

Article

Automatic Featurization Aided Data-Driven Method for Estimating the Presence of Intermetallic Phase in Multi-Principal Element Alloys

Upadesh Subedi ^{1,*} , Yuri Amorim Coutinho ² , Prafulla Bahadur Malla ³ , Khem Gyanwali ¹ 
and Anil Kunwar ^{4,*} 

¹ Thapathali Campus, Department of Automobile and Mechanical Engineering, Institute of Engineering, Tribhuvan University, Kathmandu 44600, Nepal; gyanwalikhem@ioe.edu.np

² Department of Materials Engineering, KU Leuven, Kasteelpark Arenberg 44, B-3001 Leuven, Belgium; yuri.coutinho@kuleuven.be

³ College of Architecture and Environment, Sichuan University, Chengdu 610065, China; praf_malla@hotmail.com

⁴ Scientific and Didactic Laboratory of Nanotechnology and Materials Technologies, Faculty of Mechanical Engineering, Silesian University of Technology, Konarskiego 18A, 44-100 Gliwice, Poland

* Correspondence: upadesh.subedi@thc.tu.edu.np (U.S.); anil.kunwar@polsl.pl (A.K.)

Abstract: Multi-principal element alloys (MPEAs) are characterized by a high-dimensional materials design space, and data-driven models can be considered as the best tools to describe the structure–property relationship in this class of materials. Predicting the prevalence of an intermetallic (IM) phase in a high-entropy alloy (HEA) regime of MPEAs has become a very important research direction recently. In this work, Automatic Featurization capability has been deployed computationally to extract composition and property features from the datasets of MPEAs. Data visualization has been performed, and through principal component analysis, the relative impacts of the input features on the two principal components have been specified. Artificial neural network is then trained upon the set of composition, property and phase information features. A GUI interface is subsequently developed on top of the prediction model to enable the user-friendly computer environment for detection of the IM phase in a compositionally complex alloy.

Keywords: automatic featurization; intermetallic compounds; high entropy alloys; principal component analysis; neural network analysis



Citation: Subedi, U.; Coutinho, Y.A.; Malla, P.B.; Gyanwali, K.; Kunwar, A. Automatic Featurization Aided Data-Driven Method for Estimating the Presence of Intermetallic Phase in Multi-Principal Element Alloys. *Metals* **2022**, *12*, 964. <https://doi.org/10.3390/met12060964>

Academic Editor: Diego Celentano

Received: 30 April 2022

Accepted: 31 May 2022

Published: 4 June 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The procedures and methods of phase detection in high-entropy alloys (HEAs) are at the forefront of the materials research community in today's time. After it was established by Cantor [1] and Yeh [2] in two separate research works that HEA can form suitable and desirable phases structures and can have potentially much more beneficial uses compared to traditional single principal element alloys, there have been enormous efforts from the researchers within and beyond the materials science research community regarding the improvement of the techniques of phase detection in these new multi-principal element alloys (MPEAs). The fact that HEAs are structurally different from a traditional or conventional alloy system poses several difficulties and challenges in using empirical rules for the phase detection and prediction in the former materials. So, a new method or a shift in paradigm for the phase prediction or classification is exactly the need of this hour [3].

In order to properly understand the usage of the terms—high-entropy alloys, medium-entropy alloys (MEAs), multi-principal element alloys and compositionally complex alloy—throughout this work; it is necessary to briefly distinguish them from the viewpoint of composition and entropy-based definitions. An MPEA has at least three principal elements [4] and has the configurational entropy always larger than 1 R (where R is the

universal gas constant). The compositional definition criteria stipulate that the constituent principal elements of an MPEA have a composition range constrained between 5 at % and 35 at %. Now, HEA is a subset of the alloys of the MPEA class characterized by configurational entropy larger than $1.5 R$ [5]. The number of principal elements in HEA is equal to or larger than five. MEAs on the other hand, have configurational entropy between $1 R$ and $1.5 R$, and the number of components in MEA is either three or four.

In this new era of informatics where machine learning and artificial intelligence can be found being used ubiquitously in almost every field of research and development, the material science community is also gearing up with the trend of these new state-of-the-art technologies to accelerate materials design and discovery. Especially, artificial neural network (ANN) has been very popular in the composition-based and mechanical property-based design of alloys [6,7]. Different statistical tools and machine learning tools are being used by many research works for the general phase classification of the HEAs [8]. As the most preferred phase in the alloy system is a simple solid solution (SSS) phase with a definite crystal structure, more and more research work is found to be dedicated for the detection/prediction of solid solution phase. However, in the context of the current scenario where the advancement in materials technology is accelerating at a greater pace, the beneficial aspects of other phases such as intermetallics compounds (IMC) and amorphous phases should also be considered as the topic of greater importance. Even some researchers are proposing to make a focus shift toward IMC containing HEAs [9,10].

Intermetallic (IM) phases have found applications in superconductors, high hardness materials, catalytic materials development and energy storage technology such as lithium-ion batteries, fuel-cells, etc. [3]. Although the design of IM phases in the context of conventional alloys is not a very new material research area, their design and discovery in HEAs materials is a very new research idea. The state-of-the-art knowledge of structural features and the prevalence of IMCs in HEA is considered to be quite a cumbersome and challenging process due to the need for a fully experimental study by literally making the alloy in the first place. Since MPEAs are located at or near the central region of the multi-component materials design space, they are thought to possess a unique blend of enormous desirable and exceptional properties [11,12]. These materials are often referred to as emergent materials. However, as the multi-component MPEAs lack a universal phase diagram [13], it is quite understandable that the IM phase design or discovery in such materials remains an unexplored area. The first step in studying these emergent materials is to develop robust yet lucid computer tools that can help with identifying or detecting the presence of the IM phase.

The key introductory step in the HEA materials design is to know its phase and microstructure, which then can help in identifying their potential superior qualities. The datasets are the founding source of information for knowledge development regarding the phase stability detection in HEA materials. Machine learning and statistical tools are the most appropriate techniques for determining the phase stability criteria for intermetallic phase formation in MPEAs (especially in context of HEAs). In this work, an artificial neural networks (ANN) algorithm will be implemented to functionally approximate the relationship between the IMC phase stability and the corresponding composition and property features of the MPEAs. Then, a graphical user interface application called “IMCATHEA” will be built on the top of this machine learning model to detect whether the system of high-entropy alloy contains an intermetallic compound. In our previous work [11], ANN was employed to predict all the phases (solid solution, intermetallic and amorphous phase) present in an MPEA. The pyMPEALab GUI was developed upon a 34-element framework from a dataset consisting of 1229 observations, and the output of the ANN was a multi-label classification problem. Moreover, the composition and property features for these 1229 features were entered manually in the dataset file in our previous work [11].

The machine learning model development work in the present work differs from or evolves upon our earlier work [11] in three major themes—(a) mathematical structure of the output layer of the model, (b) size of dataset, and (c) feature extraction capability.

For an ANN to address the specific need of IMC detection (presence or absence in an MPEA system), its output layer's mathematical formulation has to represent a binary label classification problem. The IMCATHEA toolkit developed in the present work will be solely based upon this purpose of IMC phase detection in a compositionally complex alloy system. Moreover, this work will encompass an expanded dataset—consisting of 1301 observations and a total of 41 elements. To enable the acceleration of the data-driven algorithm development procedure, it is imperative to develop efficient libraries that can help with automatically preprocessing the data. This work is a breakthrough from our previous work, as it will implement an Automatic Featurization capability during the data preprocessing stage that can extract the composition and property features from the information of the name of the MPEAs.

The methodology outlining the procedure of developing the IMCATHEA GUI software toolkit is schematically presented in Figure 1. As shown in the figure, the total process is categorized into two major stages. Stage I is associated with the data treatment and model training from the processed data. This is followed by the construction of the GUI interface on top of the prediction model in stage II.

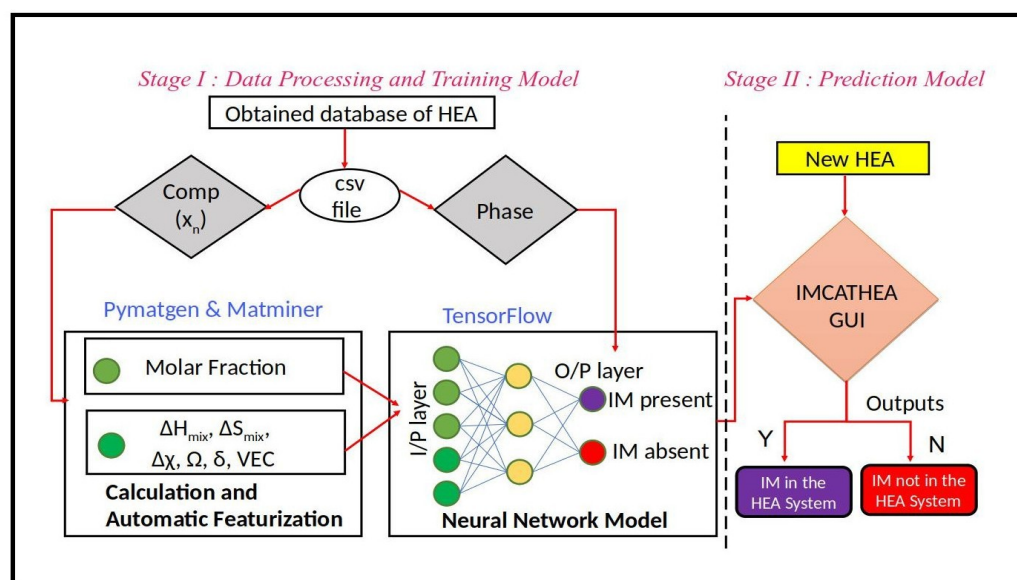


Figure 1. The construction of the digital toolkit for the prediction of the presence of an intermetallic (IMC or IM) phase in a given HEA system is complete within the two stages. Stage I involves data preprocessing and training of the neural network, and the subsequent stage II incorporates the deployment of a GUI interface built upon the prediction model. The capabilities of Pymatgen and Matminer libraries are utilized to establish the numerical fingerprints associated with the input composition features. These composition features are then utilized by the python code to compute the property features. Literally, the calculation and featurization task in stage I is enriched with automatic featurization capability.

2. Materials and Methods

This section is dedicated to outlining the details of data treatment and model training procedures (stage I). Collecting the data of multi-principal element alloys (MPEAs), featurization of data, visualization of the features and methodology for training the data using neural network model are the steps required to accomplish this stage. Although the result of model training and validation includes the components of stage I (Figure 1, it is excluded from this section and will be presented in Section 3.

2.1. Data Preparation

In data-driven models, the first and the mandatory step is the collection and processing of the datasets. The dataset consisting of the phase(s) information of 1301 MPEAs was obtained from existing literature [4,8,13–19]. The originally collected data were first obtained as a csv file with two columns—(i) name of the MPEA and (ii) thermodynamic information. As illustrated in Figure 1, the first part of data processing is the calculation and featurization of the collected data. Since the nomenclature of an MPEA consists of the information of the constituent elements and their relative proportions, it can be operated on by the Python-based libraries to extract the composition features needed by the machine learning model. The molar fraction of the constituent elements (composition features) was extracted from the data file using opensource Python libraries: Pymatgen (Python Materials Genomics) [20,21] and Matminer [22]. For the present work, the original dataset consists of a total of 41 elements (Li, Be, B, Na, Mg, Al, Si, Ca, Sc, Ti, V, Cr, Mn, Fe, Co, Ni, Cu, Zn, Ge, Sr, Y, Zr, Nb, Mo, Ru, Pd, Ag, Cd, In, Sn, Sb, La, Ce, Nd, Gd, Yb, Hf, Ta, W, Au, Bi), and so, an MPEA is digitally identified or represented by numerical values of 41 mole fraction features. A value of zero was assigned to the mole fraction of the elements that are not present in the alloy. This featurization of alloy composition was further employed to systematically compute the six thermophysical properties—atomic size difference (δ), entropy of mixing (δs_{mix} or δs_m), enthalpy of mixing (δh_{mix} or δh_m), electronegativity, valence electron concentration (VEC) and the Omega parameter (Ω). The mathematical definitions of these six properties have been provided in our previous works [11]. One hot encoding method was applied to convert phase labels into numerical fingerprints. As the purpose of this study is to design the software library capable of finding whether a given material consists of an IMC phase or not, the phase data were further simplified into two main labels: IM and Not IM. The IM class consisted of the group of materials that has intermetallic present solely existing or co-existing among other phases. On the other hand, Not IM implied that the MPEAs are devoid of IMC phases. The Python code for the calculation and featurization procedure outlined in Figure 1 is made available online at <https://github.com/subediupadesh/AutomaticFeaturizerMPEA>. It is important to note that this Python computing the alloy properties using solely the composition feature information will serve as the preprocessing calculation library in the GUI app discussed in Section 3.2. Owing to its capability to first extract the composition feature from the name of MPEA and then compute thermodynamic/physical and chemical properties using these composition features, the Python code has been termed “Automatic Featurizer for MPEA”.

The script for the automatic featurization was tested for the computation of the thermophysical and chemical properties of many MPEAs. Table 1 presents an illustration of the computation made in $\text{Ti}_{0.8}\text{CoCrFeNiCu}$ MPEA (HEA containing BCC + Laves phase). Once the name of the MPEA is entered as a syntax `mpea = 'Ti0.8CoCrFeNiCu'` in the script, the feature extraction is initiated, and the properties are subsequently calculated. It can be inferred that the values of Ω , melting temperature, enthalpy of mixing, entropy of mixing, and the atomic size difference computed by the Automatic Featurizer library for this MPEA are in reasonable agreement with the values presented in the work of Yang and Zhang [23].

Table 1. Tabulation of the values of the properties—VEC, Ω , T_m , ΔH_{mix} , ΔS_{mix} , $\Delta\chi$ and δ of $Ti_{0.8}CoCrFeNiCu$ HEA, computed using the Automatic Featurizer library in this present work. It is to be noted that the melting temperature (T_m) property of the HEA is being used in the computation of the Ω property feature, and thus, T_m data will not be considered as a separate feature in the ANN of this present work. In this table, the values of Ω , T_m , ΔH_{mix} , ΔS_{mix} , and δ for the HEA calculated in this work are compared with the corresponding values presented in Yang and Zhang [23].

Properties	Values (This Work)	Values (Ref. [23])
VEC	8.138	—
Ω	3.992	3.95
Melting temperature (T_m)	1792.68 K	1785.66 K
ΔH_{mix}	−6.6783 kJ/K	−6.75 kJ/K
ΔS_{mix}	14.87 J/(mol K)	14.89 J/(mol K)
$\Delta\chi$	0.1332	—
δ	6.5 %	5.26 %

2.2. Visualization of Data

The preliminary visualization of the data was accomplished using pairplot diagrams. The features corresponding to the thermodynamic, physical and chemical properties as well as phase information were visualized using the pairplots in Figure 2. In the top image (Figure 2a), the data corresponding to ΔH_m , ΔS_m , $\Omega (= \frac{T_f \Delta S_m}{|\Delta H_m|})$ and VEC are presented. On the other hand, the bottom image (Figure 2b) consists of the information about the features ΔH_{mix} , number of components (N), electronegativity difference ($\Delta\chi$) and atomic size difference (δ). In both images, IM and Not IM labels have been distinguished for the data points. These combined pairplots are very informative in terms of obtaining a summarized insight of the scatter of the MPEAs on the property and components coordinates. For an example, looking at the pairplot between ΔH_m and ΔS_m in Figure 2a, it can be observed that the alloy samples are uniformly scattered. However, the use of the Ω parameter against ΔH_m makes the data more skewed, suggesting that this feature will aid the ML algorithm to learn more easily. Moreover, such graphical plots are helpful for having an idea about the presence domain of IM phases. In Figure 2b, observing the plots of N versus one of the other properties— ΔH_m , $\Delta\chi$ and δ , it can be inferred that the intermetallics phase is mainly prevalent in materials with a number of components (N) equal to or less than 3 (while the materials with $N = 3$ in the dataset are medium-entropy alloys, those with $N = 2$ are not listed under the category of MPEAs). The IMC phase is also commonly present for MPEAs with the high number of components ($N = 7$), but it has a lower probability of occurrence for the materials with an intermediate number of components ($N = 4$ – 6). Hence, it can be inferred that the IM phase is less prevalent in medium-entropy alloys of $N = 4$ as compared to MEAs with $N = 3$. The probability of prevalence of the IM phase in HEAs is however the opposite trend as compared to medium-entropy alloys. IMCs are not so commonly found in HEAs, as they have fewer components ($N = 5$ and 6), whereas they become more statistically prevalent with HEAs that have a larger number of elements ($N = 7$). This is consistent with the findings reported in Senkov et al. [24]. It is quite obvious that the extremely low magnitude of configurational entropy at $N = 2$ or 3 is favorable for the prevalence of an intermetallic phase. While the rise in configuration entropy is dominant enough to suppress the intermetallic phase formation at $N = 4$, 5 and 6 , it is quite inadequate to do so at $N = 7$ or higher. At such a higher number of components, the probability of having at least one pair of elements promoting the intermetallic compound formation increases drastically.

The advanced data exploration was performed in the aftermath of preliminary visualization. As highlighted in Section 2.1, the data consist of 41 composition features, 6 properties features and 1 feature related to phase information. The data with a total of 48 features (input + output) are high-dimensional data, and the presence of an inter-correlation among these features makes it more challenging to obtain consolidated information from such type of data.

In order to perform optimum data exploration, dimensionality reduction is the solution, and principal component analysis (PCA) can help in attaining this purpose. In this study, the reduction in dimension of the data (consisting of 48 features) was achieved by representing these features as two orthogonal components (principal component 1 and principal component 2). The two principal components are uncorrelated to each other, and this helps in the easy visualization of the dataset. With these two principal components as the axes of the graphical plots, the distribution of the data points is presented in Figure 3. It is significant to note that the IM (red-colored dots) are clustered more in the central space of the graph, whereas the Not IM (green-colored dots) class are present dominantly at the extreme peripheries. The significance of the data features to the calculation of PC1 and PC2 was also assessed with the help of eigenvectors. Depending upon the sample alloys, certain elements have more weightage over other elements. The details about the relative importance of composition features over the calculation of principal components are considered under the scope of future study. With the help of the absolute values of the eigenvectors, the top three thermodynamic, physical and chemical properties that largely influence the principal components were sorted out for each PC1 and PC2. In context of PC1, VEC (−0.37), $\Delta\chi$ (0.25) and ΔS_m (−0.11) were outlined as the most significant property features. On the other hand, ΔS_m (−0.15), δ (0.14) and ΔH_m (−0.1) were the property features that are the three most important for the computation of PC2. The $\pm$ sign in front of the values for coefficients shows the direction of the vectors.

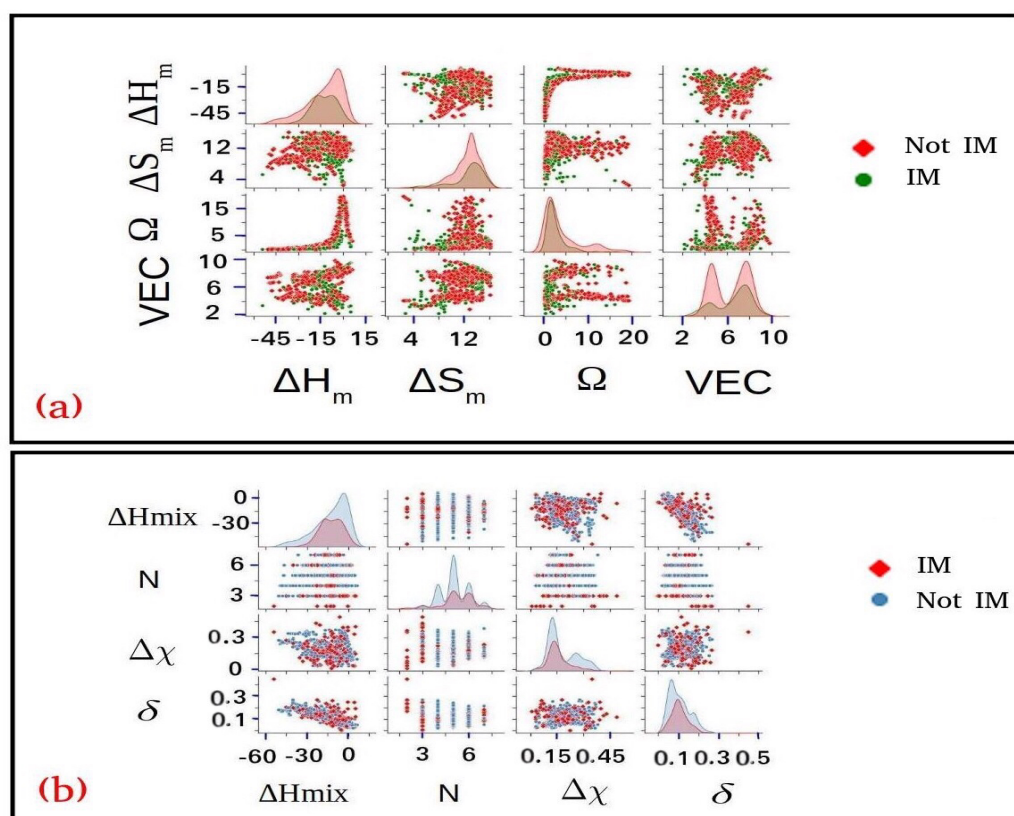


Figure 2. Pairplots for visualization of the different features of the collected dataset. The fingerprints—namely, mixing enthalpy (ΔH_{mix} or ΔH_m), mixing entropy, Ω and VEC—are plotted in (a), whereas the scatter distributions of mixing enthalpy, number of components (N), electronegativity and atomic size difference are compared in (b). The scattered data are classified as IM and Not IM. It is to be noted that IM represents that the system can have either an intermetallic compound (IMC) phase or can be the IMC in combination with other phases such as AM, SS, FCC, and BCC. Not IMC simply means that a given MPEA does not have an IMC phase.

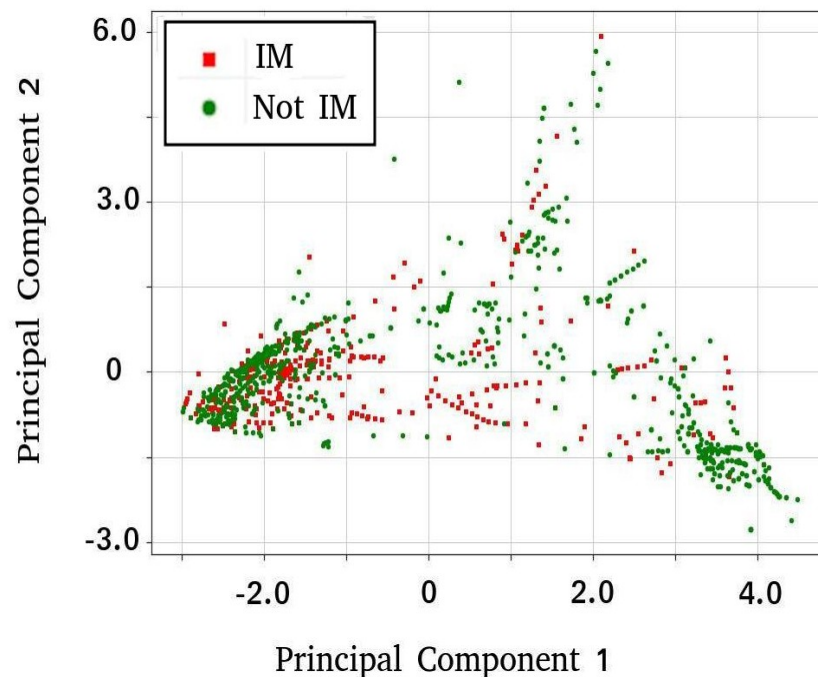


Figure 3. The dataset distribution of the IM and Not IM labels has been visualized using the two principal components (PC 1 and PC 2). The alloys consisting of intermetallic phase at full or partial proportion (IM) are represented by red-colored points, whereas the alloys devoid of intermetallic phase (Not IM) are shown as green-colored dots.

2.3. Model Construction and Training

With the overview of the base data of the high-entropy alloys and its features, it has already been understood that the materials consisting of the IM phase can be categorized separately with those materials not having these phases. Artificial neural network (ANN) has been chosen in this study as the machine learning (ML) model to perform this classification task. The neural network algorithm was designed to consist of four hidden layers (HL 1, HL 2, HL 3, and HL 4). Corresponding to the total number of composition and property features, the input layer was assigned 47 different input features. The first hidden layer (HL1) was assigned with the number of neurons equal to the number of input features (No. of neurons in HL 1 = 47). HL 2, HL 3 and HL 3 were designed to be identical in terms of their size, and each of them were allocated 10 neurons. The output layer consisted of a Sigmoid activation function. The task of classification at the output will be to separate MPEAs with an IM phase or without an IM phase, and an encoding method (IM = 1 and Not IM = 0) was utilized in the dataset (described in Section 2.1) to enable it to be machine readable by the neural network algorithm.

The neural network model of Figure 1 was compiled in TensorFlow software [25] using the train, test and validation ratio of 80:10:10. As shown in Table 2, several sets of models were trained upon the variation design of the hyperparameters—learning rate, activation functions, and optimizer function. The learning rate for the model was varied between constant values from 2.5×10^{-5} to 1×10^{-3} and increasing values within the same range with an even spacing of 2.5×10^{-5} . Similarly, several sets of combinations were made with the two activation functions used in hidden layers—namely ReLU (Rectified Linear Unit) and LeakyReLU (Leaky Rectified Linear Unit) functions. SGD (Stochastic Gradient Descent) and Adam optimizers [26] were selected as the two types of optimizer functions for tuning the models. Among all the models, the best-performing model (BPM) was defined as the one that yields the optimum accuracy values for both the train and validation data. The model with a constant learning rate of 9.5×10^{-4} using the Adam optimizer function and

having the ReLU function as the activation function for the first hidden layer (HL 1) and *LeakyReLU* for the remaining three (HL 2, HL 3, HL 4), was designated as the BPM, and its performance metrics (accuracy and loss metrics) will be presented in Section 3. In order to use the neural network model at its most optimized state, an early stopping feature was employed in the machine learning algorithm.

Table 2. Design variation in the three hyperparameters for sorting out the best performing model. Here, ReLU stands for rectified linear unit, LEAKY-ReLU is a modified rectified linear unit, Adam name is derived from adaptive moment estimation, and SGD stands for stochastic gradient decent.

Varied Hyperparameters	Range of Values/Types/Design
Activation functions in HL	Varied combinations of ReLU and LeakyReLU
Optimizers	Adam, SGD
Learning rate	range (2.5×10^{-5} – 1.0×10^{-3} : constant or stepped (step size = 2.5×10^{-5}))

3. Results and Discussion

The results and discussion section includes the result of model training (stage I) and development of application based upon the prediction model (stage II).

3.1. Performance Metrics of the BPM during Training and Validation

The training and validation accuracies for the BPM neural network model are presented in Figure 4a. As shown in the figure, both the training and validation accuracies rise with the increase in the number of epochs. It is noticeable that the validation accuracy is higher than the training accuracy during the initial model run (epoch 0 to epoch 3). On average, the rate of increase in training accuracy is larger than that of validation accuracy. At epochs in the range 4–6, the two accuracies (training accuracy = 0.9010 and validation accuracy = 0.9048) are nearly equal. At and beyond epoch 7, the training accuracy surpasses the validation accuracy. At epoch 10, the validation accuracy is 0.9096, whereas the training accuracy is 0.9197. With both the training and validation accuracies reaching above 90%, it can be inferred that the neural network model is cross-validated, and so, now, it can be readily deployed for MPEAs classification as well as the prediction of an intermetallic phase beyond the training data.

The two curves in Figure 4b represent the variations of training and validation loss of the model with the epochs. For a binary classification problem, the binary cross entropy (BCE) function is one of the commonly employed cost functions [27]. Since the IMC detection task is a binary classification problem, this work utilizes the BCE function in the model training and validation tasks. By the end of epoch 10, the training loss and validation loss for the model are 4.61×10^{-2} and 5.65×10^{-2} , respectively.

3.2. GUI Interface on Top of Prediction Model

The best performing model was selected as the prediction model. A graphical user interface (GUI) application “Intermetallics Compound at High Entropy Alloys” (Acronym = IMCATHEA) was then built over this prediction model. The application is licensed under GNU General Public License (GNU GPL), and the source code of the library can be found at <https://github.com/subediupadesh/IMCATHEA>.

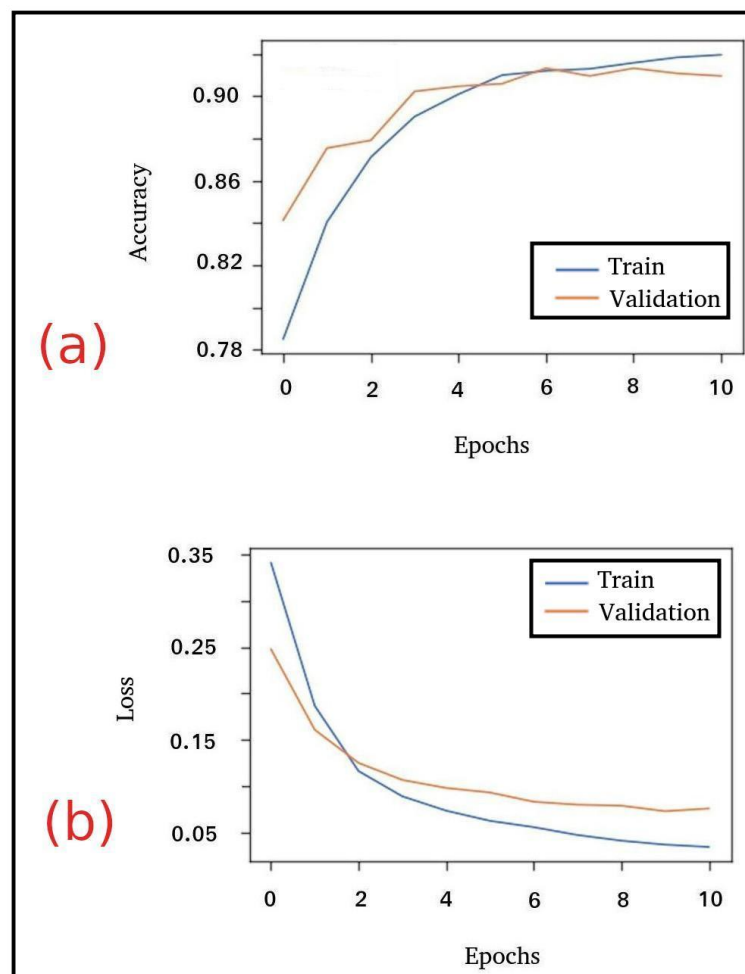


Figure 4. Accuracy metrics of the best performing model is shown in (a). The curves corresponding to the BCE loss function of the same model is presented in (b). The number of epochs is controlled by the early stopping switch activated during the neural network training procedure.

The layout of the IMCATHEA GUI platform is shown in Figure 5. As illustrated in the figure, the input part is made available at the left part of the GUI. Upon supplying the number of components in the high-entropy alloy (HEA) or multi-principal element alloys (MPEAs), the user will be able to select the constituent elements of the alloys from the dropdown list. Then, by providing the numerical values related to the composition, the user can click the “Predict IMC” button. Upon this input of the constituent elements’ composition information, the in-built “Automatic Featurizer” library placed along with the prediction model will automatically compute the total of 41 composition features and six property features. The machine-computed 47 features will be automatically supplied as input to the prediction model, which will then output the classification result. For example, in the figure, the task is to know whether the CrNbTiZr MPEA is composed of an intermetallic compound (IMC) phase. The number of components is selected as four, and the constituent elements Cr, Nb, Ti and Zr are selected from the in-built dropdown list. As the molar fraction of each of the four components is equal to 0.25, the value of 1 is supplied for every element, and the button “Predict IMC” is clicked. The automatic featurization library of the software app then automatically computes the molar fraction of constituent elements, and it assigns a 0 value of mole fraction to the remaining absentee 37 elements. With the calculation of the mole fraction of the 41 elements (composition procedures), the Automatic Featurizer then proceeds to the computation of δ , enthalpy of mixing, entropy of mixing, melting temperature (required for Ω calculation), Ω , electronegativity difference and VEC. These values are displayed on the dashboard. With all of the composition and

property features for the CrNbTiZr MPEA being made available as input, the model then is enabled to make the predictions, and for this alloy, it has been estimated by the model that the MPEA system has an IMC phase in it. Newer predictions for different alloys can be enabled by pressing the “RESTART” button.

The incorporation of Automatic Featurization capability in IMCATHEA software is viewed as an initiative to enhance the user-friendliness of the toolkit. An user can utilize the toolkit for detection of the IMC phase in an MPEA of interest without having a need to have knowledge and information about the detailed materials properties of the alloy. At the current stage of development, IMCATHEA has been tested and validated for MPEAs with $N = 3\text{--}7$. Thus, the GUI can be used to estimate the presence/absence of the IM phase in all MEAs ($N = 3, 4$). With the maximum limit of N designated as seven, this GUI therefore can be deployed for the detection of IM only for HEAs composed of five, six, or seven elements. Although the dropdown list of the GUI offers the possibility to choose upto 41 elements and thus compute properties/features for HEAs with N up to 41, the discussion on accuracy in relation to the prediction tasks for any such HEAs with N larger than seven is considered beyond the scope of present work. Upon collecting more data for HEAs with N larger than seven, future works will be focused on detecting IMC phases in MPEA with N in the range 8–10.

The screenshot shows the IMCATHEA GUI interface. At the top, there is a title bar 'IMCATHEA' and standard window controls. Below the title bar, there is a 'No. of Components' dropdown set to '4'. To the right are two buttons: 'Predict IMC' and 'RESTART'. Below these, there is a table with two columns: 'Elements' and 'Composition'. The 'Elements' column has dropdown menus for Cr, Nb, Ti, and Zr. The 'Composition' column has input fields, all set to '1'. To the right of the table is a section titled 'Properties' with a list of material properties and their corresponding values. Below this, there is a section titled 'MPEA ENTERED' with a text box containing 'CrNbTiZr'. Below that, there is a section titled 'IMC Presence or Absence' with a large red box containing the text 'System has IMC'.

Properties	Values
Atomic Size Difference (δ)	8.66 %
Mixing of Enthalpy (ΔH_{mix})	-5.18 kJ/mol
Mixing of Entropy (ΔS_{mix})	11.53 J/K/mol
Melting Temp (T_m)	2249.75 K
Omega (Ω) = $T_m \cdot \Delta S_{mix} / \Delta H_{mix} $	5.01
Electronegativity ($\Delta \chi$)	0.12
VEC	4.75

MPEA ENTERED
CrNbTiZr

IMC Presence or Absence
System has IMC

Figure 5. The layout design of IMCATHEA GUI interface.

4. Conclusions

The detection of intermetallic (IM) phase in the MPEA system is very important for the design and development of emergent materials. Considering the potential applications of intermetallic compounds and high-entropy alloys in a multidisciplinary field, it is required to develop software (on materials science) that is simpler to use on one hand and that has robust performance on the other hand. In the era of data science, the blend of simplicity and robustness becomes even more relevant. The current work focused on developing a GUI library with automatic feature calculators for materials, which is inspired by this intuition. The following conclusions have been derived from the present work:

- A Python script was written to enable the automatic extraction of composition and property features from the name of the MPEAs in the dataset of 1301 MPEAs. The phase feature was simplified into two main labels—IM (presence of intermetallic phase) and Not IM (absence of intermetallic phase).
- Pairplots and principal component analysis were utilized for the data visualization. It has been observed that intermetallic phase data were more abundant in the materials

with either components less than three (binary intermetallics and medium-entropy alloys) or high-entropy alloys with $N > 6$. In the multi-principal element alloys with N in the range four to six, the intermetallic phase was less prevalent. The significance of property features or variables in the computation of principal components was quantitatively assessed during PCA.

- iii An artificial neural network was trained upon the datasets of MPEAs. A model using ReLU and LeakyReLU activation functions at hidden layers, using the Adam optimizer function, and learning rate of 9.5×10^{-4} exhibited the training accuracy and validation accuracy of 0.9197 and 0.9096, respectively, at epoch 10. This properly cross-validated model was then chosen as the prediction model.
- iv In order to ensure the easy usage of the libraries, a GUI software named “IM-CATHEA” was built upon the automatic featurization library (preprocessor) blended together with the prediction machine learning model. The availability of an Automatic Featurizer enables the successful IM phase prediction in the alloy without having the need to manually supply the input features to the prediction model.

Author Contributions: Conceptualization, U.S. and A.K.; Methodology, U.S. and A.K.; Software, U.S., Y.A.C. and A.K.; Validation, U.S., Y.A.C., P.B.M., K.G. and A.K.; Formal Analysis, U.S. and Y.A.C.; Investigation, U.S.; Resources, A.K.; Data Curation, U.S., P.B.M. and K.G.; Writing—Original Draft Preparation, U.S.; Writing—Review and Editing, Y.A.C., P.B.M., K.G. and A.K.; Visualization, U.S., Y.A.C., P.B.M. and A.K.; Supervision, K.G. and A.K.; Project Administration, A.K.; Funding Acquisition, A.K. All authors have read and agreed to the published version of the manuscript.

Funding: This work was financially supported by the National Science Centre, Poland (SONATA BIS grant, Grant Number: 2021/42/E/ST5/00339).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The codes and minimal data required to construct a minimal working example of the IMCATHEA GUI toolkit can be found at <https://github.com/subediupadesh/IMCATHEA>. The codes illustrating the automatic featurization procedure for MPEA datasets are available at <https://github.com/subediupadesh/AutomaticFeaturizerMPEA>. The complete datasets will be made available from the authors upon reasonable request.

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

Abbreviations

The following abbreviations are used in this manuscript:

AM	Amorphous phase
ANN	Artificial Neural Network
BCC	Body-Centered Cubic
BCE	Binary Cross-Entropy
BPM	Best Performing Model
FCC	Face-Centered Cubic
GUI	Graphical User Interface
HEA	High-Entropy Alloy
IM	Intermetallics
IMC	Intermetallic Compounds
LeakyReLU	Leaky Rectified Linear Unit
MEA	Medium-Entropy Alloy
ML	Machine Learning
MPEA	Multi-Principal Element Alloy
PC 1	Principal Component 1
PC 2	Principal Component 2

PCA	Principal Component Analysis
ReLU	Rectified Linear Unit
SS	Solid Solution
SSS	Simple Solid Solution
SGD	Stochastic Gradient Descent
VEC	Valence Electron Concentration

References

1. Cantor, B.; Chang, I.T.H.; Knight, P.; Vincent, A.J.B. Microstructural development in equiatomic multicomponent alloys. *Mater. Sci. Eng. A* **2004**, *375–377*, 213–218. [\[CrossRef\]](#)
2. 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, 274. [\[CrossRef\]](#)
3. Lotfi, S.; Brgoch, J. Discovering Intermetallics Through Synthesis, Computation, and Data-Driven Analysis. *Chem. Eur. J.* **2020**, *26*, 8689–8697. [\[CrossRef\]](#)
4. Miracle, D.B.; Senkov, O.N. A critical review of high entropy alloys and related concepts. *Acta Mater.* **2017**, *122*, 448–511. [\[CrossRef\]](#)
5. Garcia Filho, F.D.C.; Ritchie, R.O.; Meyers, M.A.; Monteiro, S.N. Cantor-derived medium-entropy alloys: Bridging the gap between traditional metallic and high-entropy alloys. *J. Mater. Res. Technol.* **2022**, *17*, 1868–1895. [\[CrossRef\]](#)
6. Churyumov, A.; Kazakova, A.; Churyumova, T. Modelling of the Steel High-Temperature Deformation Behaviour Using Artificial Neural Network. *Metals* **2022**, *12*, 447. doi: [\[CrossRef\]](#)
7. Honyasz, R. Modeling the chemical composition of ferritic stainless steels with the use of artificial neural networks. *Metals* **2021**, *11*, 724. [\[CrossRef\]](#)
8. Machaka, R. Machine learning-based prediction of phases in high-entropy alloys. *Comput. Mater. Sci.* **2021**, *188*, 110244. [\[CrossRef\]](#)
9. Tsai, M.H.; Tsai, R.C.; Chang, T.; Huang, W.F. Intermetallic phases in high-entropy alloys: Statistical analysis of their prevalence and structural inheritance. *Metals* **2019**, *9*, 247. [\[CrossRef\]](#)
10. Chou, T.H.; Huang, J.C.; Yang, C.H.; Lin, S.K.; Nieh, T.G. Consideration of kinetics on intermetallics formation in solid-solution high entropy alloys. *Acta Mater.* **2020**, *195*, 71–80. [\[CrossRef\]](#)
11. Subedi, U.; Kunwar, A.; Coutinho, Y.A.; Gyanwali, K. pyMPEALab Toolkit for Accelerating Phase Design in Multi-principal Element Alloys. *Met. Mater. Int.* **2022**, *28*, 269–281. [\[CrossRef\]](#)
12. Coutinho, Y.A.; Kunwar, A.; Moelans, N. Phase-field approach to simulate BCC-B2 phase separation in the Al₃CrFe₂Ni₂ medium-entropy alloy. *J. Mater. Sci.* **2022**. [\[CrossRef\]](#)
13. 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. [\[CrossRef\]](#)
14. Agarwal, A.; Prasada Rao, A.K. Artificial Intelligence Predicts Body-Centered-Cubic and Face-Centered-Cubic Phases in High-Entropy Alloys. *JOM* **2019**, *71*, 3424–3432. [\[CrossRef\]](#)
15. Couzinié, J.P.; Senkov, O.N.; Miracle, D.B.; Dirras, G. Comprehensive data compilation on the mechanical properties of refractory high-entropy alloys. *Data Brief* **2018**, *21*, 1622–1641. [\[CrossRef\]](#)
16. 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. [\[CrossRef\]](#) [\[PubMed\]](#)
17. Guo, S.; Liu, C.T. Phase stability in high entropy alloys: Formation of solid-solution phase or amorphous phase. *Prog. Nat. Sci. Mater. Int.* **2011**, *21*, 433–446. [\[CrossRef\]](#)
18. Wang, Z.; Guo, S.; Liu, C.T. Phase Selection in High-Entropy Alloys: From Nonequilibrium to Equilibrium. *JOM* **2014**, *66*, 1966–1972. [\[CrossRef\]](#)
19. Zhou, Z.; Zhou, Y.; He, Q.; Ding, Z.; Li, F.; Yang, Y. Machine learning guided appraisal and exploration of phase design for high entropy alloys. *Npj Comput. Mater.* **2019**, *5*, 128. [\[CrossRef\]](#)
20. Ong, S.P.; Richards, W.D.; Jain, A.; Hautier, G.; Kocher, M.; Cholia, S.; Gunter, D.; Chevrier, V.L.; Persson, K.A.; Ceder, G. Python Materials Genomics (pymatgen): A robust, open-source python library for materials analysis. *Comput. Mater. Sci.* **2013**, *68*, 314–319. [\[CrossRef\]](#)
21. Ong, S.P.; Cholia, S.; Jain, A.; Brafman, M.; Gunter, D.; Ceder, G.; Persson, K.A. The Materials Application Programming Interface (API): A simple, flexible and efficient API for materials data based on REpresentational State Transfer (REST) principles. *Comput. Mater. Sci.* **2015**, *97*, 209–215. [\[CrossRef\]](#)
22. Jain, A.; Ong, S.P.; Hautier, G.; Chen, W.; Richards, W.D.; Dacek, S.; Cholia, S.; Gunter, D.; Skinner, D.; Ceder, G.; et al. Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. *APL Mater.* **2013**, *1*, 11002. [\[CrossRef\]](#)
23. Yang, X.; Zhang, Y. Prediction of high-entropy stabilized solid-solution in multi-component alloys. *Mater. Chem. Phys.* **2012**, *132*, 233–238. [\[CrossRef\]](#)
24. Senkov, O.N.; Miller, J.D.; Miracle, D.B.; Woodward, C. Accelerated exploration of multi-principal element alloys with solid solution phases. *Nat. Commun.* **2015**, *6*, 6529. [\[CrossRef\]](#) [\[PubMed\]](#)

-
25. Abadi, M.; Agarwal, A.; Barham, P.; Brevdo, E.; Chen, Z.; Citro, C.; Corrado, G.S.; Davis, A.; Dean, J.; Devin, M.; et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. *arXiv* **2016**, arXiv:1603.04467.
 26. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv* **2014**, arXiv:1412.6980.
 27. Kunwar, A.; Malla, P.B.; Sun, J.; Qu, L.; Ma, H. Convolutional neural network model for synchrotron radiation imaging datasets to automatically detect interfacial microstructure: An in situ process monitoring tool during solar PV ribbon fabrication. *Sol. Energy* **2021**, *224*, 230–244. [[CrossRef](#)]