



Machine learning-enabled framework for the prediction of mechanical properties in new high entropy alloys

Amit Singh Bundela, M.R. Rahul*

Department of Fuel Minerals and Metallurgical Engineering Indian Institute of Technology (ISM) Dhanbad, Jharkhand 826004, India



ARTICLE INFO

Article history:

Received 30 January 2022

Received in revised form 3 March 2022

Accepted 12 March 2022

Available online 15 March 2022

Keywords:

Microhardness

High entropy alloys

Feature selection

Machine learning

Principal component analysis

Materials informatics

ABSTRACT

Prediction of properties of new compositions will accelerate the material design and development. The current study uses a machine learning framework to predict the microhardness of high entropy alloys. Several feature selection algorithms are used to identify the essential material descriptors. The stability selection algorithm gives optimum material descriptors for the current dataset for the microhardness prediction. Eight different machine learning algorithms are trained and tested for microhardness prediction. The accuracy of prediction improved by reducing the higher-dimensional data to lower dimensions using principal component analysis. The current study shows the testing R^2 score of more than 0.89 for XGBoost, Random forest, and Bagging regressor algorithms. Experimental data confirms the applicability of various trained algorithms for property prediction, and for the current study, ANN shows better performance for the new experimental data.

© 2022 Elsevier B.V. All rights reserved.

1. Introduction

Alloy design with the required properties for a targeted application is challenging. The large compositional space available to explore is promising for advanced applications. The development of multicomponent alloys with equiatomic or nearly equiatomic compositions is getting wider attention due to the remarkable properties it exhibits [1–3]. The high entropy alloy (HEA) with better strength ductility combination and mechanical properties are reported for room temperature as well as cryogenic applications [2,4–6]. Several reports show the possible composition of high entropy alloys (HEAs) for high-temperature applications [7–10]. HEAs have a wide possible composition space for a particular property, and many experimental trials are required to identify the optimum composition domain [11]. The application of machine learning (ML) algorithms attracted many researchers to identify new compositions and establish their properties [12–14]. Recent reports show the next-generation research will depend on the application of ML algorithms [15]. The search for high entropy alloys with exceptional properties can be accelerated with the application of ML [16,17]. Yegi et al. [18] reported the application of multiple ML classification algorithms to identify multi-phase formation in high entropy alloys. Wen et al. [19] reported the ability to come up with composition with enhanced hardness using a

machine learning assisted framework. Zhang et al. [20] reported the application of the genetic algorithm to identify the suitable material descriptor and algorithm for phase formation in HEAs. Zhang et al. [21] reported improvement in accuracy for phase prediction using Support Vector Machine (SVM) with Kernel Principal Component Analysis (KPCA) and feature selection algorithms.

Many researchers attempt the design of high entropy alloys by incorporating several materials descriptors [19,20,22]. Huang et al. [23] reported the application of ML with phase formation parameters and solid solution hardening theoretical parameters to develop non-equiatomic high entropy with enhanced hardness using ~85 samples data. Jaafreh et al. [24] reported the design of super hard high entropy ceramic materials using the Random Forest (RF) algorithm. Sun et al. [25] predicted the high entropy alloy with desired hardness in the TiZrNbTa system using the XGBoost algorithm with reasonable accuracy. Xiong et al. [26] reported the application of RF algorithm and feature selection to establish the critical parameter for hardness and phase prediction of complex concentrated alloys. Bhandari et al. [27] show the application of RF regressor for the prediction of mechanical properties of refractory high entropy alloys at high temperatures. Li et al. [28] reported the combined molecular dynamics and machine learning framework for the design of multicomponent alloys. Nassar et al. [29] use the compositional data alone to predict the HEAs with reasonable accuracy using ML algorithms.

* Corresponding author.

E-mail address: rahulmr@iitism.ac.in (M.R. Rahul).

From these recent reports, one can confirm that various ML algorithms can be used to predict phase formation and properties. Various researchers use multiple material descriptors, and a few also explore the application of multiple feature selection algorithms. The current study discussed the applicability of various feature selection algorithms to select the important material descriptor and applied to multiple algorithms for hardness prediction of high entropy alloys. This study also compares the predictions of the algorithm with and without kernel principal component analysis. The framework adopted for the current study can be used to predict the properties of materials with enhanced accuracy.

2. Experimental details

The samples required to validate the ML framework were prepared using the vacuum arc melting setup. For the current study, seven component Eutectic high entropy alloy was considered for the validation. The pure elements of the required composition were melted using the arc stuck between the non-consumable tungsten electrode and the copper mould. The high purity argon was purged to the chamber, and the melted button was remelted six times. The arc melted button was suction cast into a 6 mm diameter rod with 70 mm length. The sample was sectioned perpendicular to the axis of the cylindrical rod and done the standard metallographic polishing on the circular surface. The Vickers microhardness was measured with 500 gf load and a dwell time of 10 s.

3. Results and discussion

The data collected from the literature and our laboratory experiments are used to develop the database. Materials descriptors such as entropy of mixing (ΔS_{mix}), enthalpy of mixing (ΔH_{mix}), Atomic size difference (δ), Valence Electron Concentration (VEC), solid solution prediction parameter (Ω), Melting temperature (T_m) and Electronegativity difference ($\Delta \chi$) are considered for the current study. Initially, the data analysis was carried out, and feature selection was employed to identify the significant material descriptors. Various feature selection algorithms are used, and the one that gives a more adjusted R^2 fit [30] value was selected to identify the significant parameters for hardness prediction.

3.1. Data analysis

The current dataset consists of 214 data points from literature and experiments [31], out of which 171 data points were used in training, and the remaining data points were used for testing. Fig. 1a shows the heatmap plotted based on the Pearson correlation coefficient (PCC) [27]. For the hardness value (HV), the dataset shows a strong positive correlation with atomic size difference (δ) with a value of 0.53, while the highest negative correlation is observed with enthalpy of mixing (ΔH_{mix}) with a value of -0.53. Fig. 1b-e shows the hexagonal binning (Hexbin) plots of microhardness values with ΔS_{mix} , VEC, δ , and ΔH_{mix} , respectively. Hexbin plots of various parameters with hardness values confirmed most of the data points are unique, and one could note that the data is distributed in a wider range. Fig. 1(d, e) confirms that while increasing the δ value, the hardness increases, and while increasing ΔH_{mix} value, the hardness value decreases. The same is evident by the positive and negative correlation values of δ and ΔH_{mix} with hardness, respectively.

3.2. Selection of significant features using feature selection algorithms

While dealing with predictive models, the study of features becomes valuable to predict with less loss. In this current study, the

feature selection algorithms [32] such as Genetic Selection, Stability Selection, Univariate linear regression test, Mutual Information selection [33], and Pearson Correlation Coefficient analysis were explored. A meta-heuristic approach for feature selection with genetic selection is applied to understand the importance of the combination of features. Genetic selection with a Logistic regression estimator (with lbgfs solver) is considered. The number of population and number of generation was taken 200 and 5 respectively. The Stability Selection uses resampling, regularization, and boosting and gives a probability for each feature to be selected when randomly resampling from data [34,35]. Stability selection with Logistic regression having L2 penalty was selected as an estimator. The threshold of stability score was set to 0.5, and the features below the threshold value were turned down. In the Univariate Linear regression test (F_{reg}), p-value was compared for each term which tests the null hypothesis that the coefficient is equal to zero (no effect). A low p-value indicates that you can reject the null hypothesis. In other words, a feature that has a low p-value is likely to be a meaningful addition to the model because changes in the feature value are related to changes in the response variable. In Mutual Information selection, dependence between 2 variables is calculated and compared. For the Pearson correlation coefficient threshold of 0.2, both positive and negative is taken, and features were selected according to it.

Fig. 2 shows the performance of the feature selection algorithms and their comparison. Fig. 2(a) shows that for the current study based on univariate linear regression test, the T_m and Ω have less importance for the prediction. Fig. 2(b &c) confirms the less importance of Ω for the current study by using the Mutual Information selection and Genetic selection. Table 1 summarizes the results from various feature selection algorithms and selected features. Using lazy predict library, selected features from each selection technique were trained and tested on these various algorithms with default hyperparameters as per lazy predict library. From various selection techniques (Fig. 2d), selected features from stability selection show maximum adjusted R^2 score for several algorithms; on the contrary, selected features from genetic selection show minimum adjusted R^2 score. Features selected using stability selection were taken into consideration for further study.

3.3. Training and testing of machine learning algorithms

The selected features and microhardness is used for training various ML algorithms. The current dataset is trained on Artificial Neural Network (ANN), Linear regression (LR), Lasso regression, Ridge regression, Random Forest Regression (RFR), Support Vector Regression (SVR), XGBoost, and Bagging Regression (BR). Hyperparameter optimization of the algorithms was done using either the gridsearch method or by visualizing the effect of various parameters and choosing the optimum value to balance the variance and bias trade-off, such that the underfitting and overfitting is avoided. Fig. 3 shows the predictions for the testing with selected features and all features using multiple algorithms. In the case of LR, Lasso, SVR and Ridge algorithm, the feature selection is not improving the R^2 fit score. For RFR, XGBoost, ANN, and BR algorithms show a reasonable improvement in R^2 value with feature selection. One could note that the XGBoost algorithm shows a maximum R^2 fit score of 0.91 with an improvement of ~0.1 by using feature selection.

For the current study, the XGBoost algorithm with all features took a max depth of 6, n estimator value of 100, and a learning rate of 0.07. For the selected feature condition, the optimum learning rate is modified to 0.5, and other optimum parameters remain unchanged. The optimized ANN for all the features having three hidden layers with 24 neurons each gives the R^2 score of 0.582. Whereas, for

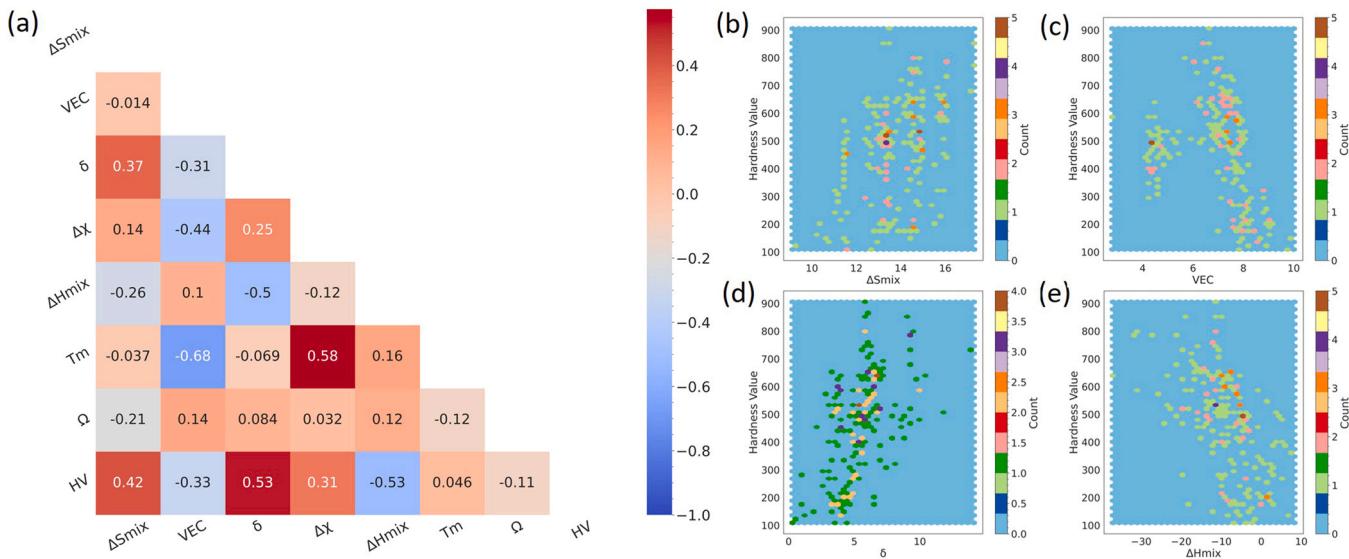


Fig. 1. a) Heatmap shows the Pearson correlation coefficient values between the features, hexagonal binning plot of b) microhardness and ΔS_{mix} , c) microhardness and VEC, d) microhardness and δ and e) microhardness and ΔH_{mix} with the number of data points represents are colour code. (For interpretation of the references to colour in this figure, the reader is referred to the web version of this article.)

five selected features after optimization and using 23 neurons each in 3 hidden layers gives an improved R^2 score of 0.651. It can be inferred that due to a lack of data points, the neural network is not unable to form extravagant connections and thus gives a low R^2 score. For improving further, the multi-dimensional data can be reduced to lower dimensions using principal component analysis.

3.4. Application of principal component analysis

The kernel principal component analysis (KPCA) was carried out and applied with machine learning algorithms with all features and selected features. Fig. 4 shows the optimum parameter variation without KPCA and with KPCA with all features for RFR and XGBoost

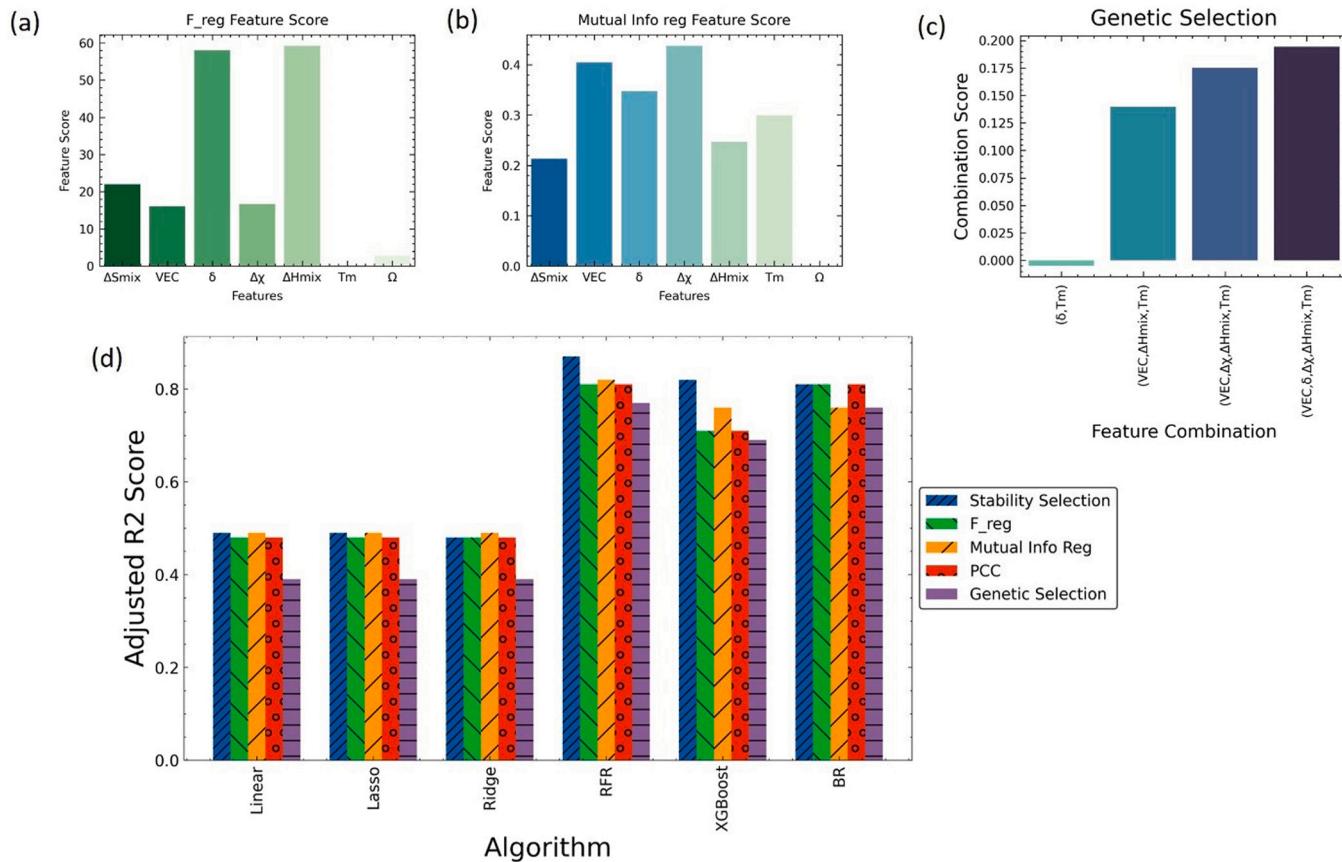


Fig. 2. Feature score for material descriptors with a) Univariate linear regression (F_reg), b) Mutual Information selection, c) Genetic selection, d) Adjusted R^2 score for feature selection algorithm with ML models.

Table 1

Results from feature selection algorithms, Yes (green) represents features to be considered, and No (red) represents features to be removed.

Feature Selection Technique	ΔS_{mix}	VEC	δ	$\Delta \chi$	ΔH_{mix}	T_m	Ω
Stability Selection	YES	YES	YES	NO	YES	YES	NO
Univariate Linear Regression	YES	YES	YES	YES	YES	NO	NO
Mutual Info Regression	YES	YES	YES	YES	YES	YES	NO
Pearson Correlation Coefficient	YES	YES	YES	YES	YES	NO	NO
Genetic Selection	NO	YES	YES	YES	YES	YES	NO

algorithms. One could note that the optimum parameter will change while using KPCA. Here, in this case, the RFR algorithm will have optimum parameter of maximum depth 10, n estimator value of 200 and the minimum number of samples in leaf of 3 without PCA, changes to maximum depth 'None', n estimator value of 100 and the minimum number of samples in the leaf of 3 with KPCA. In the case of the XGBoost algorithm, the optimum maximum depth of 6 and learning rate of 0.07 without KPCA is changed to an optimum maximum depth of 10 and a learning rate of 0.06 with KPCA.

Fig. 5 shows the predicted and true values of testing with principal component analysis using selected features and all features. Fig. 5 confirms that all features and KPCA show similar R^2 fit score for LR, Lasso, and Ridge regression. The RFR, XGBoost, and BR algorithms show a remarkable improvement in the R^2 score for selected features with KPCA. One could note that the XGBoost algorithm shows the highest R^2 fit score of 0.93 with selected features and KPCA. The ANN shows an improvement of R^2 value by applying KPCA, and still, further, improvement can be made by increasing the number of data points for training. For the current study with KPCA, the XGBoost algorithm uses the poly kernel function with max depth as 10, 100 estimators and 0.06 learning rate for all features condition; however, for selected features uses linear kernel function with a maximum depth of 5, 100 estimators and 0.09 learning rate.

Fig. 6 shows the testing performance of the studied algorithm with all features, selected features and by applying KPCA in both conditions. One could note that the RFR, BR, and XGBoost show the

R^2 score of more than 0.8 in all conditions. The comparison shows a remarkable improvement of accuracy for ANN, RFR, BR, and XGBoost algorithms using feature selection and KPCA. In the case of the SVR algorithm, KPCA's application improved the R^2 score by more than ~0.2 compared without KPCA. One could note that the current study reveals that the feature selection with the KPCA approach will give the highest testing accuracy for the most promising algorithms such as XGBoost, BR, and RFR for hardness prediction.

3.5. Comparison with experiments

The experimental verification is done using five as-cast alloys shown in Table 2. The alloys are designed in such a way that the primary FCC phase will change to primary Laves while increasing the Nb content [36]. The phase variation will ensure the remarkable variation of hardness. One could note that the composition selected for validation is not used for testing and training the algorithm. Fig. 7 shows the average experimental microhardness data and the predicted microhardness data with and without feature selection along with and without KPCA. The LR, Lasso, and Ridge algorithms are showing a reasonably good prediction for the hardness of alloy 3, and also, the error in prediction is less for other alloy systems except alloy 4. One could note that these three algorithms are able to follow the increasing trend of microhardness with Nb and the feature selection and application of KPCA improved the prediction. In the case of XGBoost, BR, SVR and RFR conditions, the variation in prediction is

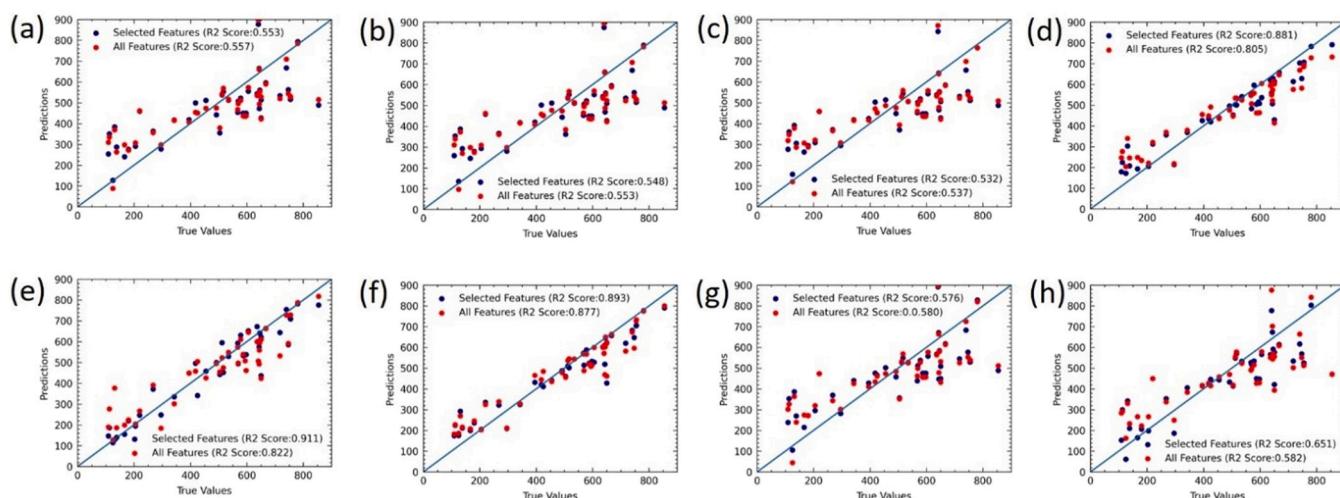


Fig. 3. True value and predicted value of test dataset with selected features and all features a) Linear regression, b) Lasso Regression, c) Ridge Regression, d) Random Forest regression, e) XGBoost Regression, f) Bagging Regressor, g) Support Vector Regressor, h) ANN.

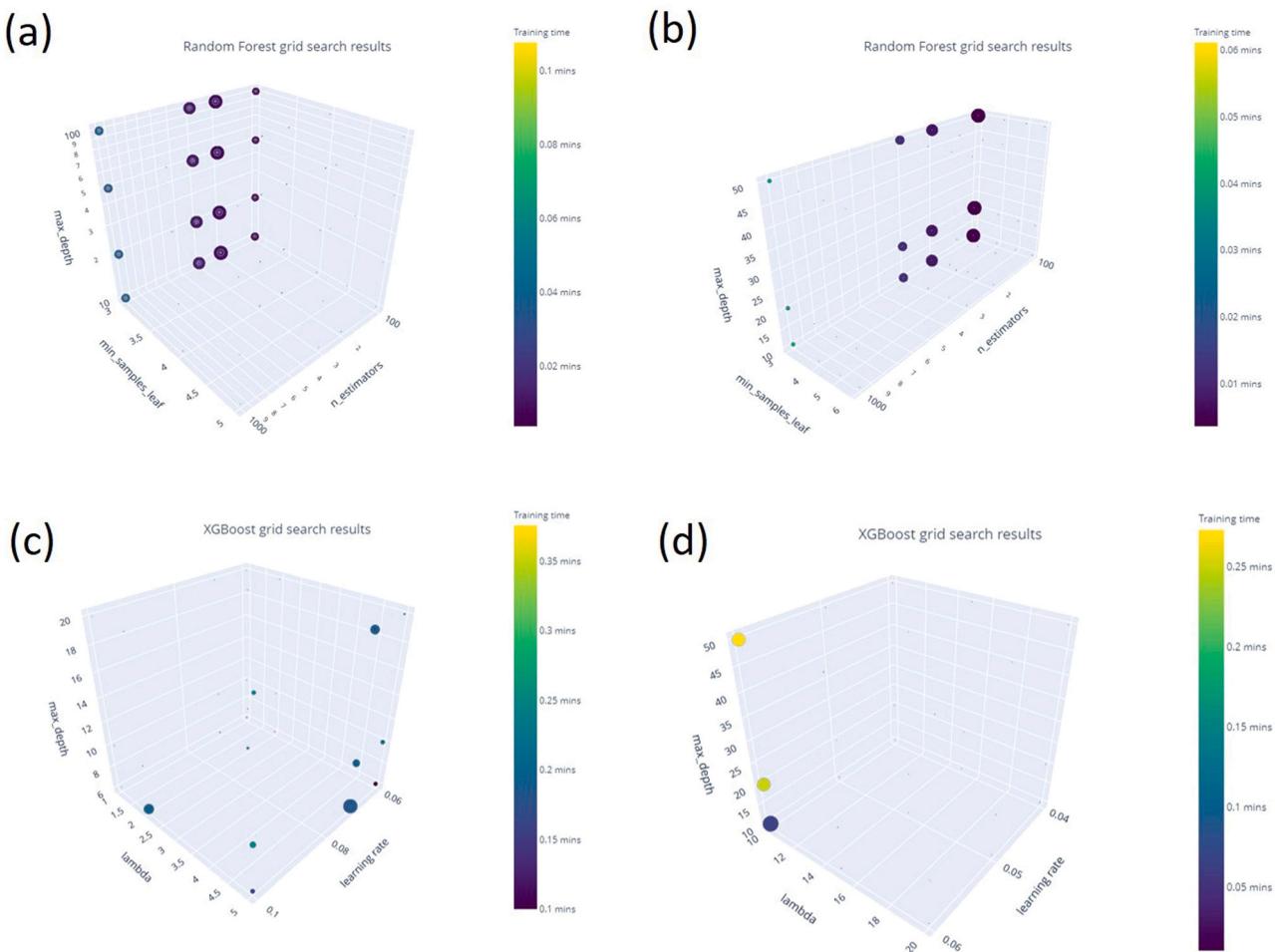


Fig. 4. Parameter optimization with all features for a) RFR without KPCA, b) RFR with KPCA, c) XGBoost without KPCA, and d) XGBoost with KPCA (Diameter of the sphere is a measure of R² score).

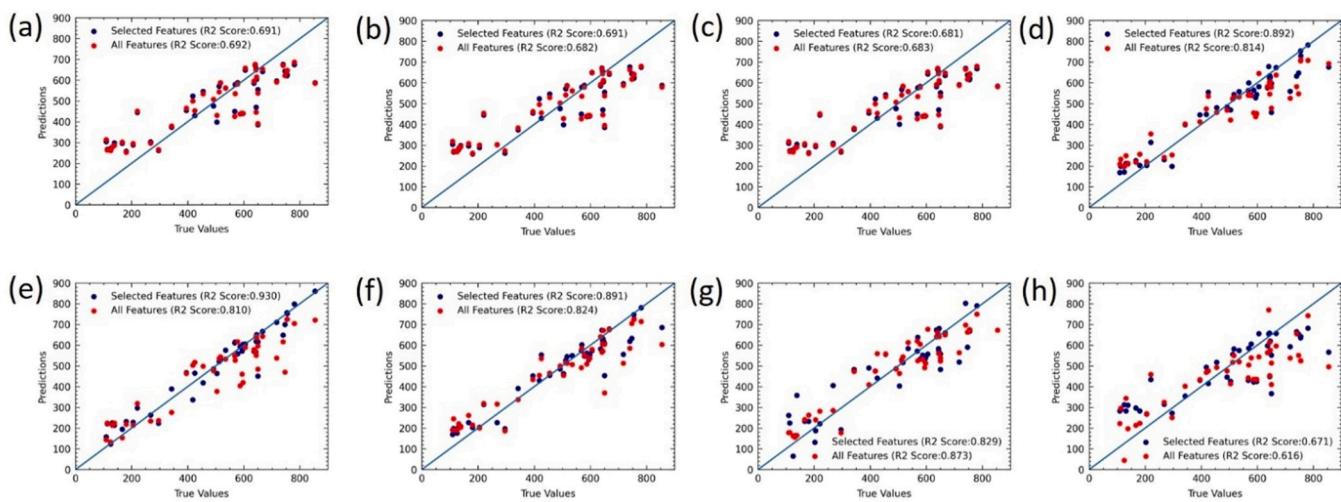


Fig. 5. True value and predicted value of test dataset with selected features and all features after applying KPCA a) Linear regression, b) Lasso Regression, c) Ridge Regression, d) Random Forest regression, e) XGBoost Regression, f) Bagging Regressor, g) Support Vector Regressor, h) ANN.

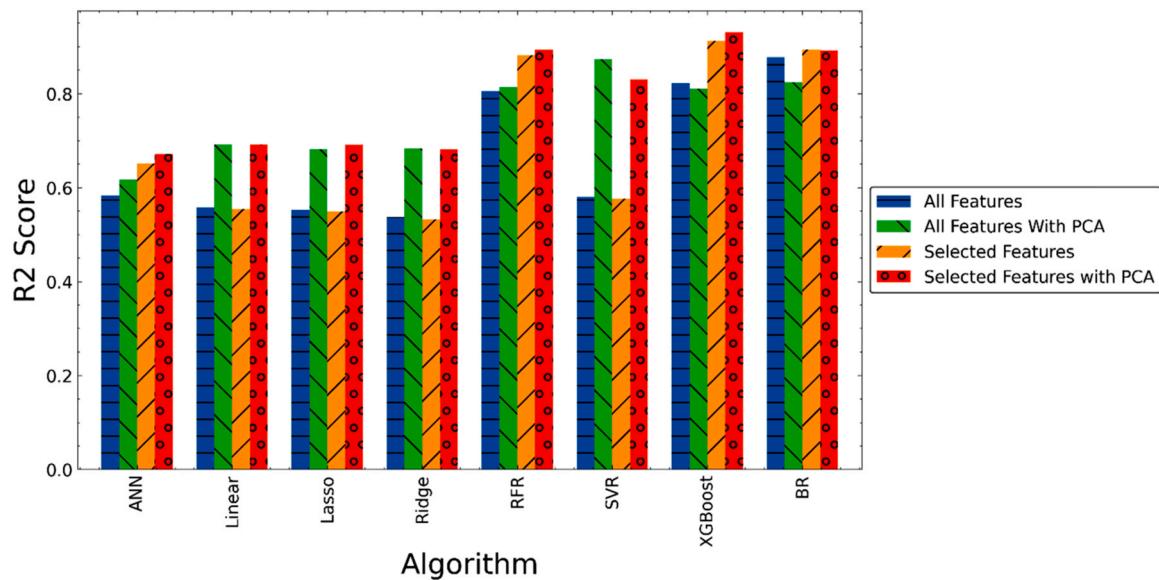


Fig. 6. Comparison of testing R^2 score for all algorithms.

Table 2

Compositions used for experimental validation.

Name of the Alloy	Composition (in at%)
Alloy 1	10 V-15Cr-5Mn-10Co-25Ni-26.34Fe-8.66Nb
Alloy 2	10 V-15Cr-5Mn-10Co-25Ni-25.5Fe-9.5Nb
Alloy 3	10 V-15Cr-5Mn-10Co-25Ni-25.3Fe-9.7Nb
Alloy 4	10 V-15Cr-5Mn-10Co-25Ni-25.27Fe-9.73Nb
Alloy 5	10 V-15Cr-5Mn-10Co-25Ni-25.2Fe-9.8Nb

high, but one could note that the variation in hardness value by applying feature selection and KPCA is quite high. This implies that these algorithms are more sensitive to feature selection and KPCA for the current study, and the prediction can be improved by exploring more data sets. In the case of ANN, the prediction with feature selection, with KPCA on all features and feature selection with KPCA, shows the reasonable well prediction for Alloy 2, 3 and 5 and with all features without KPCA able to predict the alloy 1 microhardness. One could note that the microhardness is

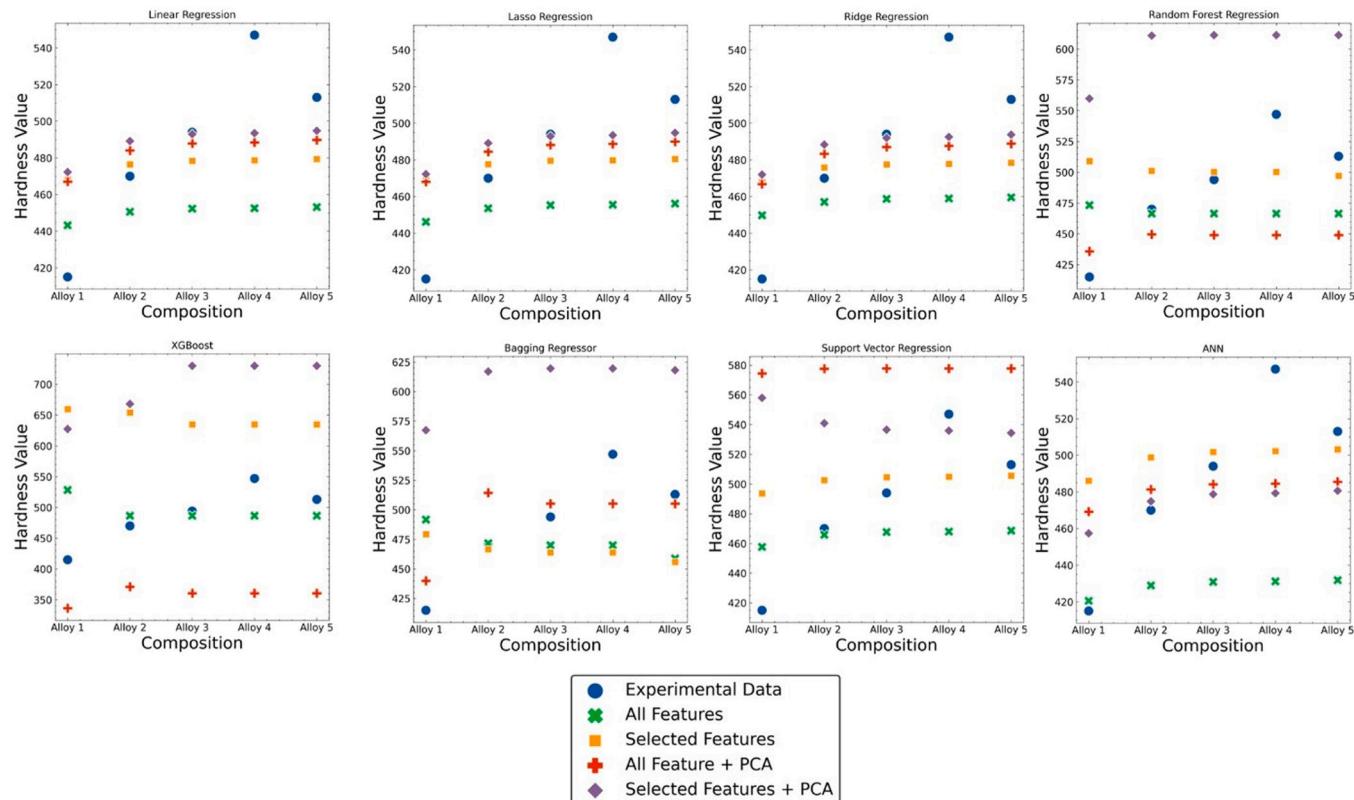


Fig. 7. Comparison of model prediction and experimental data.

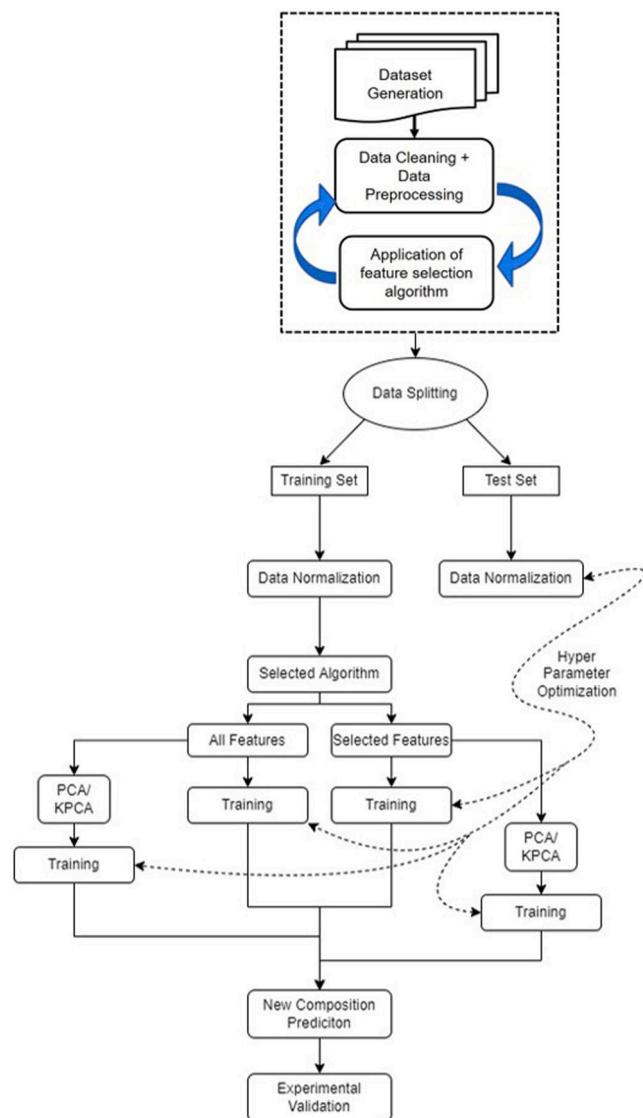


Fig. 8. Proposed framework for property prediction of new material.

underestimated with all features without KPCA for alloy 3–5, implying the applicability of feature selection and KPCA for properties prediction.

3.6. ML framework for mechanical property prediction

Fig. 8 shows the framework developed for materials property prediction for new alloys with limited data set. While using this framework, one could identify the important material descriptors for a particular property prediction and a promising algorithm for a particular dataset. The current study confirms that all materials descriptors need not be used, and the accuracy can be improved by reducing the dimensions using KPCA. One could also note that the testing accuracy may be miss leaded while coming for new predictions, so proper experimental verification is necessary for the successful application of ML. The framework also explains exploring multiple algorithms for a particular data set rather than selecting one algorithm. Based on the experimental validation, one algorithm can be used for a particular set of the alloy system. One could note that the accuracy of the algorithm can be improved by more data points which can be obtained either through experiments or by

computational techniques like CALPHAD. The verified framework is a guide for material property prediction, and further improvement can be made to improve prediction accuracy.

4. Conclusion

The current study developed a framework for material property prediction using machine learning. The dataset was developed based on the literature and experiments, and the material's descriptors were selected using a feature selection algorithm. Selecting the relevant contributing features can tremendously affect the result and training time; among all the techniques used, Stability Selection works for the given high dimensional data. Deployment of this model can reduce the tedious work involved and increase the HEA design's efficacy. The stability selection confirms that the Ω have less importance in hardness prediction for the current data set. Eight algorithms are trained and tested and found that the XGBoost algorithm shows the highest test R^2 score of 0.93 for selected features with KPCA conditions. The testing accuracy of RFR, XGBoost, BR, ANN was increased drastically by applying feature selection with KPCA. ANN shows reasonable accuracy for hardness prediction in the experimental dataset.

CRediT authorship contribution statement

Amit Singh Bundela: Data curation, Formal analysis, Investigation, Software, Visualization, Writing – original draft, Writing – review & editing. **Rahul M R:** Writing – original draft, Writing – review & editing, Conceptualization, Funding acquisition, Supervision.

Data Availability

The data used in the current study will be available from the corresponding author based on reasonable request.

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.

Acknowledgements

Rahul M R acknowledges Prof. G Phanikumar (Department of Metallurgical and Materials Engineering, IIT Madras) for providing facilities and useful discussion and Naishalkumar Shah for helpful discussion. Rahul M R acknowledges Faculty Research Scheme Project, IIT (ISM) Dhanbad (FRS(165)/2021–2022/FMME) for financial support.

References

- [1] B. Cantor, Multicomponent high-entropy Cantor alloys, *Prog. Mater. Sci.* 120 (2021) 100754, <https://doi.org/10.1016/j.pmatsci.2020.100754>
- [2] E.P. George, D. Raabe, R.O. Ritchie, High-entropy alloys, *Nat. Rev. Mater.* 4 (2019) 515–534, <https://doi.org/10.1038/s41578-019-0121-4>
- [3] T. Yang, Y.L. Zhao, B.X. Cao, J.J. Kai, C.T. Liu, Towards superior mechanical properties of hetero-structured high-entropy alloys via engineering multicomponent intermetallic nanoparticles, *Scr. Mater.* 183 (2020) 39–44.
- [4] Y.H. Jo, S. Jung, W.M. Choi, S.S. Sohn, H.S. Kim, B.J. Lee, N.J. Kim, S. Lee, Cryogenic strength improvement by utilizing room-temperature deformation twinning in a partially recrystallized VCrMnFeCoNi high-entropy alloy, *Nat. Commun.* 8 (2017) 1–8, <https://doi.org/10.1038/ncomms15719>
- [5] Z. Li, K.G. Pradeep, Y. Deng, D. Raabe, C.C. Tasan, Metastable high-entropy dual-phase alloys overcome the strength–ductility trade-off, *Nature* 534 (2016) 227–230, <https://doi.org/10.1038/nature17981>
- [6] B. Gludovatz, A. Hohenwarter, D. Catoor, E.H. Chang, E.P. George, R.O. Ritchie, A fracture-resistant high-entropy alloy for cryogenic applications, *Science* 345 (2014) 1153–1158, <https://doi.org/10.1126/science.1254581>

- [7] N.D. Stepanov, D.G. Shaysultanov, M.A. Tikhonovsky, S.V. Zherebtsov, Structure and high temperature mechanical properties of novel non-equiautomic Fe-(Co, Mn)-Cr-Ni-Al-(Ti) high entropy alloys, *Intermetallics* 102 (2018) 140–151, <https://doi.org/10.1016/j.intermet.2018.09.010>
- [8] Y.T. Chen, Y.J. Chang, H. Murakami, S. Gorsse, A.C. Yeh, Designing high entropy superalloys for elevated temperature application, *Scr. Mater.* 187 (2020) 177–182, <https://doi.org/10.1016/j.scriptamat.2020.06.002>
- [9] R. Jain, A. Jain, M.R. Rahul, A. Kumar, M. Dubey, R.K. Sabat, S. Samal, G. Phanikumar, Development of ultrahigh strength novel Co-Cr-Fe-Ni-Zr quasi-peritectic high entropy alloy by an integrated approach using experiment and simulation, *Materialia* 14 (2020), <https://doi.org/10.1016/j.mtla.2020.100896>
- [10] S. Wei, S.J. Kim, J. Kang, Y. Zhang, Y. Zhang, T. Furuhara, E.S. Park, C.C. Tasan, Natural-mixing guided design of refractory high-entropy alloys with as-cast tensile ductility, *Nat. Mater.* 19 (2020) 1175–1181, <https://doi.org/10.1038/s41563-020-0750-4>
- [11] D.B. Miracle, O.N. Senkov, A critical review of high entropy alloys and related concepts, *Acta Mater.* 122 (2017) 448–511, <https://doi.org/10.1016/j.actamat.2016.08.081>
- [12] J.M. Rickman, G. Balasubramanian, C.J. Marvel, H.M. Chan, Machine learning strategies for high-entropy alloys, *J. Appl. Phys.* 221101 (2020) 1–11, <https://doi.org/10.1063/5.0030367>
- [13] J.M. Rickman, H.M. Chan, M.P. Harmer, J.A. Smeltzer, C.J. Marvel, A. Roy, G. Balasubramanian, Materials informatics for the screening of multi-principal elements and high-entropy alloys, *Nat. Commun.* 10 (2019) 1–10, <https://doi.org/10.1038/s41467-019-10533-1>
- [14] R. Li, L. Xie, W.Y. Wang, P.K. Liaw, Y. Zhang, High-throughput calculations for high-entropy alloys: a brief review, *Front. Mater.* 7 (2020) 1–12, <https://doi.org/10.3389/fmats.2020.00290>
- [15] Q. Wang, L. Velasco, B. Breitung, V. Presser, High-entropy energy materials in the age of big data: a critical guide to next-generation synthesis and applications, *Adv. Energy Mater.* 2102355 (2021) 2102355, <https://doi.org/10.1002/aenm.202102355>
- [16] K. Kaufmann, K.S. Vecchio, Searching for high entropy alloys: a machine learning approach, *Acta Mater.* 198 (2020) 178–222, <https://doi.org/10.1016/j.actamat.2020.07.065>
- [17] G. Pilania, C. Wang, X. Jiang, S. Rajasekaran, R. Ramprasad, Accelerating materials property predictions using machine learning, *Sci. Rep.* 3 (2013) 1–6, <https://doi.org/10.1038/srep02810>
- [18] Y.V. Krishna, U.K. Jaiswal, M.R. Rahul, Machine learning approach to predict new multiphase high entropy alloys, *Scr. Mater.* 197 (2021) 113804, <https://doi.org/10.1016/j.scriptamat.2021.113804>
- [19] C. Wen, Y. Zhang, C. Wang, D. Xue, Y. Bai, S. Antonov, L. Dai, T. Lookman, Y. Su, Machine learning assisted design of high entropy alloys with desired property, *Acta Mater.* 170 (2019) 109–117, <https://doi.org/10.1016/j.actamat.2019.03.010>
- [20] 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, <https://doi.org/10.1016/j.actamat.2019.11.067>
- [21] 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, <https://doi.org/10.1016/j.matdes.2020.108835>
- [22] U.K. Jaiswal, Y. Vamsi Krishna, M.R. Rahul, G. Phanikumar, Machine learning-enabled identification of new medium to high entropy alloys with solid solution phases, *Comput. Mater. Sci.* 197 (2021) 110623, <https://doi.org/10.1016/j.commatsci.2021.110623>
- [23] X. Huang, C. Jin, C. Zhang, H. Zhang, H. Fu, Machine learning assisted modelling and design of solid solution hardened high entropy alloys, *Mater. Des.* 211 (2021) 110177, <https://doi.org/10.1016/j.matdes.2021.110177>
- [24] R. Jaafreh, Y. Seong, J. Kim, K. Hamad, Machine learning guided discovery of super-hard high entropy ceramics, *Mater. Lett.* 306 (2022) 130899, <https://doi.org/10.1016/j.matlet.2021.130899>
- [25] S. Yan, L. Zhichao, L. Xiongjun, D. Qing, X. Huamin, L. Jiecheng, S. Ruoxuan, W. Yuan, W. Hui, J. Suihe, L. Zhaoping, Prediction of Ti-Zr-Nb-Ta high-entropy alloys with desirable hardness by combining machine learning and experimental data, *Appl. Phys. Lett.* 201905 (2021), <https://doi.org/10.1063/5.0065303>
- [26] J. Xiong, S.-Q. Shi, T.-Y. Zhang, Machine learning of phases and mechanical properties in complex concentrated alloys, *J. Mater. Sci. Technol.* 87 (2021) 133–142, <https://doi.org/10.1016/j.jmst.2021.01.054>
- [27] U. Bhandari, R. Raffi, C. Zhang, S. Yang, Yield strength prediction of high-entropy alloys using machine learning, *Mater. Today Commun.* 26 (2021) 101871, <https://doi.org/10.1016/j.mtcomm.2020.101871>
- [28] J. Li, B. Xie, Q. Fang, B. Liu, Y. Liu, P.K. Liaw, High-throughput simulation combined machine learning search for optimum elemental composition in medium entropy alloy, *J. Mater. Sci. Technol.* 68 (2021) 70–75, <https://doi.org/10.1016/j.jmst.2020.08.008>
- [29] A.E. Nassar, A.M. Mullis, Rapid screening of high-entropy alloys using neural networks and constituent elements, *Comput. Mater. Sci.* 199 (2021) 110755, <https://doi.org/10.1016/j.commatsci.2021.110755>
- [30] M. Douglas C, *Design and Analysis of Experiments*, third ed., John Wiley & Sons Ltd, 2012.
- [31] 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, <https://doi.org/10.1016/j.dib.2018.11.111>
- [32] C. Yang, C. Ren, Y. Jia, G. Wang, M. Li, W. Lu, A machine learning-based alloy design system to facilitate the rational design of high entropy alloys with enhanced hardness, *Acta Mater.* 222 (2022) 117431, <https://doi.org/10.1016/j.actamat.2021.117431>
- [33] H. Zhou, X. Wang, R. Zhu, Feature selection based on mutual information with correlation coefficient, *Appl. Intell.* (2021), <https://doi.org/10.1007/s10489-021-02524-x>
- [34] N. Meinshausen, P. Bühlmann, Stability selection, *J. R. Stat. Soc. Ser. B Stat. Method.* 72 (2010) 417–473, <https://doi.org/10.1111/j.1467-9868.2010.00740.x>
- [35] R.D. Shah, R.J. Samworth, Variable selection with error control: another look at stability selection, *J. R. Stat. Soc. Ser. B Stat. Method.* 75 (2013) 55–80, <https://doi.org/10.1111/j.1467-9868.2011.01034.x>
- [36] N. Shah, M.R. Rahul, G. Phanikumar, Accelerated design of eutectic high entropy alloys by ICME approach, *Metall. Mater. Trans. A* 52 (2021) 1574–1580, <https://doi.org/10.1007/s11661-021-06218-4>