



Data Article

Machine learning-based prediction of phases in high-entropy alloys: A data article [☆]



Ronald Machaka ^{a,b,*}, Glenda T. Motsi ^a, Lerato M. Raganya ^{a,b},
Precious M. Radingoana ^a, Silethelwe Chikosha ^a

^a Advanced Materials Manufacturing, Manufacturing Cluster, Council for Scientific and Industrial Research, P.O. Box 395, Pretoria, 0001, South Africa

^b School of Mining, Metallurgy and Chemical Engineering, University of Johannesburg, Johannesburg, South Africa

ARTICLE INFO

Article history:

Received 10 May 2021

Revised 25 August 2021

Accepted 27 August 2021

Available online 16 September 2021

Dataset link: [Dataset for High-Entropy Alloys Phases \(Reference data\)](#)

Keywords:

High entropy alloys

HEA microstructures

Phases

Machine learning

Deep learning

Material informatics

ABSTRACT

A systematic framework for choosing the most determinant combination of predictor features and solving the multiclass phase classification problem associated with high-entropy alloy (HEA) was recently proposed [1]. The data associated with that research paper, titled “Machine learning-based prediction of phases in high-entropy alloys”, is presented in this data article. This dataset is a systematic documentation and comprehensive survey of experimentally reported HEA microstructures. It contains microstructural phase experimental observations and metallurgy-specific features as introduced and reported in peer-reviewed research articles. The dataset is provided with this article as a supplementary file. Since the dataset was collected from experimental peer-reviewed articles, these data can provide insights into the microstructural characteristics of HEAs, can be used to improve the optimization HEA phases, and have an important role in machine learning, material informatics, as well as in other fields.

[☆] Given his role as Editorial Board Member of this journal, Ronald Machaka had no involvement in the peer-review of this article and has no access to information regarding its peer-review.

DOI of original article: [10.1016/j.commatsci.2020.110244](https://doi.org/10.1016/j.commatsci.2020.110244)

* Corresponding author.

E-mail address: rmachaka@csir.co.za (R. Machaka).

Social media: [@RonaldMachaka](#) (R. Machaka)

© 2021 The Authors. Published by Elsevier Inc.
This is an open access article under the CC BY license
(<http://creativecommons.org/licenses/by/4.0/>)

Specifications Table

Subject	Materials Science
Specific subject area	High Entropy Alloys (HEA)
Type of data	Text file, tables and figures
How data were acquired	The dataset was gathered from experimental observations reported in peer-reviewed research articles. Each entry of the 1362 dataset entries is therefore an experimental observation as reported in the literature.
Data format	Raw, tabulated and plotted
Parameters for data collection	The following wide range of parameters were collected: <ol style="list-style-type: none"> i. Identification of the multi-component alloy, ii. the number of elements making up the alloy system, iii. the valence electron concentration, iv. the atomic size difference, v. the difference in Pauling negativities, vi. the enthalpy of mixing for a alloy system, vii. the entropy of mixing of a multi-component alloy system, viii. how the alloy was synthesis and processing conditions (post-process working and heat treatment), ix. the experimentally observed and reported, and x. the bibliographical references against each entry.
Description of data collection	Secondary data (i.e. composition-specific features, alloy processing and post-processing parameters, and the resulting phases) were collected. Some typical empirical HEA design parameters were calculated using well-known approaches. Data was processed using Excel and R, a language and environment for statistical computing, for purposes of visualization and data analysis.
Data source location	Unit: <i>Advanced Materials Engineering, Manufacturing Cluster</i> Institution: <i>Council for Scientific and Industrial Research (CSIR)</i> Address: <i>Scientia Campus, 627 Meiring Naude Rd, Brummeria, Pretoria 0185</i> Country: <i>South Africa</i> GPS: <i>25°44'35.2"S 28°16'52.3"E</i>
Data accessibility	Data is provided with this article
Related research articles	Research articles associated with research data presented are. <p>[1] Ronald Machaka, <i>Machine learning-based prediction of phases in high-entropy alloys</i>, <i>Computational Materials Science</i>, Vol. 188, 2021,11024 https://doi.org/10.1016/j.commatsci.2020.110244</p>

Value of the Data

- This dataset documents synthesis routes, processing conditions (post-process working and heat treatment), and the resulting microstructural observations which can be valuable for researchers in the field of Materials Science in the development of experiments.
- This dataset specifically contains experimentally reported HEA microstructures which provides enough observations to train and test machine learning and deep learning algorithms.
- Different machine learning and material informatics computational methods can be applied to this dataset in order to extract insights and trends not immediately available from individual studies thereby advancing the real-world applications of these alloys.

1. Data Description

The data presented in this article are related to the research articles [1,2]. The data is presented in the supplementary data file.

1.1. The dataset

This dataset is a systematic documentation and comprehensive survey of experimentally reported HEA microstructures. The dataset was constructed from microstructural observations reported in peer-reviewed experimental HEA research articles; it is built upon datasets prior published by Miracle et al. [3], Couzinié et al. [4], and Ye et al. [5]. The dataset presents metallurgy-specific features and microstructural phases experimentally observed.

The dataset, provided with this article as supplementary material, has seventeen columns and 1422 entries.

- Columns 1 and 2 correspond to the identification of the dataset entry and the multi-component alloy system – *Alloy_ID* and *Alloy*, respectively. Composition-specific features can further be developed from the *Alloy* specification
- Columns 3 to 27 correspond to the elemental compositions of the multi-component alloy while Column 28 corresponds to the number of elements making up the multi-component alloy system (*Num_of_Elem*).
- Columns 29 to 37 correspond to some typical empirical HEA design parameters [6,7] such as the density estimate (*Density_calc*), the enthalpy of mixing for a multi-component alloy system (dH_{mix}), entropy of mixing of a multi-component alloy system (dS_{mix}), melting temperature estimate (T_m), valence electron concentration (*VEC*), atomic size difference (δ), and difference in Pauling negativities (χ). While this article is limited to these typical parameters for succinctness, others can also be developed from the *Alloy* specifications [7–13].
- Columns 38 to 45 correspond to metallurgy-informed alloy processing and post-processing parameters indicating that: how the alloy was synthesised (*Synthesis_Route*); hot- or cold-worked (*Hot-Cold_Working*); undergone homogenization processing at temperature (*Homogenization_Temp*) and time (*Homogenization_Time*); undergone annealing processing at temperature (*Annealing_Temp*) and time (*Annealing_Time*); and/or undergone some quenching processing (*Quenching*).
- The values tabulated in Column 50 represent the microstructure or phases corresponding to the multi-component alloy as reported in the literature. Each dataset entry clearly documents the microstructural observations reported for the respective alloy as either single solid solution phases (face-centered cubic – designated *FCC* and body-centered cubic – designated *BCC*), dual-phase solid solutions (designated *FCC+BCC*), other intermetallic, laves, and martensitic phases (designated *Im*).
- The last column, Column 52 provides the bibliographical references against the entries, for transparency.

Table 1 gives a summarised description of these features.

2. Experimental Design and Methods

The dataset is built upon datasets prior published by Miracle et al. [3], Couzinié et al. [4], and Ye et al. [5]. It is constructed from HEA microstructural observations reported in experimental peer-reviewed research articles.

Selected HEAs reported in the literature through the end of August 2020 make up the dataset presented in this article. The as-constructed dataset has at least 1362 multicomponent alloys. After removing entries missing some data and eliminating glassy, amorphous, and hexagonal close-packed phases, the as-constructed dataset is reduced to 1362 multicomponent alloys. Supplementary materials documents accompanying this article contain the full list of alloys and references.

The equations used to estimate each of the selected empirical design parameters are summarised in Table 1. Fig. 1 shows the distribution of phase classes in the dataset (the atomic size difference, δ as a function of *VEC*) after Refs [7,14]. The visualization also shows that solid solution phases are particularly more sensitive to empirical parameters than IM phases [1,2].

Table 1

Descriptions of the empirical and metallurgy-specific features cites ML-based studies that attempted predicting HEA phases therefrom.

Symbol	Description of Feature	References
Num_of_Elem	Number of elements in a multi-component alloy system	
δ	<p>A parameter describing the atomic size mismatch or difference in a multi-component alloy system</p> $\delta = \sqrt{\sum_{i=1}^n c_i \cdot \left(1 - \frac{r_i}{\bar{r}}\right)^2}$ <p>where c_i and r_i is the atomic percentage and atomic radius of the i^{th} component and \bar{r} is the average atomic radius of the components of the alloy, respectively.</p>	[7–13]
X	<p>Pauling negativities mismatch for multi-component alloy system</p> $X = \sqrt{\sum_{i=1}^n c_i \cdot (\chi_i - \bar{\chi})^2}$ <p>where c_i and χ_i is the atomic percentage and Pauling electronegativity of the i^{th} component and $\bar{\chi}$ is the mean value of electronegativity for a multi-component alloy system, respectively.</p>	[7–13]
VEC	<p>The valence electronic concentration of a multi-component alloy system calculated on the basis of the rule of mixtures approach</p> $VEC = \sum_{i=1}^n c_i \cdot VEC_i$ <p>where c_i and VEC_i are the atomic percentage and the valence electron concentration of the i^{th} component, respectively.</p>	[7–13]
ΔS_{mix}	<p>The entropy of mixing of a multi-component alloy system calculated as follows</p> $\Delta S_{mix} = -R \cdot \sum_{i=1}^n c_i \cdot \ln c_i$ <p>where R ($= 8.314 \text{ JK}^{-1} \text{ mol}^{-1}$) is the universal gas constant and c_i is the atomic percentage of the i^{th}, component.</p>	[8–12]
ΔH_{mix}	<p>Enthalpy of mixing for a multi-component alloy system</p> $\Delta H_{mix} = \sum_{i=1, i \neq j}^n 4 c_i c_j \Delta H_{mix}^{ij}$ <p>where c_i and c_j is the atomic percentages of the i^{th} and j^{th} components, respectively. ΔH_{mix}^{ij} is mixing enthalpy of binary liquid alloys accessible from conventional tables prepared by Takeuchi and Inoue [15] based on Miedema's atomistic model [16].</p>	[8–13]
Synthesis_Route	A categorical feature indicating that the alloy was synthesised via vacuum melted (AC), powder metallurgy (PM), or otherwise	
Hot-Cold_Working	A categorical feature indicating that the alloy has been subjected to a cold/hot working treatment. (nan, CW – cold worked, HW – hot worked, HIP – hot-isostatically pressed)	
Homog_Temp	A feature indicating the temperature at which the alloy was subjected to a homogenization treatment (in °C)	
Homog_Time	A feature indicating the duration for which the alloy was subjected to a homogenization treatment (in minutes).	
Annealing_Temp	A feature indicating the temperature at which the alloy was subjected to a annealing treatment (in °C)	
Annealing_Time	A feature indicating the duration for which the alloy was subjected to a annealing treatment (in minutes).	
Quench_Proc	Categorical feature indicating that the alloy has been subjected to an quenching heat treatment	
Microstructure	Experimentally observed microstructure(s), namely BCC_SS (body-centered cubic solid solutions), FCC_SS (face-centered cubic single solid solutions) FCC+BCC_SS (dual-phase solid solutions) and Im (intermetallic, intermetallic, laves, martensitic, and other phases but excluding glassy or amorphous and hexagonal close-packed – HCP)	

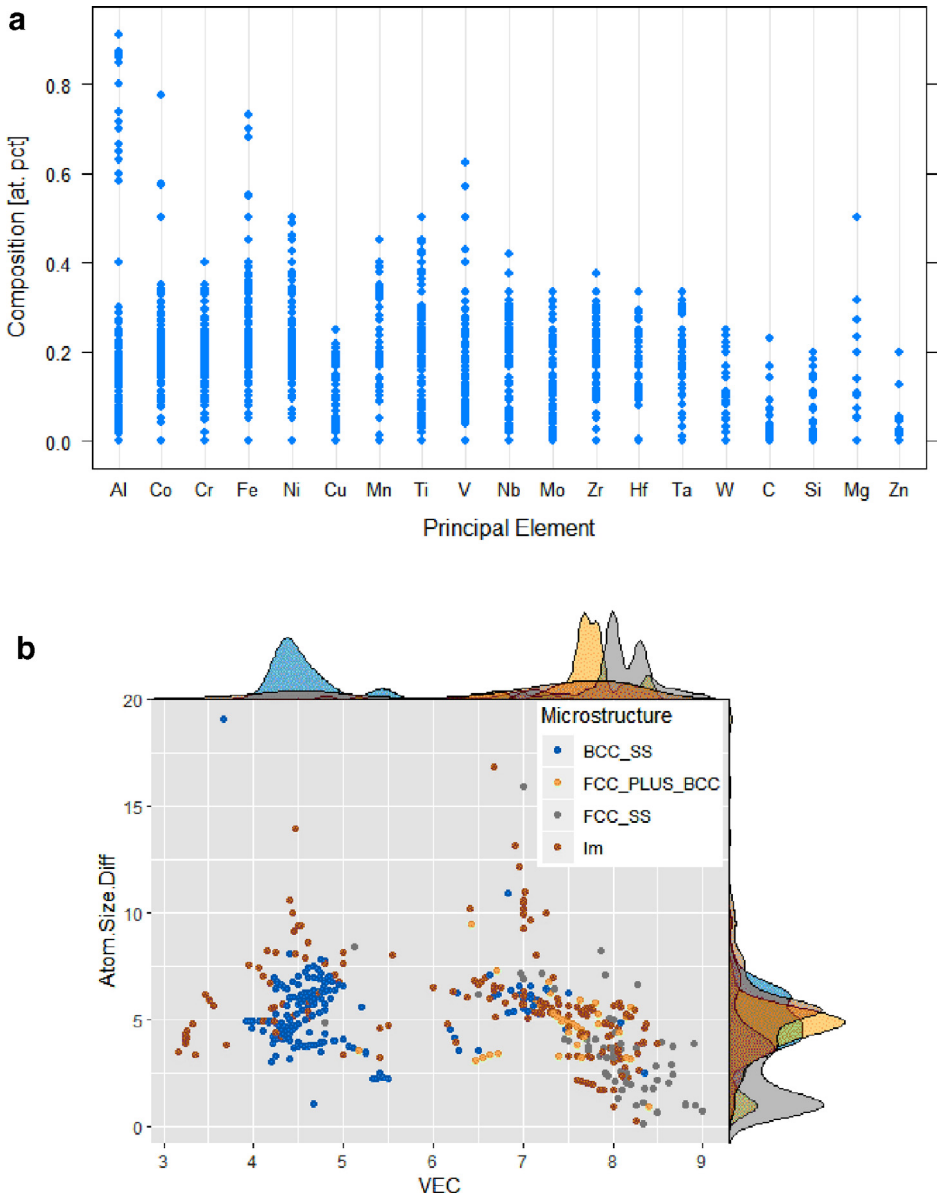


Fig. 1. Distribution visualization of the (a) HEA elements and (b) the full dataset phase distribution in the VEC- δ space.

Ethics Statement

None.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships which have, or could be perceived to have, influenced the work reported in this article.

Data availability

[Dataset for High-Entropy Alloys Phases \(Reference data\)](#) (Mendeley Data)

CRediT Author Statement

Ronald Machaka: Conceptualization, Funding acquisition, Resources, Methodology, Visualization, Supervision, Writing – original draft, Writing – review & editing; **Glenda T. Motsi:** Data curation, Investigation, Writing – original draft, Writing – review & editing; **Lerato M. Raganya:** Data curation, Investigation, Writing – original draft, Writing – review & editing; **Precious M. Radingoana:** Data curation, Investigation, Writing – original draft, Writing – review & editing; **Silethelwe Chikocha:** Data curation, Supervision, Investigation, Writing – original draft, Writing – review & editing.

Acknowledgments

The authors greatly appreciate the financial support received from the National Research Foundation through the Y-Rated Researcher Support Grant [Unique Grant No.: 120-413] and the NRF-DST Professional Development Programme (PDP) for Postdoctoral Research Fellowships Block Grant.

Supplementary Materials

Supplementary material associated with this article can be found in the online version at doi:[10.1016/j.dib.2021.107346](https://doi.org/10.1016/j.dib.2021.107346).

References

- [1] R. Machaka, Machine learning-based prediction of phases in high-entropy alloys, *Comput. Mater. Sci.* (2020).
- [2] R. Machaka, Machine learning based prediction of phases in high-entropy alloys, *SSRN Electron. J.* (2020), doi:[10.2139/ssrn.3530328](https://doi.org/10.2139/ssrn.3530328).
- [3] S. Gorsse, M.H. Nguyen, O.N. Senkov, D.B. Miracle, Database on the mechanical properties of high entropy alloys and complex concentrated alloys, *Data Br.* 21 (2018) 2664–2678, doi:[10.1016/j.dib.2018.11.111](https://doi.org/10.1016/j.dib.2018.11.111).
- [4] J.-P. Couzinié, O.N. Senkov, D.B. Miracle, G. Dirras, Comprehensive data compilation on the mechanical properties of refractory high-entropy alloys, *Data Br.* 21 (2018) 1622–1641, doi:[10.1016/j.dib.2018.10.071](https://doi.org/10.1016/j.dib.2018.10.071).
- [5] Y.F. Ye, Q. Wang, J. Lu, C.T. Liu, Y. Yang, High-entropy alloy: challenges and prospects, *Mater. Today.* 19 (2016) 349–362, doi:[10.1016/j.mattod.2015.11.026](https://doi.org/10.1016/j.mattod.2015.11.026).
- [6] D.J.M. King, S.C. Middleburgh, A.G. McGregor, M.B. Cortie, Predicting the formation and stability of single phase high-entropy alloys, *Acta Mater.* 104 (2016) 172–179, doi:[10.1016/j.actamat.2015.11.040](https://doi.org/10.1016/j.actamat.2015.11.040).
- [7] Z. Tang, S. Zhang, R. Cai, Q. Zhou, H. Wang, Designing high entropy alloys with dual fcc and bcc solid-solution phases: structures and mechanical properties, *Metall. Mater. Trans. A.* 50 (2019) 1888–1901, doi:[10.1007/s11661-019-05131-1](https://doi.org/10.1007/s11661-019-05131-1).
- [8] Y. Zhang, C. Wen, C. Wang, S. Antonov, D. Xue, Y. Bai, Y. Su, Phase prediction in high entropy alloys with a rational selection of materials descriptors and machine learning models, *Acta Mater.* 185 (2020) 528–539, doi:[10.1016/j.actamat.2019.11.067](https://doi.org/10.1016/j.actamat.2019.11.067).
- [9] N. Islam, W. Huang, H.L. Zhuang, Machine learning for phase selection in multi-principal element alloys, *Comput. Mater. Sci.* 150 (2018) 230–235, doi:[10.1016/j.commatsci.2018.04.003](https://doi.org/10.1016/j.commatsci.2018.04.003).
- [10] Z. Zhou, Y. Zhou, Q. He, Z. Ding, F. Li, Y. Yang, Machine learning guided appraisal and exploration of phase design for high entropy alloys, *Npj Comput. Mater.* 5 (2019) 128, doi:[10.1038/s41524-019-0265-1](https://doi.org/10.1038/s41524-019-0265-1).

- [11] H. Zhang, G. Zhou, S. Li, X. Fan, Z. Guo, T. Xu, Y. Xu, X. Chen, D. Dai, Q. Qian, Application of fuzzy learning in the research of binary alloys: revisit and validation, *Comput. Mater. Sci.* 172 (2020) 109350, doi:[10.1016/j.commatsci.2019.109350](https://doi.org/10.1016/j.commatsci.2019.109350).
- [12] S. Li, H. Zhang, D. Dai, G. Ding, X. Wei, Y. Guo, Study on the factors affecting solid solubility in binary alloys: an exploration by Machine Learning, *J. Alloys Compd.* 782 (2019) 110–118, doi:[10.1016/j.jallcom.2018.12.136](https://doi.org/10.1016/j.jallcom.2018.12.136).
- [13] N. Qu, Y. Chen, Z. Lai, Y. Liu, J. Zhu, The phase selection via machine learning in high entropy alloys, *Procedia Manuf.* 37 (2019) 299–305, doi:[10.1016/j.promfg.2019.12.051](https://doi.org/10.1016/j.promfg.2019.12.051).
- [14] W. Huang, P. Martin, H.L. Zhuang, Machine-learning phase prediction of high-entropy alloys, *Acta Mater.* 169 (2019) 225–236, doi:[10.1016/j.actamat.2019.03.012](https://doi.org/10.1016/j.actamat.2019.03.012).
- [15] A. Takeuchi, A. Inoue, Classification of bulk metallic glasses by atomic size difference, heat of mixing and period of constituent elements and its application to characterization of the main alloying element, *Mater. Trans.* 46 (2005) 2817–2829, doi:[10.2320/matertrans.46.2817](https://doi.org/10.2320/matertrans.46.2817).
- [16] F.R. De Boer, R. Boom, W.C.M. Mattens, A.R. Miedema, *Cohesion in Metals*, North-Holland, Netherlands, 1988.