same applies when using Euclidean distances except in that case the SOM step will be replaced by building a Euclidean space, as shown in Fig. 7.

4.6. Results and discussion

The results obtained from the GA that implemented SOMs in the cost function showed that electrical resistivity, molar volume, and electron affinity are the properties responsible for the phase formation in $\{AlCoCrFeNiTiCu\}$ based HEAs. This result was consistent on average, in the sense that different repetitions of running the optimization -with different SOM initial weights- yielded

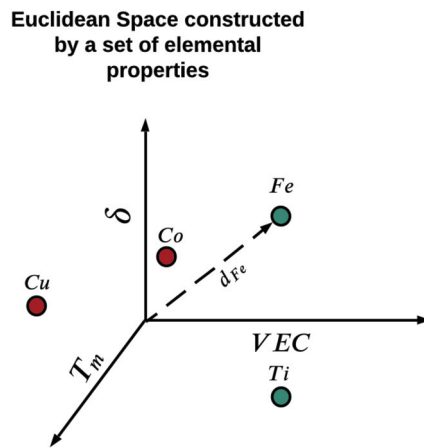


Fig. 7. A Euclidean space constructed by $\{VEC, \delta, T_m\}$. The color indicates the elements grouping.

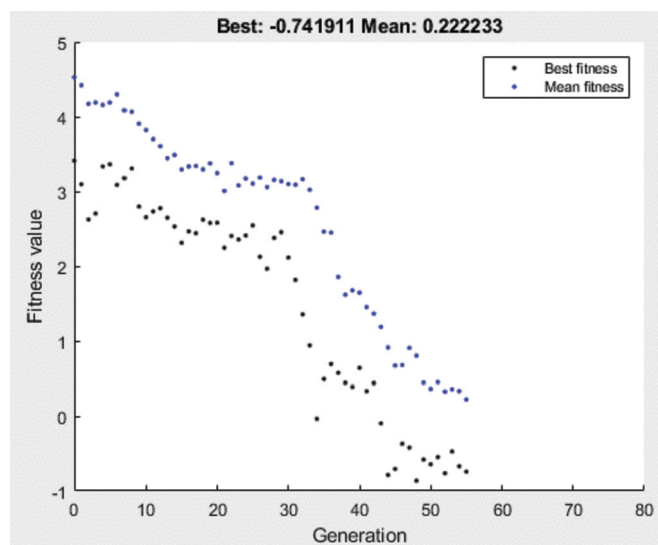


Fig. 8. The cost function value during the progress of the optimization algorithm.

this solution the most when the GA optimization reached a cost $cost_{tot}$ such that $cost_{tot} \in [0, -1]$. The progress of the optimization using SOMs is shown in Fig. 8. Meanwhile, when the Euclidean space method was used, the results showed that density, specific heat, and electron affinity are the properties that drive the phase formation. Both methods report the electron affinity. In addition, both molar volume and density are related as they both indicate the space that the material occupies. Moreover, the results obtained by the Euclidean space method were consistent.

Pei et al. [16] found that molar volume, bulk modulus, and melting temperature are the factors that affect phase formation in HEAs. It is seen that molar volume is a common property between the two findings. It should be mentioned that bulk modulus and melting temperature were part of the solution in a few runs when SOMs were used with a smaller weight factor (w) (Equation (7)). The properties encoded by the solution are believed to be a proxy to a more fundamental properties on the atomic level.

5. Conclusion

In this work, a neural network (NN) was trained on 225 experimental samples to predict the phase of $\{AlCoCrFeNiTiCu\}$ HEAs. The NN showed an $F1 - score$ of 97.4% and 95.6% for FCC and BCC phase prediction, respectively. On the contrary, IM phase prediction had a lower $F1 - score$ of 83.8%. Then, the model was used to generate a total of 8,000 hypothetical samples. Herein, the generated data were used to train another NN model. The model resulted in high accuracy when tested on the experimental dataset, thus suggesting the validity of the generated samples. Pearson correlation coefficient along with pruned tree models applied to the generated samples was used to develop simple rules governing the phase formation in $\{AlCoCrFeNiTiCu\}$ system of HEAs. For FCC phase formation, the result revealed that the phase could be predicted – with 87% accuracy- by only knowing the concentration of

both *Al* and *Cu*. Similar -and less accurate- rules were also developed for BCC and IM phase formation. Furthermore, for any phase, the chemical elements could be grouped into a class that supports that phase formation and a class that discourages it. Based on that, elemental properties responsible for the stabilizing/destabilizing categorization are also responsible for the phase formation. Therefore, an optimization problem based on that assumption was formulated such that it searches for the set of elemental properties that maximizes the distance between the chemical elements from different groups and minimizes it for elements belonging to the same group. SOMs and Euclidean spaces were used to introduce that notion of distance/topology to the problem, and GA optimization was used to search for the optimum solution. The solution suggests that resistivity, molar volume, and electron affinity are the properties responsible for the phase formation in {*AlCoCrFeNiTiCu*} based HEAs.

CRedit authorship contribution statement

Junaidi Syarif: Contributed reagents, materials, analysis tools or data; Analyzed and interpreted the data; Wrote the paper.

Mahmoud Bakr: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Wrote the paper.

Ali Bou Nassif: Contributed reagents, materials, analysis tools or data; Wrote the paper.

Data availability

The raw and processed data required to reproduce these findings are provided as supplementary material.

Declaration of competing interest

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

Appendix A. Supplementary material

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.heliyon.2023.e12859>.

References

- [1] Y. Zhang, M. Gao, J. Yeh, P. Liaw, Y. Zhang, High-entropy alloys: fundamentals and applications, 2016.
- [2] Y. Jien-Wei, Recent progress in high entropy alloys, *Ann. Chim. Sci. Mat* 31 (6) (2006) 633–648.
- [3] H.-P. Chou, Y.-S. Chang, S.-K. Chen, J.-W. Yeh, Microstructure, thermophysical and electrical properties in alxcoyfez (0 ≤ x ≤ 2) high-entropy alloys, *Mater. Sci. Eng. B* 163 (3) (2009) 184–189.
- [4] M.-H. Chuang, M.-H. Tsai, W.-R. Wang, S.-J. Lin, J.-W. Yeh, Microstructure and wear behavior of alxco1. 5crfeni1. 5tiy high-entropy alloys, *Acta Mater.* 59 (16) (2011) 6308–6317.
- [5] C.-C. Juan, K.-K. Tseng, W.-L. Hsu, M.-H. Tsai, C.-W. Tsai, C.-M. Lin, S.-K. Chen, S.-J. Lin, J.-W. Yeh, Solution strengthening of ductile refractory hfmoxnbtatiz high-entropy alloys, *Mater. Lett.* 175 (2016) 284–287.
- [6] X.-W. Qiu, Y.-P. Zhang, L. He, C.-g. Liu, Microstructure and corrosion resistance of alcrfecuco high entropy alloy, *J. Alloys Compd.* 549 (2013) 195–199.
- [7] C.-J. Tong, M.-R. Chen, J.-W. Yeh, S.-J. Lin, S.-K. Chen, T.-T. Shun, S.-Y. Chang, Mechanical performance of the al x cocrufeni high-entropy alloy system with multiprincipal elements, *Metall. Mater. Trans. A* 36 (5) (2005) 1263–1271.
- [8] X.-W. Qiu, C.-G. Liu, Microstructure and properties of al2crfecocutinix high-entropy alloys prepared by laser cladding, *J. Alloys Compd.* 553 (2013) 216–220.
- [9] D. Miracle, High entropy alloys as a bold step forward in alloy development, *Nat. Commun.* 10 (1) (2019) 1–3.
- [10] S. Guo, Phase selection rules for cast high entropy alloys: an overview, *Mater. Sci. Technol.* 31 (10) (2015) 1223–1230.
- [11] Y.-c. Liu, S.-y. Yen, S.-h. Chu, S.-k. Lin, M.-H. Tsai, Mechanical and thermodynamic data-driven design of al-co-cr-fe-ni multi-principal element alloys, *Mater. Today Commun.* 26 (2021) 102096.
- [12] V.K. Soni, S. Sanyal, K.R. Rao, S.K. Sinha, A review on phase prediction in high entropy alloys, *Proc. Inst. Mech. Eng., Part C, J. Mech. Eng. Sci.* (2021) 095440622111008935.
- [13] P. Singh, A.V. Smirnov, A. Alam, D.D. Johnson, First-principles prediction of incipient order in arbitrary high-entropy alloys: exemplified in ti0. 25crfenialx, *Acta Mater.* 189 (2020) 248–254.
- [14] S. Hernandez, Development of methods for reducing the cost of density functional theory and time-dependent density functional theory, Ph.D. thesis, UCLA, 2015.
- [15] S.Y. Lee, S. Byeon, H.S. Kim, H. Jin, S. Lee, Deep learning-based phase prediction of high-entropy alloys: optimization, generation, and explanation, *Mater. Des.* 197 (2021) 109260.
- [16] Z. Pei, J. Yin, J.A. Hawk, D.E. Alman, M.C. Gao, Machine-learning informed prediction of high-entropy solid solution formation: beyond the hume-rothery rules, *npj Comput. Mater.* 6 (1) (2020) 1–8.
- [17] L. Zhang, H. Chen, X. Tao, H. Cai, J. Liu, Y. Ouyang, Q. Peng, Y. Du, Machine learning reveals the importance of the formation enthalpy and atom-size difference in forming phases of high entropy alloys, *Mater. Des.* 193 (2020) 108835.
- [18] N. Islam, W. Huang, H.L. Zhuang, Machine learning for phase selection in multi-principal element alloys, *Comput. Mater. Sci.* 150 (2018) 230–235.
- [19] J. Qi, A.M. Cheung, S.J. Poon, High entropy alloys mined from binary phase diagrams, *Sci. Rep.* 9 (1) (2019) 1–10.
- [20] Z.Y.A. Al-Shibaany, N. Alkhafaji, Y. Al-Obaidi, A.A. Atiyah, Deep learning-based phase prediction of high-entropy alloys, in: *IOP Conference Series: Materials Science and Engineering*, vol. 987, IOP Publishing, 2020, p. 012025.
- [21] M. Bakr, J. Syarif, I.A.T. Hashem, Prediction of phase and hardness of heas based on constituent elements using machine learning models, *Mater. Today Commun.* 31 (2022) 103407.
- [22] D. Dai, T. Xu, X. Wei, G. Ding, Y. Xu, J. Zhang, H. Zhang, Using machine learning and feature engineering to characterize limited material datasets of high-entropy alloys, *Comput. Mater. Sci.* 175 (2020) 109618.

- [23] D. Klimenko, N. Stepanov, R. Ryltsev, S. Zherebtsov, Phase prediction in high-entropy alloys with multi-label artificial neural network, *Intermetallics* 151 (2022) 107722.
- [24] Q. Han, Z. Lu, S. Zhao, Y. Su, H. Cui, Data-driven based phase constitution prediction in high entropy alloys, *Comput. Mater. Sci.* 215 (2022) 111774.
- [25] T. Gray, Periodic table, <https://periodictable.com/>, created by Theodore Gray, with assistance from Nick Mann, and in partnership with Max Whitby of RGB Research.