

Review | Received 28 April 2026; Revised 7 September; Accepted 21 September 2026; Published 14 October 2026
<https://doi.org/10.55092/am20260013>

A systematic review of machine learning approaches for phase stability and mechanical property prediction in high-entropy alloys



J Sharath Kumar*, Rakesh Kumar and Rajeev Verma*

Industrial & Production Engineering Department, Dr B. R. Ambedkar National Institute of Technology
Jalandhar, Jalandhar, India

* Correspondence authors; E-mails: jsk.ip.21@nitj.ac.in (J.S.K.); vermar@nitj.ac.in (R.V.).

Highlights:

- A PRISMA-based systematic review identified 58 ML studies on HEA prediction.
- ML approaches for phase stability and mechanical property prediction are systematically reviewed.
- Random forest and boosting models are widely applied to HEA phase and hardness prediction.
- Conventional HEA systems dominate, while lightweight and refractory HEAs remain underexplored.
- Dataset size, diversity, descriptors, and target properties strongly influence model performance.

Abstract: High entropy alloys (HEAs) have attracted significant research interest owing to their vast compositional space and diverse achievable properties. However, exploring this extensive design space and experimentally determining phase constituents and mechanical properties are time-consuming processes. Machine learning (ML) provides a data-driven approach to identifying relationships among alloy composition, phase, microstructure, and mechanical properties. This review was conducted in accordance with Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines. A systematic Web of Science search identified 301 records; 58 studies were selected for detailed analysis after screening. The review examined various ML models and datasets for predicting the HEA phase and mechanical properties. Most ML models have been developed to study conventional HEAs composed of Al, Co, Cu, Fe, Ni, and Mn, whereas lightweight, refractory, eutectic, and other novel HEAs remain underexplored. Prediction performance depends not only on the ML model but also on dataset size and diversity, input parameters, and target properties. Future studies should focus on developing larger, more compositionally diverse datasets, adopting transparent methods, and integrating ML models with experimental verification to improve the reliability and transferability of HEA design.

Keywords: high-entropy alloys (HEAs); machine learning; PRISMA; phase prediction; mechanical properties; systematic review; microstructure; ML models



Copyright©2026 by the authors. Published by ELSP. This work is licensed under Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium provided the original work is properly cited.

1. Introduction

Alloy design has been a key means of tailoring material properties since the Bronze Age. The vast majority of traditional alloy systems (e.g., Fe-, Ni-, or Ti-based) are based on a single principal alloying element, with small amounts added to enhance strength through mechanisms such as precipitation hardening or grain-boundary strengthening [1]. In extreme environments, however, their applications are intrinsically limited by dislocation saturation in Fe-based alloys, grain-boundary sliding in Ni-based superalloys, and surface crack-induced fatigue in Ti-based alloys [2]. Such restrictions show that a new strategy for alloy design is required.

In the early 21st century, materials scientists introduced a radically different approach to alloy design: high-entropy alloys (HEAs), which introduced a different alloying strategy in 2004 [3]. Unlike conventional alloys, which are based on a single principal element with minor alloying additions, HEAs contain multiple principal elements in comparable concentrations that form a random solid solution [4]. These materials were termed HEAs because their high configurational entropy was initially considered a key factor in stabilising the solid-solution phases. What makes HEAs especially interesting is their ability to combine properties that are often difficult to achieve together, such as high strength, good chemical stability, and reasonable workability [5]. More importantly, this multi-element design approach opens up a vast, largely unexplored compositional space, far beyond what is possible with conventional alloy systems.

1.1. Core effects of HEAs and their mechanism of action

The unique behaviour of HEA is characterised by four core effects: high entropy, slow diffusion, lattice distortion, and the cocktail effect [6]. These are not separate mechanisms but rather overlapping effects of the multi-component nature of HEAs and should not be viewed in isolation.

High entropy effect: HEAs contain a lot of configurational entropy because of the random mixing of two or more major elements, which reduces the Gibbs free energy of the system:

$$\Delta G_{\text{mix}} = \Delta H_{\text{mix}} - T\Delta S_{\text{mix}}$$

$$\Delta S_{\text{mix}} \geq 1.6R$$

The entropy contribution is important at high temperatures and may favour simple solid solutions over intermetallic phases. This effect is, however, overstated. Entropy is not the sole determinant of phase stability in HEAs; enthalpy of mixing, atomic size mismatch, and electronic considerations are equally critical [7].

Slow diffusion effect: Since diffusion in HEAs is chemically disordered, the local bonding conditions and energy barriers are distributed, which can sluggishly slow down atomic diffusion. This is often linked to improved thermal stability and resistance to grain growth. That said, the behaviour of diffusion depends on composition over a wide range, and the assumption of generally slow diffusion is not necessarily true.

Severe lattice distortion effect: Lattice distortion due to the difference between the atomic sizes of the constituent elements. This gives rise to internal stress fields that inhibit dislocation motion and are a

source of strengthening. This distortion is not localised as in conventional alloys, but is spread throughout the lattice.

Cocktail effect: The cocktail effect is often applied to refer to the synergistic effect of various factors in material properties. Practically speaking, it indicates that, in many cases, HEAs exhibit non-linear and even surprising behaviour. But it is not a clear physical process or feature; rather, it is an adjective describing complex interactions that are not yet fully understood.

These four core effects of HEAs provide the fundamental basis for understanding their formation and properties. Based on these concepts, several thermodynamic and empirical parameters have been proposed to predict phase formation in HEAs, as shown in Figure 1.

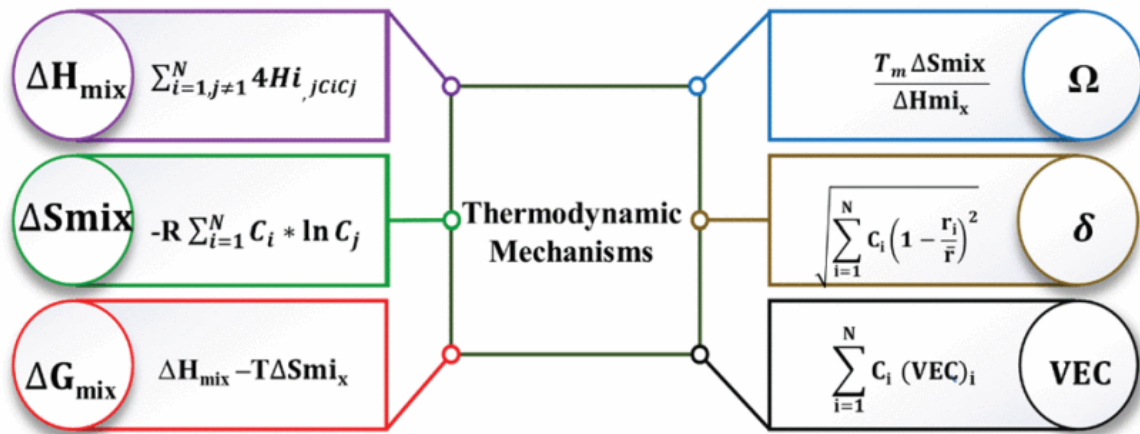


Figure 1. Thermodynamic parameters influencing phase formation [8]

1.2. Challenges in modelling high-entropy alloys

The modelling of HEAs is not easy, primarily due to the sensitivity of such systems to composition and the size of their design space. HEAs can be highly responsive to small compositional changes, whereas traditional alloys tend to exhibit gradual changes in properties in response to the same changes. Minor compositional changes, even a few atomic per cent, can affect phase stability or the microstructure, or dramatically alter properties such as hardness or strength [9]. In practice, it suggests that a small change in composition can move the alloy into a different regime, and that predicting this regime might be hard and not intuitive. The large dimensionality of the compositional space of HEAs exacerbates the challenge. The number of combinations of several major components increases exponentially; therefore, it is nearly impractical to try all combinations experimentally (or computationally). Consequently, the alloy design of HEAs is no longer a matter of fine-tuning but rather a large, complex search space. Traditional modelling techniques are still important, but are restricted in this regard. Traditional modelling techniques remain significant, but they are limited in this respect. Estimations of phase stability on a thermodynamic basis are commonly performed using methods such as Calculation of Phase Diagram (CALPHAD), and Molecular Dynamics (MD) is useful for modelling the behaviour of atoms at the atomic scale, such as diffusion and deformation [10,11]. Density Functional Theory (DFT) and other first-principles methods can provide greater insight into phase energetics and electronic structure [12]. Nevertheless, implementing these techniques in HEAs becomes more difficult as the number of elements and possible configurations increases. In general, the high compositional sensitivity

and the enormous design space make HEAs challenging to model with traditional methods alone. That is why data-driven techniques, especially machine learning, have become a focus; they provide a means of dealing with non-linear relations that are complex and of investigating the compositional space more effectively.

1.3. Merits of machine learning for high-entropy alloys

Over the past decade, machine learning (ML) and deep learning have witnessed explosive growth. The essence of ML is to capture and model complex patterns that are hard to describe by traditional methods. They are built in a systematic way that learns from data and are enhanced by the availability of large volumes of data generated in experiments and simulations, as well as by high-performance computing and sophisticated algorithms. This combination makes the ML approach particularly well-suited to addressing the challenges of theoretical modelling of HEAs [13]. There is an increasing need for ML, as reflected in the growing number of relevant publications. Importantly, ML models offer a significant computational advantage: they are much faster than DFT calculations and can be as accurate [14]. This enables simulations at scales of millions of atoms, much larger than what is possible with conventional DFT. This is especially crucial for the study of the extraordinary mechanical properties of HEAs, as these are related to complex ordering, defects and microstructures, which demand accurate nanoscale simulations. In addition to simulations, ML can also be a powerful predictive tool. Data-driven learning can reveal nonlinear relationships and features between structure and properties that are difficult to infer by traditional reasoning. By examining these relationships, the researchers were able to study and explore HEAs more efficiently. The advantages of ML discussed above have spurred a new paradigm of research on HEAs that is centred on data, with concrete examples of its utility in atomistic simulations, predictive modelling of physical properties, and materials design.

Based on the literature, designing a novel HEA with specific properties is challenging and often requires multiple trials to achieve the desired phase and mechanical properties. To limit the process, many researchers are exploring ML. However, training an ML algorithm requires larger data. So, this systematic review mainly focuses on various ML algorithms used to study novel HEAs and their mechanical properties, as well as data set collection. However, the study has been limited to research on as-cast HEA microstructure, phase, and mechanical properties. The review has been conducted using Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines.

2. Methodology

The review was conducted in accordance with the PRISMA guidelines to ensure clarity, transparency, and reproducibility throughout the study selection process. A systematic literature search was performed in the Web of Science database, covering publications available up to 20th March 2026. No filters were applied other than the English language. Various search terms were used in different combinations, as listed in Appendix A, based on the aim of the review: to identify ML studies related to the prediction of HEA phase and mechanical properties. Using different combinations of search terms, 301 records were found; 127 duplicates were identified and manually removed from the title and Digital Object Identifier (DOI) fields in Microsoft Excel by sorting. Five records were removed because they were retrieved articles or had incomplete information. After the initial identification, there were 169 records.

The next screening was conducted by title; those not relevant to the scope of the review (e.g., HEA coatings, review articles, wear, and corrosion studies) were removed. Following title screening, 102 articles were identified as potentially relevant and retained for full-text assessment. Eight articles had full data that could not be retrieved, leaving 94 articles for eligibility assessment. These articles were further screened based on their abstracts and conclusions. Studies that did not focus on HEA microstructures, phase prediction, hardness, or yield strength were excluded at this stage. Following the final eligibility assessment, 58 articles met the inclusion criteria and were included in the review. The overall process of study identification, screening, eligibility assessment, and inclusion is illustrated in the PRISMA flow diagram, as shown in Figure 2.

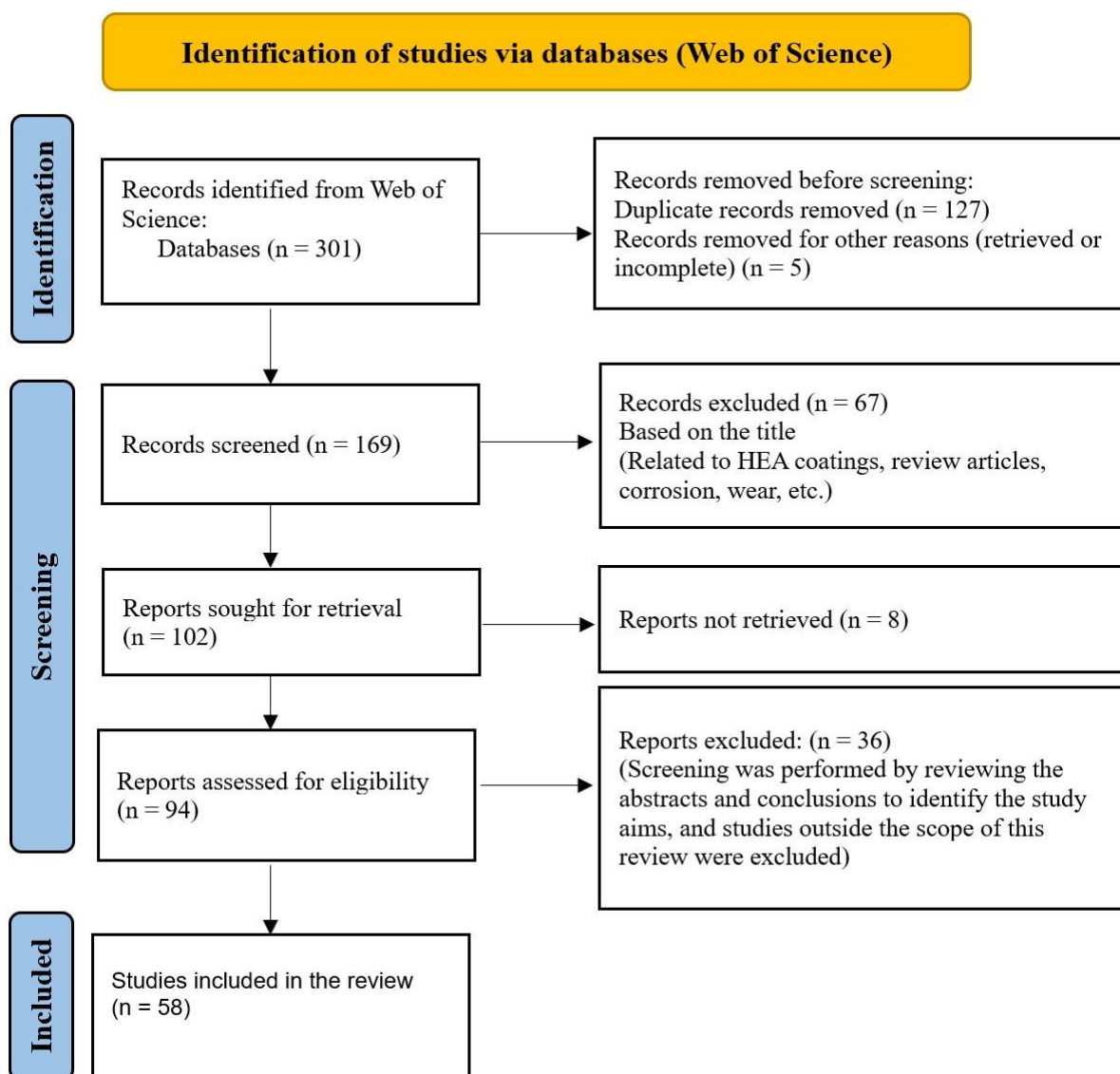


Figure 2. PRISMA flow diagram of study selection [15].

3. ML algorithms

Several machine learning models have been investigated for the study of HEA, and there are some clear trends in the literature. The model is usually determined by the size of the data set, the type of prediction (phase vs property) and the complexity of the relationships [16]. The most widely used model is Random

Forest (RF), which trains multiple decision trees on a subset of the data and then aggregates their predictions [17]. This averaging effect helps reduce overfitting and stabilise the model, particularly for HEA data, which is often small and noisy. Other popular boosting-based approaches include gradient boosting and XGBoost, which often achieve among the highest predictive accuracies [18]. Unlike Random Forest, boosting builds trees sequentially, trying to correct the errors of the previous ones. Because of this sequential learning, these models can learn more complex patterns and hence work well at predicting the phase and estimating the mechanical properties.

Support Vector Machine (SVM) and its regression counterpart (SVR) are typically employed, especially when the dataset is small or limited [19]. The models are designed to find an optimal boundary that separates or fits the data in a high-dimensional space. They can also model non-linear relationships using kernel functions, which are well-suited to HEA problems with numerous descriptors that are related in complex ways. Artificial Neural Networks (ANNs), Multi-Layer Perceptrons (MLPs), and Deeper Neural Networks (DNNs) are typically used to predict properties such as hardness, yield strength, and elastic properties [20]. The models are trained by tuning the weights between layers and are therefore capable of learning highly nonlinear relationships. However, these are highly data-dependent and can overfit with limited data. Other studies use the K-Nearest Neighbours (KNN) algorithm, which is normally used for simpler classification tasks, such as phase prediction [21]. It uses the nearest data points to its feature space to make projections. It is effective on low-dimensional, well-structured data, but its performance degrades as the number of features increases or as the data becomes noisy. Recent studies reveal a trend towards ensemble models and hybrid models (Stacking models, Voting models). These approaches amalgamate several algorithms to enhance prediction stability and accuracy. They can be especially useful at controlling the complexity of multi-phase HEA systems by leveraging the advantages of the different models. Gaussian Process Regression (GPR) is occasionally used in regression. Unlike other models, it not only makes predictions but also estimates uncertainty, which can be helpful when deciding on new alloy compositions. Finally, simpler models like Logistic Regression, Decision Trees, and Naive Bayes are sometimes used as baselines. Although they are simple to apply and interpret, they tend not to work as well as ensemble or boosting techniques in the context of the complexity of HEA data. The evaluation of ML models is performed using various metrics for regression and classification problems, as shown in Table 1.

The size of the dataset, the number of features, and the prediction goals are important factors in the ML algorithm used in HEA research. As is evident in the literature, no model is consistently better suited than any other, but some trends are evident. Random Forest and boosting-based models consistently outperform, particularly for phase prediction and hardness estimation, while SVM-based models are effective on smaller datasets. To capture these trends, Table 2 presents detailed information on the reported studies, including alloy systems, dataset sizes, input features, machine learning models, and prediction performance. This list can be used directly to compare different modelling strategies and to recognise correlations among data characteristics, modelling choices, and predictive accuracy in HEAs.

Table 1. Key evaluation metrics employed for regression and classification models [8].

Category	Evaluation Metrics	Formulas	Description
Regressors	R^2	$1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2}$	The coefficient of determination (R^2) represents the proportion of variance in the dependent variable explained by the independent variables in the model.
	MAE	$\frac{1}{n} \sum_{i=1}^n y_i - \hat{y}_i $	Mean absolute error (MAE) represents the average absolute difference between the actual and predicted values, with lower values indicating better prediction accuracy.
	RMSE	$\sqrt{\frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2}$	Root mean square error (RMSE) is the standard deviation of prediction errors.
	MAPE	$\frac{1}{n} \sum_{j=1}^n \frac{ (Pr_j - Ac_j) }{Pr_j} \times 100\%$	Mean absolute percentage error (MAPE) is the average percentage error between actual and predicted values.
Classifiers	Precision	$\frac{TP}{TP + FP}$	True positives divided by all positive predictions.
	Accuracy	$\frac{\sum_{i=1}^n TC_i}{Total\ Prediction}$	The proportion of correct predictions is unreliable for imbalanced data.
	F1-score	$\frac{2 \times Precision \times Recall}{Precision + Recall}$	Harmonic mean of precision and recall.

Table 2. Summary of ML models and key evaluation metrics used in predicting HEA phase and mechanical properties.

Collection of datasets	HEA configuration	Size of dataset	Input Features	ML Model(s)	Performance Metrics	Predictions/Application Scenarios	References
Experimental/Literature	AlCoCrCuFeNi	A total data set of 323 alloy phases, and for micro hardness, a 176-sample data set	7 parameters (VEC, δ , $\Delta\chi$, ΔS_{mix} , ΔH_{mix} , T_m and Ω)	Gaussian noise augmented data And support vector machine (SVM)	The regression model for medium-noise samples achieved the highest prediction accuracy, with an R^2 value of 0.954	HEA phase classification and hardness prediction regression model	[22]
	AlCoCuCrFeNi	A total data set of 205 alloys, which contains Hardness data	7 parameters (δ r, γ , E, G, δG , Peierls-Nabarro factor (F), lattice distortion energy (μ), and energy term in strengthening model (A))	Hardness prediction models (SVM, Random Forest (RF), XGBoost, Artificial neural network (ANN), Lasso, Ridge), and other methods are for optimising atomic molar ratios for high hardness (SVM, ANN, Lasso, Ridge)	- For hardness prediction, the RF model showed a prediction accuracy of 94.16% - For composition optimisation, the SVM model showed high hardness with values higher than 800 HV, and the model prediction error is 2.01%	To predict the hardness of HEA (AlCoCuCrFeNi) and optimise the composition for achieving high hardness HEAs	[23]
	AlCoCrFeMnNi	A total data set of 632 alloys from experimental data with hardness	Alloy compositions (Al, Cr, Fe, Co, Cu, Mn, Mo, and Ni)	Neural Network (NN)	The ML proposed 3 compositions with hardness above 600 HV	To study and predict the Hardness of HEAs	[24]
	AlCoCuCrFeNi	A total data set of 305 alloys	6 parameters (VEC, δ , $\Delta\chi$, ΔS_{mix} , ΔH_{mix} , and y) along with Elements	SVM, Bayesian Ridge (BR), RF and these models were combined with Particle Swarm Optimization (PSO)	PSO improved the prediction accuracy. The PSO-RF model showed 96.61% prediction accuracy	To study HEA compositions and hardness. In this, 10 HEA compositions were predicted to have high hardness	[25]

Table 2. Cont.

Collection of datasets	HEA configuration	Size of dataset	Input Features	ML Model(s)	Performance Metrics	Predictions/Application Scenarios	References
Experimental/Literature	AlCoCrCuFeNi	A total data set for training 116 alloys and cross-validation from 39 data sets	Alloy compositions	Back Propagation Neural Network (BPNN)	• BPNN model showed a better prediction accuracy of 99.54 %	• To study and predict the hardness of HEAs	[26]
	AlCoCrFeNi	A total hardness dataset of 399 alloys collected from the literature. (120 BCC, 87 FCC, and 82 IM)	5 parameters (VEC, δ , $\Delta\chi$, ΔS_{mix} , ΔH_{mix})	K-Nearest Neighbours (KNN), SVM, Decision Tree (DT), Logistic Regression (LR) and RF	• RF model showed the best prediction accuracy of 87% • SVM model showed a better prediction accuracy of more than 0.9	• To study and predict the crystallite phases of refractory HEAs • ΔH_{mix} and VEC parameters play an important role in predicting the phase of HEAs	[27]
	AlCrCuFeNiSi	A total hardness data set of 309 phase alloys collected from the literature	Nine features were used for phase prediction and ten features for hardness prediction, selected from a total of 65 features	For phase prediction, Multi-layer Perceptron (MLP), RF, GBT, NB and AdaBoost models were used; For hardness prediction, SVR, RF, ANN and XGBoost models were used	• RF model showed best accuracy for phase prediction with a value of 94.6% • XGBoost model showed the best accuracy for phase prediction with a value of 95.5%	• To study and predict the crystallite phases and hardness of HEAs	[28]
	AlCoCrFeNiMoTi	A total data set of 483 from the literature	7 parameters (VEC, δ , T_m , ΔS_{mix} , ΔH_{mix} , and Ω)	MLP, RF, SVM, XGBoost, and DT	• XGBoost showed a better prediction accuracy with a correlation coefficient of 0.95	• To study and predict the hardness of HEAs (AlCoCrFeNiMoTi)	[29]

Table 2. Cont.

Collection of datasets	HEA configuration	Size of dataset	Input Features	ML Model(s)	Performance Metrics	Predictions/Application Scenarios	References
Experimental/Literature	CoCuFeNi _x	A total data set of 664 from the literature. (267 BCC, 199 FCC and 198 FCC+BCC)	7 parameters (VEC, δ , T_m , ΔS_{mix} , ΔH_{mix} , and Ω)	LR, DT, GBDT, SVM, RF-SVM and ANN	ANN showed a better prediction in mixed phases	- To study and predict the phase of HEAs - VEC showed a 0.3 importance score in the predicting phase	[30]
	HEAs (Al, Cr, Co, Fe, Ni, Cu, Ti, V, Mn, Si, Mo, Nb, Zr, W, Sn, Ta, and Hf)	A data set of 2043 samples about 17 constituent elements	5 parameters (VEC, δ , $\Delta\chi$, ΔS_{mix} , and ΔH_{mix})	KNN	- KNN with accuracy of 93% - Prediction accuracy for the individual phases (BCC, FCC, HCP, IM, and others) exceeded 97%	For phase prediction and selection of HEAs	[31]
	HEAs (21 elements: Ni, Cr, Fe, Co, Al, Ti, Mo, Nb, Cu, Zr, V, Mn, Ta, Si, Hf, W, B, Sn, Zn, Mg, and Pd)	A data set of 2231 entries	A set of 64 artificial parameters derived from 19 basic elemental parameters	RF, GBT, ADA, and STA	- RF with accuracy of 71.3% (F1 score 0.657) - GBT showed accuracy of 72.2% (F1 score 0.678) - ADA showed accuracy of 52.5% (F1 score 0.473) - STA showed accuracy of 72.03% (F1 score 0.665)	To study and predict crystal structure phases	[32]
	HEAs (10 elements: Cu, Cr, Ni, Al, Mn, V, Co, Fe, Mo, and Ti)	A data set of 490 HEA samples' hardness values	NR	RF, XGBoost, and GPR	- RF with accuracy of 90% - XGBoost showed an accuracy of 89.9% - GPR showed an accuracy of 69.51%	To predict the materials hardness of HEA	[33]

Table 2. Cont.

Collection of datasets	HEA configuration	Size of dataset	Input Features	ML Model(s)	Performance Metrics	Predictions/Application Scenarios	References
Experimental/Literature	HEAs (14 elements: Al, Co, Ni, Cr, Fe, Cu, Mn, Ti, V, Mo, Nb, Hf, Zr and Ta)	A total data set of 544 alloys (248 IM, 131 SS and 165 AM)	7 parameters (VEC, δ , $\Delta\chi$, ΔS_{mix} , ΔH_{mix} , T_m and K)	RF, GBDT, and LR	• SVC model showed a better prediction accuracy of 98% • The GAN model was used to generate a dataset. This also improved the model's accuracy in IM phases	• To predict crystallite phases of HEAs	[34]
	HEAs (composed of 30 elements)	A total of 1345 alloys for training from the open-source literature. (213 BCC, 135 FCC, 82 FCC+BCC, 275 Multi-phase)	5 parameters (VEC, δ , $\Delta\chi$, ΔS_{mix} , and ΔH_{mix}) and alloy compositions	XGBoost, AdaBoost, Gradient Boosting, LightGBM, and CatBoost	• XGBoost model showed a notable prediction accuracy of 94.05 %	• To study and predict the mixed and SS phases of HEAs	[35]
	HEAs (composed of 19 elements) (Al, Co, Fe, Cr, Si, Ti, V, Mn, Ni, Cu, Zn, Zr, Nb, Mo, Sn, Hf, Ta, W, and Re)	A total data set for training 487 alloys from publications and experiments. (235 BCC, 196 FCC, and 55 FCC+BCC)	21 features (Thermodynamic, atomic size, and electronic parameters)	12 models SVC, ET, ADA, RF, GBDT, LGB, XGBoost, XGBRF, SGC, CAT, VC, and ANN-MLP	• ANN classifier MLP model showed robustness and prediction accuracy with an F1 score of 0.94	• To study and predict the SS phases of HEAs	[36]
	HEAs 15 elements (Al, Ti, V, Mn, Fe, Cr, Co, Ni, Cu, Zr, Nb, Hf, Mo, Ta and W)	A total data set of 85 alloys with hardness data collected from the literature	33 features (Phase formation parameters, SSH theoretical parameters, other elemental parameters)	MLP, LibSVM, RF, and KNN	• RF model showed prediction on the hardness of HEAs with SS	• To study and predict the hardness of SS. • RF model showed prediction matched with experimental for FeNiCuCo and CrMoNbTi HEAs	[37]

Table 2. Cont.

Collection of datasets	HEA configuration	Size of dataset	Input Features	ML Model(s)	Performance Metrics	Predictions/Application Scenarios	References
Experimental/Literature	HEAs (Al, Fe, Nb, Co, Cr, Cu, Ni, Ti, Mo, Ta, Mn, Zr, V, Si, Be, Sn, Hf, and W)	A total of 296 alloy data sets were collected from the literature. (111 BCC, 99 FCC, and 86 BCC+FCC)	12 features (Thermodynamic, atomic size, and electronic parameters)	MLP, SVM, DT, RF and KNN	• A two-feature selection strategy implemented. With fewer inputs, the RF model achieved higher accuracy, up to 0.98	• To predict the crystallite phase structure of HEAs with high accuracy, which will improve the discovery of new HEAs with desired mechanical properties	[38]
	HEAs	302 data set in each class (SS, IM and mixed)	6 parameters (VEC, δ , $\Delta\chi$, ΔS_{mix} , ΔH_{mix} , and T_m)	KNN and RF	• KNN with accuracy of 92.31% • Random Forest Classifier with an accuracy of 91.21%	• To predict the phase formation ability of HEAs	[39]
	HEAs (Al, Cu, Co, Cr, Fe, Ni, Mn, Ti, V, Nb, Mo, Zr, Hf, Ta and their derivatives)	A total data set of 611. The dataset comprised 285 FCC HEAs, 95 FCC + BCC HEAs, 192 FCC + IM HEAs, and 39 FCC + BCC + IM HEAs	7 parameters (VEC, δ , $\Delta\chi$, ΔS_{mix} , ΔH_{mix} , and T_m) Elemental compositions	MLP, DT, RF, GB, KNN, XGBoost and SVM	• RF model showed best accuracy with a prediction score of 0.913	• To study and predict crystal phase structure	[40]
	HEAs	A total data set of 656 alloys (306 BCC, 194 FCC, 156 FCC+BCC phases)	9 parameters (VEC, δ , $\Delta\chi$, ΔS_{mix} , ΔH_{mix} , and T_m)	SVM, GBDT, MLP, and LR	• SVM model showed a good F1 score of 0.95, and the remaining models have 0.948, 0.9, and 0.896 for GBDT, MLP and LR, respectively	• To study and predict the phase and residual strain in HEAs	[41]

Table 2. Cont.

Collection of datasets	HEA configuration	Size of dataset	Input Features	ML Model(s)	Performance Metrics	Predictions/Application Scenarios	References
Experimental/Literature	HEAs	A total data set of 601 alloys (131 SS, 165 Amorphous (AM), 248 IM, 6 AM+IM, and 51 IM+SS)	13 parameters (VEC, δ , $\Delta\chi$, ΔS_{mix} , ΔH_{mix} , and T_m)	RF, SVM and stacked ensemble	- Stacked ensemble predicted over 90% for SS, IM, and AM phases - A mixed-phase SVM model showed better prediction near to 50%	To study and predict the crystallite phase structure of HEAs	[42]
	HEAs	A total data set of 511 alloys (177 FCC, 173 BCC, and 166 FCC+BCC)	13 parameters (VEC, δ , $\Delta\chi$, ΔS_{mix} , ΔH_{mix} , Mean and Standard Cohesive energy, bulk modulus and T_m)	DT, RF, XGBoost, Voting, Stacking	All models showed an accuracy of more than 90%	To predict solid solution phase formation in HEAs	[43]
	HEAs	A total data set of 322 alloys (258 SS, 31 IM, and 33 AM)	5 parameters (VEC, δ , $\Delta\chi$, ΔS_{mix} , and ΔH_{mi})	KNN, SVM, LR, DT, RF and GNB	DT and SVM models showed the best prediction accuracy of 93.84%, and the remaining models showed a prediction accuracy above 90%	To select HEA with desirable properties, mainly the crystal structure	[21]
	HEAs	A total data set of 482 alloys	5 parameters (VEC, δ , $\Delta\chi$, ΔS_{mix} , and T_m)	SVM, KNN, Ensemble Classification EC, ANN and CART	- ANN model showed better prediction accuracy with a low number of features - ANN showed prediction accuracy upto 90%, whereas other models showed above 70%	To study and predict the crystallite phase structure of HEAs	[44]

Table 2. Cont.

Collection of datasets	HEA configuration	Size of dataset	Input Features	ML Model(s)	Performance Metrics	Predictions/Application Scenarios	References
Experimental/Literature	HEAs	A total data set of 426 alloys (180 quinarys, 189 senaries, and 57 septenaries)	7 parameters (VEC, δ , $\Delta\chi$, ΔS_{mix} , ΔH_{mix} , and Ω)	KNN, DT, LR, RF and Dempster-Shafer (DS)	The hybrid model DS showed better prediction accuracy in all categories than the remaining models	To study and predict the crystallite phase structure of HEAs	[45]
	HEAs	A total data set of 118 alloys	5 parameters (VEC, δ , $\Delta\chi$, ΔS_{mix} , and ΔH_{mix})	SVM, KNN, ANN, and a new model	The proposed model showed higher predictive accuracy than existing models. This model showed an accuracy of 92%	To predict crystallite phases of HEAs	[46]
	HEAs	A total data set of 5692 alloys from experimental data. This consists of 50 elements and 11 phase categories	12 parameters (Thermodynamic, atomic size, and electronic parameters)	XGBoost, RF, ANN, and SVC	- XGBoost and RF showed higher prediction accuracy across all phases (>85%). - In the SS prediction, all models showed above 95% levels	To predict crystallite phases of HEAs	[47]
	HEAs and MPEAs	A total data set for training 1193 alloy compositions categorised into different phases (BCC, FCC, FCC+BCC, IM, BCC+IM, FCC+IM, and BCC+FCC+IM) using DFT	5 parameters (VEC, δ , $\Delta\chi$, and Ω)	SVM, RF, DNN, XGBoost, and LR	DNN and XGBoost models showed a better prediction accuracy, exceeding 95 %	To study and predict the mixed and SS phases of HEAs	[14]

Table 2. Cont.

Collection of datasets	HEA configuration	Size of dataset	Input Features	ML Model(s)	Performance Metrics	Predictions/Application Scenarios	References
Experimental/Literature	HEAs	A total hardness data set of 261 alloys of single-phase solid solution HEAs	32 parameters	MLP, SVR, KNN, RF, and SGR	RF model showed a correlation coefficient up to 0.91 on an independent test	To study and predict the hardness of SS	[48]
	HEAs	A total of 902 alloys were collected from 12 data sets. (405 quinary, 359 senary, and 138 senary HEAs)	Meta features collected from different data sets. Majorly 5 parameters	DT, KNN, SVM, and RF Proposed new meta-learning model	- Septenary HEAs showed better prediction accuracy near to 95.8% - Meta-learning showed better accuracy than other models	To study and predict the crystallite phases of HEAs	[49]
	HEAs	A total dataset of 1364 items from 430 articles was collected from the literature. (896 SS(354 FCC, 441 BCC, and 101 BCC+FCC))	36 parameters (Thermodynamic, atomic size, material properties and electronic parameters)	DT, LDA, NB, GLMNET, RF, ANN, KNN and SVM	RF model showed better prediction, 97.5 %	To study and predict the phase of HEAs	[50]
	HEAs	A total of 214 data points from the literature, of which 171 were used for training.	7 parameters (VEC, δ , T_m , ΔS_{mix} , ΔH_{mix} , and Ω)	ANN, LR, Lasso, RF, SVR, XGBoost, and BR	XGBoost models showed a better hardness prediction with a correlation coefficient of 0.93	To study and predict the hardness of HEAs	[51]

Table 2. Cont.

Collection of datasets	HEA configuration	Size of dataset	Input Features	ML Model(s)	Performance Metrics	Predictions/Application Scenarios	References
Experimental/Literature	HEAs	A total dataset of 404 alloy phases (IM and AM phases)	Thermodynamic, atomic size, and electronic parameters)	KNN, RF, SVM, XGBoost, Adaboost, LR, and DT	• XGBoost model showed a better prediction accuracy	• To study and predict the phase of HEAs	[52]
	HEAs	A total dataset of 775 alloys from experimental. For hardness 427 samples with 18 parameters	28 parameters (Thermodynamic, atomic size, and electronic parameters)	NN, and SVM	• NN showed a higher hardness prediction accuracy than other models	• To study and predict the crystallite phases and hardness of HEAs	[53]
	Lightweight HEAs (containing Al, Mg, and Li)	A data set of 97 samples (with densities below 6 g/cm ³)	6 parameters (VEC, δ , $\Delta\chi$, ΔS_{mix} , ΔH_{mix} , and T_m)	14 different models were Trained in that 3 models with the best accuracy were selected (RF, Gradient Boosting Classifier (GBC), Cat Boost Classifier (CBC))	• RF with accuracy of 86.33% (F1 score 0.796) • GFC showed accuracy of 91.66% (F1 score 0.8923) • CFC showed accuracy of 88.55% (F1 score 0.834)	• The study aims to establish ML to discover LHEAs with simple solid solutions	[54]
	Refractory HEA (RHEA) (NbTiVZr)	A data set of 75	5 parameters (VEC, δ , $\Delta\chi$, ΔS_{mix} , and ΔH_{mix})	SVR, XGBoost, GBR, RF, GPR, ANN, and MLR	• SVR model showed the best performance with a prediction score of 0.91	• To study and predict the hardness of RHEAs	[20]
	RHEAs (W, Nb, Ta, Hf, Mo, Ti, Ni, Zr, V and Al, etc.)	A total data set of 227 alloys.	32 parameters derived from elemental features (E, G, B, etc.)	DT, LR, KNN, SVR, Lasso, and Ridge	• The LGBM model has shown better prediction accuracy with a score of 0.977	• For optimisation of composition and to predict yield strength	[55]
	RHEAs (Ti, Ta, Zr, Hf, V, Nb, Cr, Mo and W)	A total data set of 227 alloys	Elemental compositions	BPNN, SVM, Boosting Ensemble, and DT	• The BPNN model has shown better prediction accuracy, 92.68%	• To predict crystallite phases of RHEAs	[56]

Table 2. Cont.

Collection of datasets	HEA configuration	Size of dataset	Input Features	ML Model(s)	Performance Metrics	Predictions/Application Scenarios	References
Experimental/Literature	RHEAs (low-, medium-, high entropy alloys) (Cr, Hf, Mo, Nb, Ta, Ti, V, W, and Zr)	A total hardness data set of 2982 alloys	7 parameters (VEC, δ , $\Delta\chi$, ΔS_{mix} , ΔH_{mix} , T_m and γ)	SVM, KNN, DT and EL	SVM model showed a prediction accuracy of 92%	To study and predict the crystallite phases of refractory HEAs	[57]
	RHEAs (CrTaTiV-based HEAs)	A total data set of 224. (11 ternary, 94 quaternary, 93 quinary, 22 six-component, and 4 seven-component alloys)	5 features (ΔH_{mix} , E, D, μ , δ , and VEC)	Ridge, SVR, Lasso, ANN, RF, AdaBoost, and XGBoost	RF model showed prediction accuracy with $R^2 = 0.8$	To study and predict the hardness of RHEAs, mainly CrTaTiV-based HEAs	[58]
	RHEAs	A data set of 29,269	11 parameters (r_A/r_C , δ , $\Delta\chi_{\text{pauling}}$, $\Delta\chi_{\text{Mulliken}}$, ΔS_{mix} , ΔH_{mix}) (phase diagrams with the help of CALPHAD)	SVR	- The model exhibited a mean prediction error below 20% for yield strength at 20 °C and 600 °C, whereas the error exceeded 20% at 800 °C. - The model exhibited a mean prediction error below 16.6% for ductility at 20 °C, whereas the error exceeded 50% at 800 °C	To study and predict Yield strength and ductility at various temperatures (20 °C, 600 °C, 800 °C) for RHEAs	[19]
	Lightweight RHEAs (Al, Ti, Cr, Nb, V, Zr, Mo, and Hf)	A total data set of 92 alloys from 28 publications	11 parameters	RF, MLP, KNN, SVC, and LR	The SVC model showed a better prediction accuracy.	To study and predict the hardness and corrosion resistance of HEAs	[59]

Table 2. Cont.

Collection of datasets	HEA configuration	Size of dataset	Input Features	ML Model(s)	Performance Metrics	Predictions/Application Scenarios	References
Experimental/Literature	Eutectic HEAs (FeCoNiCrMnVNb)	A total data set of 477, in which 53 data sets (49 HEAs and 4 EHEAs collected from literature) were used for training	NM	ANN	ANN model showed a good correlation coefficient of 0.98	- To study and predict the microstructure of EHEAs - To study and predict the micro-hardness of EHEAs	[60]
	High Entropy Carbides (HECs)	A total data set of 177 (26 of HECs and 151 belong to ternary carbides) (Young's modulus and hardness)	14 parameters	RF, SVR, and ANN	All models achieved good accuracy, exceeding 0.92	To predict mechanical properties (Young's modulus and Hardness)	[17]
	High Entropy Oxide Ceramics (HEOs)	A data set of 112 samples of experimental data	5 parameters (r_A/r_C , δ , $\Delta\chi_{\text{pauling}}$, $\Delta\chi_{\text{Mulliken}}$, and ΔS_{mix})	XGBoost, NB, and RF	- XGBoost with accuracy of 98.7% (F1 score 0.975) - RF showed accuracy of 98.9% (F1 score 0.974) - NB showed accuracy of 98.1% (F1 score 0.93)	To study and predict crystal structures of HEOs	[61]
	(NbZrTiCr)B ₂ Stable Alkaline earth elements diboride (AlB ₂)	133 data sets in each class (SS and mixed)	6 parameters (r_A/r_C , δ , $\Delta\chi_{\text{pauling}}$, $\Delta\chi_{\text{Mulliken}}$, ΔS_{mix} , and ΔH_{mix})	KNN	- KNN with accuracy of 94.1% - Logistic regression showed an accuracy of 91.2% - Gaussian Naïve Bayes showed an accuracy of 91.1%	To predict the phase structure of high-entropy boride composition	[62]

Table 2. Cont.

Collection of datasets	HEA configuration	Size of dataset	Input Features	ML Model(s)	Performance Metrics	Predictions/Application Scenarios	References
Experimental/Literature	WC-HEAs (W, Co, C, Cr, Fe, Mo, Ti, Ni, Mn, Cu, and Al)	For Vickers hardness prediction, 75 WC-HEAs data and For fracture toughness prediction, 70 WC-HEAs data. Both datasets were collected from the literature	Alloy compositions	SVM, RF, and ANN	- For Vickers hardness prediction, the SVM model showed a prediction accuracy of 82.22%, and the remaining models showed above 70% - For the prediction of fracture toughness, the RF model showed a prediction accuracy of 83.5%, and the remaining models showed above 75%	To predict the Vickers hardness and fracture toughness of WC-HEAs	[63]
	NbTaW-based HEA	A total data set for training 62 alloys	36 features (15 elemental and 21 physical descriptors)	LR, KNN, BR, KR, RF, and SVM	SVM model showed a better prediction accuracy of 87.2 % with 12 features, while increasing the number of features to 13, the prediction accuracy improved to 92.37 %	To study and predict the hardness of RHEAs, mainly NbTaW-based HEAs	[64]
DFT	HEAs	A total dataset of 401 alloys by DFT. (174 SS, 54 IM, 173 SS+IM)	5 parameters (VEC, δ , $\Delta\chi$, ΔS_{mix} , ΔH_{mix})	KNN, SVM, and ANN	- The ANN model showed better predictive accuracy at 74.3%. - ANN model showed better prediction at SS and SS+IM	To study and predict the phase of HEAs.	[65]

Table 2. Cont.

Collection of datasets	HEA configuration	Size of dataset	Input Features	ML Model(s)	Performance Metrics	Predictions/Application Scenarios	References
DFT	HEAs	A total dataset of 1252 alloys using DFT. (625 single-phase and 627 multi-phase)	6 parameters	New model	New rules showed a less accurate 73%	To study and predict the phase of HEAs	[66]
MD	HEAs with Graphene	A total data set of 700	A raw data set containing 59 features	XGBoost, AdaBoost, LGBost, DT, and RF	- All models achieved an accuracy of more than 0.9 for Young's modulus - All models achieved an accuracy of more than 0.85 for toughness	To predict tensile properties (E, Y and Toughness)	[18]
	CoCuCrFeNi	A total data set of 1000 alloys	62 features derived from 6 variables	XGBoost, RF, DT, LGBost, and AdaBoost	Four of the five models achieved prediction accuracy of more than 90%. The XGBoost model showed the best prediction accuracy	To predict the Young's modulus and toughness of HEAs	[67]
CALPHAD	TiZrNbTa-based HEA	A total dataset of 100 alloy hardness values from the literature and CALPHAD	10 parameters (Physical and chemical properties)	XGBoost	The XGBoost model showed a better prediction accuracy of 87.6% in the prediction of RHEAs	To study and predict the hardness of refractory HEAs	[68]
Hybrid	High-entropy carbide ceramics (HECCs)	A total data set of 251 samples (171 from experimental and literature collection, and another 80 generated from the borderline-SMOTE method)	6 parameters (VEC, average metallic radius (r_{Me}), δ , $\Delta\chi_{pauling}$, $\Delta\chi_{Mulliken}$, ΔS_{mix} , and Z^*)	RF, SVM, LR, LOG and XGB	- RF with accuracy of 94.1% - LOG showed accuracy of 83.15% - Gaussian Naïve Bayes showed an accuracy of 83.15%	To predict the phase formation ability of HECCs	[16]

Table 2. Cont.

Collection of datasets	HEA configuration	Size of dataset	Input Features	ML Model(s)	Performance Metrics	Predictions/Application Scenarios	References
NR	MoNbTaTiW and HfMoNbTaTiZr	NR	7 parameters (VEC, B, ΔS_{mix} , Y, density, T_m and Testing temperature)	RF	RF model showed predicted yield strength at higher temperature (1200 °C) with low error for both alloys	To predict the yield strength of HEAs at high temperatures	[69]
NR	HEAs	NR	3 parameters (VEC, δ , and ΔH_{mix})	KNN, Multinomial Regression (MR), RF and XGBoost	XGBoost and RF models showed the best prediction accuracy, 72%	To study and predict the crystallite phase structure of HEAs	[70]

Abbreviations: HEAs—High Entropy Alloys, MPEA—Multi-Principal Element Alloy, HEC—High Entropy Carbide, HEO—High Entropy Oxide, HECC—High Entropy Carbide Ceramic, RHEA—Refractory High Entropy Alloy, EHEA—Eutectic High Entropy Alloy, SS—Solid Solution, IM—Intermetallic, AM—Amorphous, BCC—Body Centered Cubic, FCC—Face Centered Cubic, HCP—Hexagonal Close Packed, VEC—Valency Electron Concentration, δ —Atomic Size difference, $\Delta\chi$ —Electro Negativity difference, ΔS_{mix} —Mixing Entropy, ΔH_{mix} —Mixing Enthalpy, T_m —Melting Temperature, Ω —Solid Solution Prediction parameter, E—Young's Modulus, Y—Yield Strength, B—Bulk Modulus, δ_r —Atomic Radius Ratio, G—Shear Modulus, δG —Shear Modulus difference, μ —Lattice Distortion Energy, F—Peierls-Nabarro factor, A—energy term in strengthening model, RF—Random Forest, SVM—Support Vector Machine, SVR—Support Vector Regression, KNN—K-Nearest Neighbors, ANN—Artificial Neural Network, MLP—Multi-Layer Perceptron, BPNN—Back Propagation Neural Network, DNN—Deep Neural Network, DT—Decision Tree, LR—Logistic Regression, NB—Naive Bayes, GNB—Gaussian Naive Bayes, GBDT—Gradient Boosting Decision Tree, SVC—Support Vector Classifier, XGBoost/XGB—Extreme Gradient Boosting, AdaBoost—Adaptive Boosting, LGBM/LightGBM—Light Gradient Boosting Machine, CatBoost—Category Boosting, SGR—Stochastic Gradient Regression, BR—Bayesian Ridge, EL—Ensemble Learning, PSO—Particle Swarm Optimization, SMOTE—Synthetic Minority Over-Sampling Technique, MD—Molecular Dynamics, DFT—Density Functional Theory and NR—Not Reported.

4. Discussion

Table 2 shows an overall summary of the studies included in this review, presenting the reported data by data collection source and HEA configuration, then by dataset size, input features, machine learning models and prediction performance. This review is useful not only for identifying developments in machine learning methods in HEA research but also for understanding trends in dataset composition and researchers' preferences. One observation from Table 2 is that the same group of elements, namely Al, Fe, Co, Cr, Cu, Ni, and Mn, was repeated in different combinations or alone. In Phase prediction and Hardness estimation studies, AlCoCrCuFeNi, AlCoCrFeMnNi and similar compositions reappear [24,25]. These compositions are among the most thoroughly examined conventional HEA systems and have thus yielded relatively extensive experimental and literature data. The frequent choice of these is due to their known behaviour in phase formation and their capacity to provide a useful trade-off among strength, hardness, ductility, and phase stability. These well-studied works, therefore, provide a solid basis for data to develop and test machine learning models.

Although these are typical compositions, they also suggest a constraint in the existing data landscape: They tend to be reused. While the number of elements and alloy families available for HEA design is extremely large, many existing datasets cluster around a relatively small subset of them. This concentration can lead to models that are accurate in some well-investigated compositional regions but not in others. As the number of larger datasets in Table 2 increases, more elements such as Ti, V, Nb, Mo, Zr, Hf, Ta, W, and Si are added to the compositions, suggesting an attempt to broaden the compositional space [56]. Beyond these traditional HEA classes, there have been some studies on more specific and less-studied HEA classes. These are light-weight HEAs, refractory HEAs, light-weight refractory HEAs and eutectic HEAs. For example, machine learning has been explored to identify compositions of lightweight HEAs with densities below 6 g/cm³ and desirable phase characteristics [59]. Other refractory HEAs have been studied, including Nb, Ti, V, Zr, W, Ta, Hf, and Mo, with researchers focusing on hardness, phases, yield strength, high-temperature properties, and other factors [57]. Likewise, there is a body of literature on high-entropy carbides, oxides, borides, and other compositionally complex materials, and data-driven techniques are increasingly being applied to materials beyond the scope of conventional metallic HEAs.

Several widely varying machine learning models have been used to predict phases and mechanical properties and are listed in Table 2. The most common models in the reviewed studies are used either alone or in combination with other models, such as SVM/SVR, RF, and ANN [60]. They are applied repeatedly, highlighting the need to account for the non-linear relationships among alloy composition, thermodynamic/physicochemical descriptors, microstructure, and mechanical properties. To compare several algorithms on the same dataset, several studies have used SVM, RF, DTs, KNNs, LR, gradient boosting, AdaBoost, and XGBoost. Comparative approaches such as these can help determine which model yields the highest predictive accuracy on a specific dataset, rather than assuming that a single model will work for every dataset in the HEA prediction problem. The rise in the use of ensemble and boosting models such as RF, XGBoost, and gradient boosting reflects a trend towards modelling complex nonlinear relationships between composition-based and derived material descriptors. Optimisation techniques, such as PSO, have also been used in some studies to enhance prediction and identify the most promising alloy compositions when combined with traditional ML models [25].

An important thing to note from Table 2 is the relationship between the characteristics of the available dataset and model performance. Models with limited data sets, such as studies, typically use SVMs, ANNs, or other models that can learn nonlinear relationships from the limited data [27]. In contrast, larger datasets in the literature can enable more extensive comparisons among ensemble, boosting, neural network, and conventional statistical models. Comparisons should be made with caution, however, because reported performance values are not consistent across studies due to differences in data, data descriptors, targeted properties, data splitting, and evaluation methods. Thus, the reported accuracy of a single model is not necessarily indicative of its overall superiority over models built on other datasets. It illustrates the significance of data size, compositional diversity, feature selection, validation method, and model interpretability in evaluating the reliability of ML-based HEA prediction.

In conclusion, Table 2 shows the increasing use of ML across HEA systems and prediction tasks. However, research remains focused on a limited number of established HEA systems, while differences in datasets and evaluation methods make direct comparison difficult. Larger and more diverse datasets, consistent validation, and robust feature and model selection are therefore needed to support the reliable application of ML in next-generation HEA design.

5. Conclusion

This review summarises the application of machine learning to phase prediction and mechanical property estimation in HEAs, with particular emphasis on microstructure, hardness, and yield strength. The literature shows growing use of data-driven approaches to relate alloy composition, thermodynamic and physicochemical descriptors, phase constitution, microstructure, and mechanical properties. Several ML models have been used, including SVMs, RFs, ANNs, and boosting-based algorithms, and some studies have demonstrated their ability to accurately predict and accelerate the development of novel alloys. Based on an analysis of existing data, the research primarily focuses on conventional HEA systems containing Al, Co, Cr, Cu, Fe, Ni, and Mn. These HEA systems have produced relatively large datasets from experiments and literature, which served as a solid basis for creating ML models. However, this focus leaves much of the broader HEA compositional space unexplored. Studies on lightweight, refractory, eutectic, and specialised HEAs show that ML can also be applied to these systems, although the available datasets are still limited. In addition, the review highlights that reported accuracy alone is not sufficient to evaluate ML performance. The size of the dataset, the diversity of its composition, the choice of input features, the target property, the chosen data-splitting strategy and the validation method play a significant role in the predictions. Therefore, it is important to consider the results from this model when comparing results from other studies. Research on ML-based approaches for HEAs would be more reliable and reproducible if greater emphasis were placed on independent validation, consistent evaluation metrics, and transparent reporting of dataset characteristics. In summary, machine learning is a promising tool for accelerating phase prediction, mechanical property prediction, and composition screening in HEAs. Future studies will benefit from compositionally diverse, higher-quality and larger datasets, especially for underrepresented HEA classes and for unexplored compositions. The combination of machine learning with experimental validation, physically meaningful descriptors, interpretable modelling, and optimisation-based alloy design can further enhance the confidence of data-driven approaches and aid in the design of high-performance HEAs with desired properties.

Declaration of generative AI and AI-assisted technologies

During the preparation of this manuscript, the authors used generative AI tools only to improve language and readability. Specifically, the authors used Grammarly and ChatGPT solely to polish the language when drafting some sections of the manuscript. The authors take full responsibility for the content of the manuscript.

Acknowledgments

The authors gratefully acknowledge the MHRD, Government of India, for providing research scholarship support to J Sharath Kumar.

Authors' contribution

Conceptualization, R.V.; methodology, J.S.K. and R.V.; investigation, J.S.K.; resources, R.K. and R.V.; writing—original draft preparation, J.S.K.; writing—review and editing, J.S.K. and R.V.; supervision, R.K. and R.V. All authors have read and agreed to the published version of the manuscript.

Conflicts of interest

Rajeev Verma holds the position of Guest Editor for *Advanced Manufacturing* and has not peer-reviewed or made any editorial decisions for this paper. The authors declare no other conflicts of interest.

Appendix A

Systematic search terms:

("High Entropy Alloy" OR "HEA" OR "Multi-principal element alloy" OR "Complex Concentrated Alloy*") AND

("Machine Learning" OR "Deep Learning" OR "Neural Network*" OR "Artificial Intelligence" OR "AI" OR "Data-driven") AND

("Solid Solution" OR "Phase Prediction" OR "Mechanical Property*" OR "Hardness" OR "Yield Strength")

References

- [1] Kumar J, Rao DK, Kumar P. Composite manufacturing: properties, application, challenges, and future trend. In *Industry 4.0 in Composite Manufacturing Industry for Sustainable Development*. Boca Raton: CRC Press, 2026. pp. 344–371.
- [2] He C, Zhang Z, Wang B, Zhang H, Ren Z, *et al.* Twinning behavior and texture evolution of pure titanium during corrugated-flat rolling under wide temperature range. *J. Alloys Compd.* 2025, 024:180198.
- [3] Murty BS, Yeh JW, Ranganathan S, Bhattacharjee PP. High-entropy alloys: basic concepts. In *High-Entropy Alloys*. Amsterdam: Elsevier, 2019. pp. 13–30.

- [4] Kumar JS, Kishore KN, Kumar R, Verma R. Evaluation of microstructure and phase formation in nanocrystalline FeCoCuNbMo high-entropy alloy synthesised by mechanical alloying. *Faraday Discuss.* 2026, 264:132–150.
- [5] Chen J, Zhou X, Wang W, Liu B, Lv Y, *et al.* A review on fundamental of high entropy alloys with promising high-temperature properties. *J. Alloys Compd.* 2018, 760:15–30.
- [6] Wang X, Guo W, Fu Y. High-entropy alloys: emerging materials for advanced functional applications. *J. Mater. Chem. A* 2021, 9(2):663–701.
- [7] Murty BS, Yeh JW, Ranganathan S, Bhattacharjee PP. Alloy design and phase selection rules in high-entropy alloys. In *High-Entropy Alloys*. Amsterdam: Elsevier, 2019. pp. 51–79.
- [8] Sivaraman S, Radhika N, Abubaker Khan M. Machine learning-driven prediction of wear rate and phase formation in high entropy alloy coatings for enhanced durability and performance. *IEEE Access* 2025, 13:33956–33975.
- [9] Anupam A, Kumar P, Singh H. Microstructure and mechanical properties of high-entropy alloy coatings. In *High-Entropy Alloy Coatings*. Boca Raton: CRC Press, 2025. pp. 61–92.
- [10] Xie L, Brault P, Thomann AL, Yang X, Zhang Y, *et al.* Molecular dynamics simulation of Al-Co-Cr-Cu-Fe-Ni high entropy alloy thin film growth. *Intermetallics* 2016, 68:78–86.
- [11] Zhang C, Gao MC. CALPHAD modeling of high-entropy alloys. In *High-Entropy Alloys: Fundamentals and Applications*. Cham: Springer International Publishing, 2016. pp. 399–444.
- [12] Biermair F, Razumovskiy VI, Ressel G. Influence of alloying on thermodynamic properties of AlCoCrFeNiTi high entropy alloys from DFT calculations. *Comput. Mater. Sci.* 2022, 202:110952.
- [13] Machaka R, Motsi GT, Raganya LM, Radingoana PM, Chikosha S. Machine learning-based prediction of phases in high-entropy alloys: a data article. *Data Brief* 2021, 38:107346.
- [14] Mahanta BK, Rawat P, Kumar S, Singh VK, Singh SK. Experimental and machine learning approaches to phase prediction in high entropy alloy and medium entropy alloy. *Mater. Chem. Phys.* 2026, 347:131403.
- [15] Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, *et al.* The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 2021, 372:n71.
- [16] Zhang X, Sun J, Zhang Y, Fan K, Zhang Z, *et al.* Machine learning for phase prediction of high entropy carbide ceramics from imbalanced data. *npj Comput. Mater.* 2026, 12(1):22.
- [17] Zhou Q, Xu F, Gao C, Zhang D, Shi X, *et al.* Machine learning-assisted mechanical property prediction and descriptor-property correlation analysis of high-entropy ceramics. *Ceram. Int.* 2023, 49(4):5760–5769.
- [18] Wu Q, Gao T, Liu G, Ma Y. Machine learning-assisted prediction of mechanical properties of high-entropy alloy/graphene nanocomposite. *Mater. Today Commun.* 2024, 40:109663.
- [19] Klimenko D, Stepanov N, Li J, Fang Q, Zharebtsov S. Machine learning-based strength prediction for refractory high-entropy alloys of the Al-Cr-Nb-Ti-V-Zr system. *Materials* 2021, 14(23):7213.
- [20] Hosseini S, Vaghefi E, Lee C, Prorok B, Mirkoochi E. Machine learning-enabled prediction and optimization of hardness for Nb-Ti-V-Zr refractory high entropy alloy. *Mater. Today Commun.* 2024, 40:109607.
- [21] Mandal P, Choudhury A, Mallick AB, Ghosh M. Phase prediction in high entropy alloys by various machine learning modules using thermodynamic and configurational parameters. *Metals Mater. Int.* 2023, 29(1):38–52.

- [22] Ye Y, Li Y, Ouyang R, Zhang Z, Tang Y, *et al.* Improving machine learning based phase and hardness prediction of high-entropy alloys by using Gaussian noise augmented data. *Comput. Mater. Sci.* 2023, 223:112140.
- [23] Ren W, Zhang Y, Wang W, Ding S, Li N. Prediction and design of high hardness high entropy alloy through machine learning. *Mater. Des.* 2023, 235:112454.
- [24] Chang YJ, Jui CY, Lee WJ, Yeh AC. Prediction of the composition and hardness of high-entropy alloys by machine learning. *JOM* 2019, 71(10):3433–3442.
- [25] Chen C, Ma L, Zhang Y, Liaw PK, Ren J. Accelerating the design of high-entropy alloys with high hardness by machine learning based on particle swarm optimization. *Intermetallics* 2023, 154:107819.
- [26] Paturi UMR, Ishtiaq M, Lakshmi Narayana P, Maurya AK, Choi SW, *et al.* Evaluating machine learning models for predicting hardness of AlCoCrCuFeNi high-entropy alloys. *Crystals* 2025, 15(5):404.
- [27] He Z, Zhang H, Cheng H, Ge M, Si T, *et al.* Machine learning guided BCC or FCC phase prediction in high entropy alloys. *J. Mater. Res. Technol.* 2024, 29:3477–3486.
- [28] Li T, Zhang M, Xu H, Wang Z, Zhao X. Phase identification and hardness prediction of AlCrFeNiCuSi high-entropy alloys guided by explainable machine learning. *JOM* 2026, 78(1):697–710.
- [29] Bai F, Huang Y, Wang J, Zhou H, Wu X, *et al.* Machine learning of hardness and composition in Al-Co-Cr-Fe-Ni-Mo-Ti high-entropy alloys. *Mater. Chem. Phys.* 2025, 343:131068.
- [30] Jaiswal UK, Vamsi Krishna Y, Rahul MR, Phanikumar G. Machine learning-enabled identification of new medium to high entropy alloys with solid solution phases. *Comput. Mater. Sci.* 2021, 197:110623.
- [31] Qu N, Liu Y, Zhang Y, Yang D, Han T, *et al.* Machine learning guided phase formation prediction of high entropy alloys. *Mater. Today Commun.* 2022, 32:104146.
- [32] Liu G, Wu Q, Ma Y, Huang J, Xie Q, *et al.* Machine learning-based phase prediction in high-entropy alloys: further optimization of feature engineering. *J. Mater. Sci.* 2025, 60(8):3999–4019.
- [33] Li X, Sun J, Chen X. Machine learning-based prediction of high-entropy alloy hardness: design and experimental validation of superior hardness. *Trans. Indian Inst. Metals* 2024, 77(11):3973–3981.
- [34] Wu S, Song Z, Wang J, Niu X, Chen H. Enhanced phase prediction of high-entropy alloys through machine learning and data augmentation. *Phys. Chem. Chem. Phys.* 2025, 27(2):717–729.
- [35] Dey D, Das S, Pal A, Dey S, Raul CK, *et al.* Improved machine learning framework for prediction of phases and crystal structures of high entropy alloys. *J. Alloys Metall. Syst.* 2025, 9:100144.
- [36] Gao S, Gao Z, Zhao F. Machine learning and visualization assisted solid solution strengthening phase prediction of high entropy alloys. *Mater. Today Commun.* 2023, 35:105894.
- [37] Huang X, Jin C, Zhang C, Zhang H, Fu H. Machine learning assisted modelling and design of solid solution hardened high entropy alloys. *Mater. Des.* 2021, 211:110177.
- [38] Wang J, Liu K, Lei Z, Li X, Liu L, *et al.* Machine-learning-assisted phase prediction in high-entropy alloys using two-step feature selection strategy. *Acta Metall. Sin. (Engl. Lett.)* 2025, 38(8):1261–1274.
- [39] Risal S, Zhu W, Guillen P, Sun L. Improving phase prediction accuracy for high entropy alloys with machine learning. *Comput. Mater. Sci.* 2021, 192:110389.

- [40] Jain R, Jain S, Dewangan SK, Boriwal LK, Samal S. Machine learning-driven insights into phase prediction for high entropy alloys. *J. Alloys Metall. Syst.* 2024, 8:100110.
- [41] Chang H, Tao Y, Liaw PK, Ren J. Phase prediction and effect of intrinsic residual strain on phase stability in high-entropy alloys with machine learning. *J. Alloys Compd.* 2022, 921:166149.
- [42] Mishra A, Kompella L, Sanagavarapu LM, Varam S. Ensemble-based machine learning models for phase prediction in high entropy alloys. *Comput. Mater. Sci.* 2022, 210:111025.
- [43] Gao J, Wang Y, Hou J, You J, Qiu K, *et al.* Phase prediction and visualized design process of high entropy alloys via machine learned methodology. *Metals* 2023, 13(2):283.
- [44] Zhao D, Pan SP, Zhang Y, Liaw PK, Qiao J. Structure prediction in high-entropy alloys with machine learning. *Appl. Phys. Lett.* 2021, 118(23):231904.
- [45] Hou S, Sun M, Bai M, Lin D, Li Y, *et al.* A hybrid prediction frame for HEAs based on empirical knowledge and machine learning. *Acta Mater.* 2022, 228:117742.
- [46] Choudhury A, Kumar S. Breaking data barriers: advancing phase prediction in high entropy alloys through a new machine learning framework. *Can. Metall. Q.* 2025, 64(3):903–913.
- [47] Peivaste I, Jossou E, Tiamiyu AA. Data-driven analysis and prediction of stable phases for high-entropy alloy design. *Sci. Rep.* 2023, 13(1):22556.
- [48] Gao Z, Zhao F, Gao S, Xia T. Machine learning prediction of hardness in solid solution high entropy alloys. *Mater. Today Commun.* 2023, 37:107102.
- [49] Hou S, Li Y, Bai M, Sun M, Liu W, *et al.* Phase prediction of high-entropy alloys by integrating criterion and machine learning recommendation method. *Materials* 2022, 15(9):3321.
- [50] Machaka R. Machine learning-based prediction of phases in high-entropy alloys. *Comput. Mater. Sci.* 2021, 188:110244.
- [51] Bundela AS, Rahul MR. Machine learning-enabled framework for the prediction of mechanical properties in new high entropy alloys. *J. Alloys Compd.* 2022, 908:164578.
- [52] Bobbili R, Ramakrishna B. Prediction of phases in high entropy alloys using machine learning. *Mater. Today Commun.* 2023, 36:106674.
- [53] Bakr M, Syarif J, Hashem IAT. Prediction of phase and hardness of HEAs based on constituent elements using machine learning models. *Mater. Today Commun.* 2022, 31:103407.
- [54] Li S, Liu R, Yan H, Li Z, Li Y, *et al.* Machine learning phase prediction of light-weight high-entropy alloys containing aluminum, magnesium, and lithium. *Metals* 2024, 14(4):400.
- [55] Ding S, Wang W, Zhang Y, Ren W, Weng X, *et al.* A yield strength prediction framework for refractory high-entropy alloys based on machine learning. *Int. J. Refract. Metals Hard Mater.* 2024, 125:106884.
- [56] Qu N, Zhang Y, Liu Y, Liao M, Han T, *et al.* Accelerating phase prediction of refractory high entropy alloys via machine learning. *Phys. Scr.* 2022, 97(12):125710.
- [57] Wang Y, Liao M, Qu N, Li X, Yang D, *et al.* Machine learning assisted phase prediction of refractory multi-component alloys based on data from low-, medium-, and high-entropy alloys. *Int. J. Refract. Metals Hard Mater.* 2025, 132:107263.
- [58] Li L, Wang S, Su C, Guo S. Machine learning assisted design of CrTaTiV-based refractory high entropy alloys with enhanced hardness. *Int. J. Refract. Metals Hard Mater.* 2025, 133:107400.
- [59] Gao T, Gao J, Yang S, Zhang L. Data-driven design of novel lightweight refractory high-entropy alloys with superb hardness and corrosion resistance. *npj Comput. Mater.* 2024, 10(1):256.

- [60] Jain R, Dewangan SK, Umre P, Kumar V, Samal S. Microstructure evolution and an ANN approach for microhardness prediction of suction cast FeCoNiCrMnVNb eutectic high-entropy alloys. *Trans. Indian Inst. Metals* 2021, 74(11):2671–2679.
- [61] Liu J, Wang A, Gao P, Bai R, Liu J, *et al.* Machine learning-based crystal structure prediction for high-entropy oxide ceramics. *J. Am. Ceram. Soc.* 2024, 107(2):1361–1371.
- [62] Mitra R, Bajpai A, Biswas K. Machine learning based approach for phase prediction in high entropy borides. *Ceram. Int.* 2022, 48(12):16695–16706.
- [63] Fan C, Sun J, Ning S, Li X, Zhao J. Prediction of hardness and fracture toughness in high-entropy alloy bonded cemented carbides through machine learning. *J. Mater. Eng. Perform.* 2026, 35(2):1073–1087.
- [64] Deng D, Han J, Wen H, Tong Y. Hardness prediction of NbTaW-based high-entropy alloys with machine learning. *J. Mater. Eng. Perform.* 2026, 35(27):28200–28207.
- [65] Huang W, Martin P, Zhuang HL. Machine-learning phase prediction of high-entropy alloys. *Acta Mater.* 2019, 169:225–236.
- [66] Pei Z, Yin J, Hawk JA, Alman DE, Gao MC. Machine-learning informed prediction of high-entropy solid solution formation: beyond the Hume-Rothery rules. *npj Comput. Mater.* 2020, 6(1):50.
- [67] Gao T, Wu Q, Chen L, Liang Y, Han Y. Prediction of mechanical properties of high entropy alloys based on machine learning. *Phys. Scr.* 2025, 100(4):046013.
- [68] Sun Y, Lu Z, Liu X, Du Q, Xie H, *et al.* Prediction of Ti-Zr-Nb-Ta high-entropy alloys with desirable hardness by combining machine learning and experimental data. *Appl. Phys. Lett.* 2021, 119(20):201905.
- [69] Bhandari U, Rafi MR, Zhang C, Yang S. Yield strength prediction of high-entropy alloys using machine learning. *Mater. Today Commun.* 2021, 26:101871.
- [70] Oñate A, Sanhueza JP, Zegpi D, Tuninetti V, Ramirez J, *et al.* Supervised machine learning-based multi-class phase prediction in high-entropy alloys using robust databases. *J. Alloys Compd.* 2023, 962:171224.