Rational design of high-entropy ceramics based on machine learning – A critical review

Zhang, Jun; Xiang, Xuepeng; Xu, Biao; Huang, Shasha; Xiong, Yaoxu; Ma, Shihua; Fu, Haijun; Ma, Yi; Chen, Hongyu; Wu, Zhenggang; Zhao, Shijun

Published in:
Current Opinion in Solid State and Materials Science

Published: 01/04/2023

Document Version:
Post-print, also known as Accepted Author Manuscript, Peer-reviewed or Author Final version

Published version (DOI):
10.1016/j.cossms.2023.101057

Publication details:
Zhang, J., Xiang, X., Xu, B., Huang, S., Xiong, Y., Ma, S., Fu, H., Ma, Y., Chen, H., Wu, Z., & Zhao, S. (2023). Rational design of high-entropy ceramics based on machine learning – A critical review. *Current Opinion in Solid State and Materials Science*, *27*(2), Article 101057. https://doi.org/10.1016/j.cossms.2023.101057

Rational design of high-entropy ceramics based on machine learning – a critical review

Jun Zhang[1], Xuepeng Xiang[1], Biao Xu[1], Shasha Huang[1], Yaoxu Xiong[1], Shihua Ma[1], Haijun Fu[1], Yi Ma[2], HongYu Chen[3], Zhenggang Wu[4], Shijun Zhao[1,*]

[1] Department of Mechanical Engineering, City University of Hong Kong, Hong Kong, 999077, China.

[2] Key Laboratory for Light-Weight Materials, Nanjing Tech University, Nanjing, Jiangsu, China.

[3] College of Mechanical Engineering, Zhejiang University of Technology, Hangzhou, China.

[4] College of Materials Science and Engineering, Hunan University, Changsha, Hunan, China.

* Corresponding author: *shijzhao@cityu.edu.hk*

Highlights

- The unique machine-learning workflow used for the rational design of high-entropy ceramics is summarized and discussed.
- Current applications of machine learning in the design of high-entropy ceramics are systematically reviewed and future opportunities are highlighted.
- The bottleneck of present machine-learning applications is identified as the shortage of available database and data imbalance, as well as the limitation of model generalization.

Abstract

High-entropy materials provide a versatile platform for the rational design of novel candidates with exotic performances. Recently, it has been demonstrated that high-entropy ceramics (HECs), depending on their compositions, show great application potential because of their superior structural and functional properties. However, the immense phase space behind HECs significantly hinders the efficient design and exploitation of high-performance HECs through traditional trial-and-error experiments and expensive *ab-initio* calculations. Machine learning (ML), on the other hand, has become a popular approach to accelerate the discovery of HECs and screen HECs with exceptional properties. In this article, we review the recent progress of ML applications in discovering and designing novel HECs, including carbides, nitrides, borides, and oxides. We thoroughly discuss different ingredients that are involved in ML applications in HECs, including data collection, feature engineering, model refinement, and prediction performance improvement. We finally provide an outlook on the challenges and development directions of future ML models for HEC predictions.

Keywords: Machine Learning, High-Entropy Ceramics, Phase Stability, Mechanical Properties, Deep Learning, Single-Phase Synthesizability

1. Introduction

Ceramics are one of the earliest materials that human produced for their daily usage because of their simple forge procedures and high mechanical strength. In modern society, new techniques render them even superior properties. Carbides, borides, and nitrides with high hardness are widely used as cutting tools, thermal and electric coatings, etc. [1, 2]. On the other hand, emerging technologies also put forward higher requirements for future ceramics. For example, as one of the cleanest forms of energy, nuclear energy produced in the next-generation reactors will need to operate under high temperatures, high pressures, and more destructive irradiation damages [3-7]. Novel ceramic materials, either as the fuel cladding or structural materials, should be developed by incorporating multiple merits to withstand these extreme service conditions.

To achieve these goals, not only do the conventional materials need to be fine-tuned, but revolutionized principles for the rational design of materials should be pursued. The concept of high-entropy alloys (HEAs), which consist of more than four principal components [8] or with configurational mixing entropy higher than 1.61R (R: the gas constant) [9], was independently introduced in 2004 by Yeh et al. [8] and Cantor et al. [10]. They found that the increased number of components helps stabilize multiple element alloys, even containing several conventionally immiscible components. Thus, the vast composition space provides ample opportunities and advanced platforms for tuning the chemical and mechanical properties of HEAs. Because of the so-called cocktail effects after mixing, HEAs usually exhibit exceptional properties compared to their constitutive components, attracting extensive attention in the last decades [11, 12].

Inspired by the idea of HEAs, the high-entropy nitride film (Fe-Co-Ni-Cr-Cu-Al-Mn-N-O) was deposited on the HEAs with low crystallinity by reactive sputtering [13]. Later, bulk ceramic materials stepped into a high-entropy era in 2015 when Rost et al. [14] introduced the concept of entropy stabilization to multi-component oxides. By delicately selecting the precursory oxides with distinct crystal structures, electronegativity, or cation coordination, the single-phase stability of (Mg,Co,Ni,Cu,Zn)O and its subsystems were investigated through changing their components and concentrations. Their results show that only the oxide with five cations is a single phase, and the departure from equiatomic metallic compositions increases the transformation temperature from multiple phases to a single phase, signifying the decreased single-phase stability. Afterward, high-entropy borides [15], carbides [16-20], nitrides [21], sulfides [22], oxynitrides [23, 24], and carbonitrides [25, 26] were experimentally synthesized. These newly developed ceramics show great potential in irradiation resistance [5, 27-29], thermoelectricity [22], superconductivity [30],

cutting machines [1, 2], coatings [1, 2], catalyst [23, 24, 31], and energy storage [32-34]. For example, the sluggish diffusion of species induced by the rugged energy surface in high-entropy oxides makes them promising to be used as hydrogen-resistant materials [27]. Besides, it is found that the distortion of local geometry in $(Zr_{0.25}Ta_{0.25}Nb_{0.25}Ti_{0.25})C$ can effectively hinder the growth of dislocation loops, deferring void formation and elemental segregation [5]. The properties, syntheses, characterizations, and applications of HECs have been thoroughly discussed in previous reviews [1, 2, 35-38].

Despite the excellent performance of HECs, the huge phase space behind HECs makes their phase exploration with traditional trial-and-error experiments nearly impossible. Traditional rules fail to distinguish single-phase high-entropy carbide ceramics (HECCs) [39]. For example, Zhao [40] reported that the lattice distortion (δ) might be unsuitable for predicting the single-phase synthesizability of HECs directly. Zhang et al. [39] showed that the single-phase synthesizability of HECCs cannot be underpinned by the mixing enthalpy or the Ω-δ map ($\Omega = \frac{T_m \Delta S}{\Delta H}$, where T_m is the melting temperature). Although a high mixing entropy is beneficial to stabilize single-phase high-entropy compounds [8, 14], it is not the governing factor in most cases [41]. For example, Otto et al. [41] fabricated series of HEAs with five principal components and found that replacing one element in CoCrFeMnNi (keeping the number of components) can form multi-phase materials. For HECs, distinct phase formation behaviors were also reported even with the same number of components at the cation sublattice [17, 42]. The limited power of mixing entropy is attributed to the incomplete description of the system randomness, where local ordering, vibrational, electronic, and magnetic moment disordering are totally ignored. Therefore, increasing mixing entropy alone is inadequate to achieve stable single-phase structures. Besides the difficulties in predicting the phase formation, exploiting the mechanical properties of HECs is also challenging owing to the vast phase space involved. One may intuitively infer the properties of HECs from their binary precursors. However, experimental studies have shown that the properties of HECs may surpass the rule-of-mixture (ROM) values [25, 26, 43], rendering the conventional experience unreliable in high-entropy scenarios. Tang et al. [44] correlated the mechanical properties of HECs with bonding parameters, and they found large deviations from the fitted scaling relations. Although theoretical simulations are powerful tools to aid the design of advanced HECs, accurate *ab-initio* calculations are too expensive to afford for exploring the diverse phase compositions in HECs [45, 46]. Cluster expansion is an efficient way to obtain the total energy by normally incorporating fewer than 5-body interactions, significantly accelerating the energy prediction of compositionally disordered systems. Though, heavy ab-initio calculations are required to reach accurate interaction parameters

for a variety of many-body configurations due to chemical complexity [32]. What's more, empirical interatomic potentials for describing HECs are extremely scarce due to their complicated bonding interactions. These new challenges significantly impede the discovery and design of high-performance HECs.

Recently, material science has embraced machine learning (ML) methods [47], such as artificial neural network (ANN) [48-50], random forest (RF) [51-53], support vector machine (SVM) [50], residual network (RESNET) [54, 55], etc. These algorithms provide an alternative way for rapid designing and high-throughput prediction of novel materials. ML can correlate the input descriptors and the output labels in a non-linear manner within a high-dimensional space, making it more accurate than conventional rule predictions from linear combinations of traditional descriptors. The applications of machine learning in material science generally have been reviewed by Morgan et al. [56] and Schmidt et al. [57]. Previous reviews of the machine learning application for developing high-entropy materials are mainly focused on HEAs [55, 58-64]. A timely examination of machine-learning applications for HECs is thus of great significance.

In this work, we provide a comprehensive review of the applications of the machine learning approach in the field of HECs. Firstly, the general workflow of HEC prediction with ML models is briefly discussed. Next, the shortage of available databases and current data sources from literature are introduced. Then, we highlight the feature engineering including feature selection and input pretreatments. Afterward, we discuss the methods associated with model generalization, model interpretability, and transferability. Additionally, applications of different ML models for predicting HECs are discussed. In the end, we summarize the challenges in the current ML approach and provide our outlook on future development.

2. Machine learning workflow

The workflow of ML for predicting HECs is shown in Fig. 1, which includes data collection, database organization, feature engineering, model training and selection, model application, and experimental validation. For the supervised ML training for HEC, the data should first be collected from literature or generated by new simulations/experiments. Then the collected data points are processed in an organized format. In this procedure, the collected dataset needs to be cleaned, in which HEC samples with problematic input features or labels should be removed. For example, the partially reported experimental details may lead to datums with missing input dimensions, which cannot be passed into a supervised model. Also, ill-converged data from simulations may mislead

the ML models. After data collection, two different treatments for these data samples can be conducted. The database can directly be passed into ML models by coding with numeric, such as atomic numbers or one-hot code. Alternatively, the properties of collected samples can be transformed with feature engineering, which helps to guide the ML models. Here, feature engineering comprises feature selection and pretreatments of input data. With the input dataset, ML models with varied hyperparameters are trained and evaluated, and the best one with well-trained parameters is deployed for prediction. The predicted HECs are filtered, and selected predictions can finally be synthesized and characterized experimentally for validation purposes. These newly synthesized samples can be augmented into the database and used to improve the ML performance further. In the following, we discuss the current status of each procedure specifically applied to HECs.

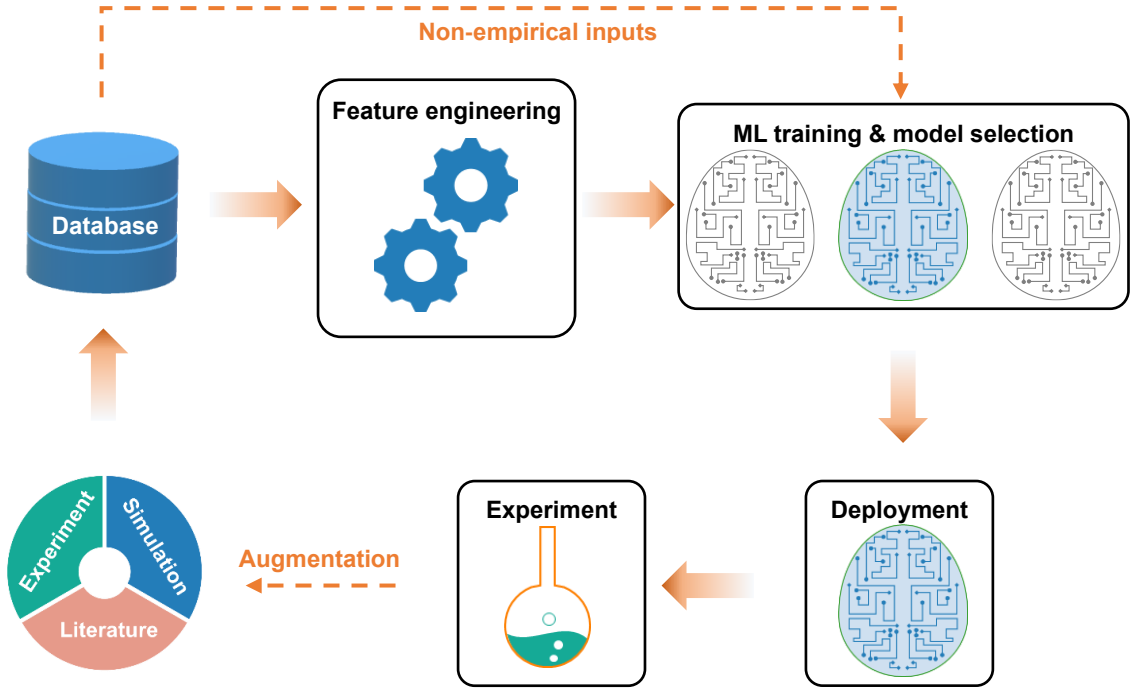

Fig. 1. Flowchart of machine-learning applied for high-entropy ceramics.

3. Database construction

As a data-driven approach, ML longs for a high-quality database. The size of the available dataset has been the bottleneck of ML applications in materials science for a long period. The materials genome initiative launched in 2011 aims to double the speed of material discovery, manufacture, and deployment in a shorter time [65, 66]. One of its key infrastructures is the open database. Since then, various material databases have been constructed, such as Materials Project [67], AFLOWLIB

[68], Open Quantum Material Database [69, 70], Open Catalysts Project [71], etc. However, these datasets are largely built from ordered or simple structures. A database incorporating chemically-disordered high-entropy materials is still unavailable. The shortage of the HEC database can also be partially attributed to the short time since its first discovery. In fact, there are only hundreds of experimentally verified equal or nearly equal metallic composition compounds up to date [35]. Therefore, ML applications for HECs are limited by the small dataset available. Because of this, the ML-aided approach bears great potential in exploring the phase space of HECs.

Though we are at a very early stage in database construction for HECs, several reviews [1, 2, 35-38] have collected most of the published HECs. For example, Akrami et al. [35] listed more than 700 published HECs until February 2021, among which ~300 materials are high-entropy oxide ceramics (HEOCs). The structures of HEOCs can be spinel, perovskite, fluorite, rock-salt, amorphous, or orthorhombic. In contrast, the vast majority of discovered carbides and di-borides are found with rock-salt and hexagonal structures, respectively. Such diverse structures for HECs also pose a great challenge for ML predictions, and most current ML models are mainly focused on one type of HECs. Another key challenge of building a high-quality database lies in the shortage of multi-phase samples [11, 39, 51], as researchers tend to consider materials with mixed phases as negative results and ignore them.

4. Feature engineering

4.1 Feature selection

Before passing the HEC compositions into an ML model, compositional information should be parameterized with continuous or discrete variables, so that the ML models can map the input features to the output labels. Depending on the pathways in obtaining the input features/descriptors, input features can be classified into three categories (as shown in Fig. 2).

This first type of input feature is experience independent, in which the compositions are represented by numerics only. For example, the one-hot code represents the categorical variables with zeros and ones. The dimension of each sample is the number of elements in the system or the length of the periodic table. For the *i*-th chemical element, the *i*-th value in this vector is set to be 1, and all other numbers are set to be zero. Taking a system containing three elements as an example, the precursory element can be codified with [1 0 0], [0 1 0], and [0 0 1], respectively. Then, these three vectors can be weighted with elemental concentrations to represent the composition. As this representation delivers little physical and chemical information, a large number of samples and a complicated

model are needed to digest the training data. The second type of input feature is atomic/precursory-based descriptors, such as atomic radius, mass, electronegativity, and valence electron concentration (*VEC*). These properties can be obtained directly from the periodic table, where no additional calculations are required. Note that these features are widely used in ML models for designing HECs [39, 51, 72]. The third type of feature is more expensive than the other two as it requires costly experiments or simulations, such as the melting temperature of HECs, mixing enthalpy, lattice distortion, entropy forming ability (*EFA*), and energy above the hull. These variables are inappropriate for the input because they are expensive to obtain when generalizing the ML model in predicting new materials. In fact, some of these features were used as the labels of HECs because they can deliver sophisticated information on the target HECs. For example, Kaufmann et al. [51] predicted the *EFA* values with the ML method, in which HECCs with *EFA* lower than 45 atom/eV are regarded as multi-phase materials.

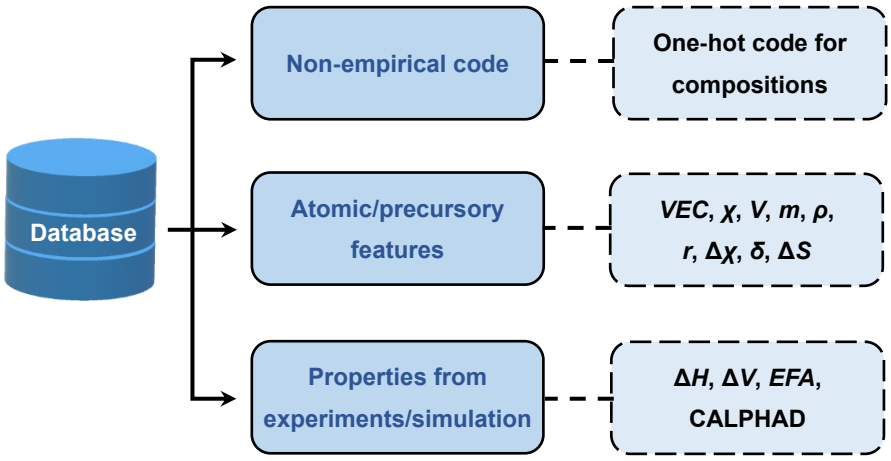

Fig. 2. Three different ways to obtain the input features. The full name of each descriptor can be found in Table 1.

All these descriptors commonly used in ML models for predicting HECs are listed in Table 1. The atomic radius difference is calculated by:

$$\delta = \sqrt{\sum_{i=1}^{n} c_i \left(1 - \frac{r_i}{\bar{r}}\right)^2}, \quad (1)$$

where c_i, r_i, and $\bar{r}$ are the concentration of i-th precursor, the atomic radius of i-th precursor, and the average atomic radius ($\bar{r} = \sum_{i=1}^{n} c_i r_i$), respectively. For HECCs, it is reported that the local lattice distortion affects the miscibility by lowering the mixing enthalpy [40]. Also, defect formation energies and mechanical properties are modulated by charge redistribution induced by local

distortions [40]. Similar to the average atomic radius and radius difference, other atomic/precursory properties (*VEC*, χ, *V*, *m*, ρ, and *r*) can be converted as the averages and the deviations. The average properties are calculated as:

$$\overline{prop} = \sum_{i=1}^{n} c_i prop_i, \quad (2)$$

where *prop* denotes precursory properties. These average values can capture the HEC properties based on the mean-field concept. Besides, the deviations can be used to quantify the differences among precursors, which can be calculated by:

$$\delta_{prop} = \sqrt{\sum_{i=1}^{n} c_i (1 - \frac{prop_i}{\overline{prop}})^2}. \quad (3)$$

Note that the electronegativity difference ($\Delta\chi$) is slightly different from Equation (3):

$$\Delta\chi = \sqrt{\sum_{i=1}^{n} c_i (\chi_i - \bar{\chi})^2}. \quad (4)$$

The mixing entropy is usually calculated by:

$$\Delta S = -R \sum_{i=1}^{n} c_i \ln c_i, \quad (5)$$

where *R* is the gas constant. The ΔS will be maximized with an equal atomic ratio. Additionally, property changes introduced by mixing (such as ΔH and ΔV) are calculated as:

$$\Delta prop = prop_{\mathrm{HEC}} - \overline{prop}, \quad (6)$$

where $prop_{\mathrm{HEC}}$ is the property value of HEC after mixing. Such changes can quantify the differences between HECs and their precursors.

Table 1 Descriptors used in ML models for HEC design.

Features	Symbol	Classification
One-hot code	Binaries	Type 1
VEC	*VEC*	Type 2
Electronegativity	χ	Type 2
Volume	*V*	Type 2
Mass	*m*	Type 2
Density	ρ	Type 2
Atomic radius	*r*	Type 2
Electronegativity difference	$\Delta\chi$	Type 2
Lattice distortion	δ	Type 2
Mixing entropy	ΔS	Type 2

Mixing enthalpy	ΔH	Type 3
Volume change	ΔV	Type 3
EFA	*EFA*	Type 3
CALPHAD	-	Type 3

Some descriptors shown in Table 1 are important empirical parameters that have been proven to be related to the phase ability and mechanical properties of HECs. The *EFA* [17] parameter was proposed to capture the configurational disorder of HECCs by probing the energy distribution spectrum near the ground state. Specifically, *EFA* is parameterized by:

$$EFA(N) \equiv \{\sigma[spectrum(H_i(N))]_{T=0}\}^{-1}, \quad (7)$$

where $\sigma[spectrum(H_i(N))]_{T=0}$ is the standard enthalpy deviation of configurations at 0 Kelvin. The higher *EFA* value (narrow energy spectrum) signifies higher configurational disorder (or higher entropy) because it is easy for a system to access different configurations, which leads to higher probabilities of single-phase formation (Fig. 3). This concept was successfully applied to guide the synthesis of single-phase HECCs [17, 42], and a critical value between 45 and 50 atom/eV is reported to differentiate single- and multi-phase HECCs. Though *EFA* provides a facile tool to quantify the configurational disorder of high-entropy materials, it is generally unattainable by experiments, and it needs thousands of *ab-initio* calculations to converge a stable enthalpy spectrum for one composition. Thus, an ML model was developed to predict the *EFA* of HECCs directly to ease this dilemma [51]. Some principles used to dictate the phase formation of other materials can also be borrowed to indicate the single-phase stability of HECs. For example, the conventional Hume-Rothery rules [73] state that solid solutions must be formed by precursors with similar atomic radius (< 15%), crystal structure, valency, and electronegativity. In particular, for HEAs, the atomic radius difference is a key influential factor in their phase stability and mechanical properties [74, 75]. Also, the *VEC* is widely used to pinpoint the phase of HEAs [76]. Though it is still challenging to dictate the single-phase stability of HECs with linear combinations of these indicators (Hume-Rothery rules and *VEC*), they can be passed into an ML model to guide the training process. It has been found that the deviations of precursory *VEC*, electronegativity, and atomic mass play crucial roles in ML models for predicting the single-phase synthesizability of HECCs [39].

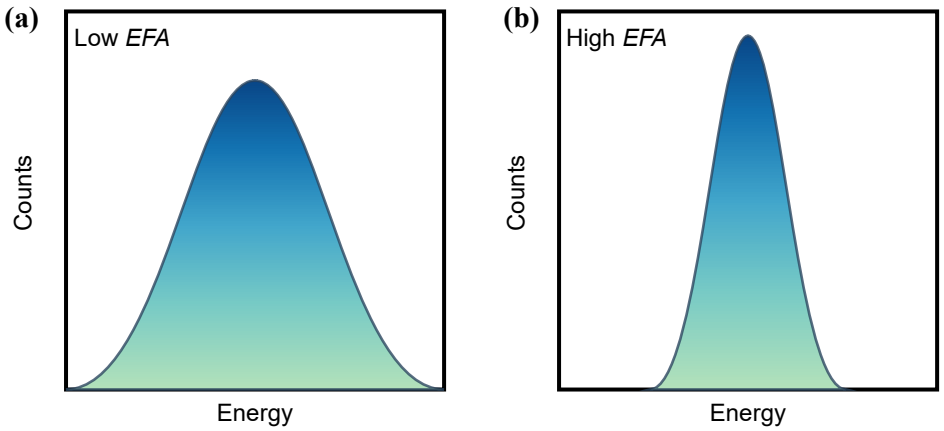

Fig. 3. Schematic representation of *EFA* descriptor. Energy spectrums with (a) low *EFA* and (b) high *EFA*.

Besides phase formation, *VEC* values are also widely used to understand the trend of mechanical properties of ceramics (Fig. 4). Jhi et al. [77] found that the ceramics with rock-salt structures may reach the maximum hardness when *VEC*=8.4 per cell. In addition, Balasubramanian et al. [78] reported that *VEC* plays a critical role in the mechanical behavior of rock-salt ceramics. The ductility can be enhanced by increasing *VEC* [78, 79], which is attributed to the richness of electrons populating on d-t_{2g} orbitals [77, 80]. Critical values of *VEC*=10.6 and *VEC*=10 were proposed for the stable-to-unstable and brittleness-to-ductile transitions, respectively. Though the *VEC* can be correlated with the mechanical properties of muti-component ceramics (≤ 3 components), these trends are difficult to be extrapolated to high-entropy ceramics (≥ 6 components), as lattice distortion and atomic arrangements are more complex. Yet, *VEC* can be a powerful descriptor in ML models as valuable knowledge and noises can be learned and excluded, respectively.

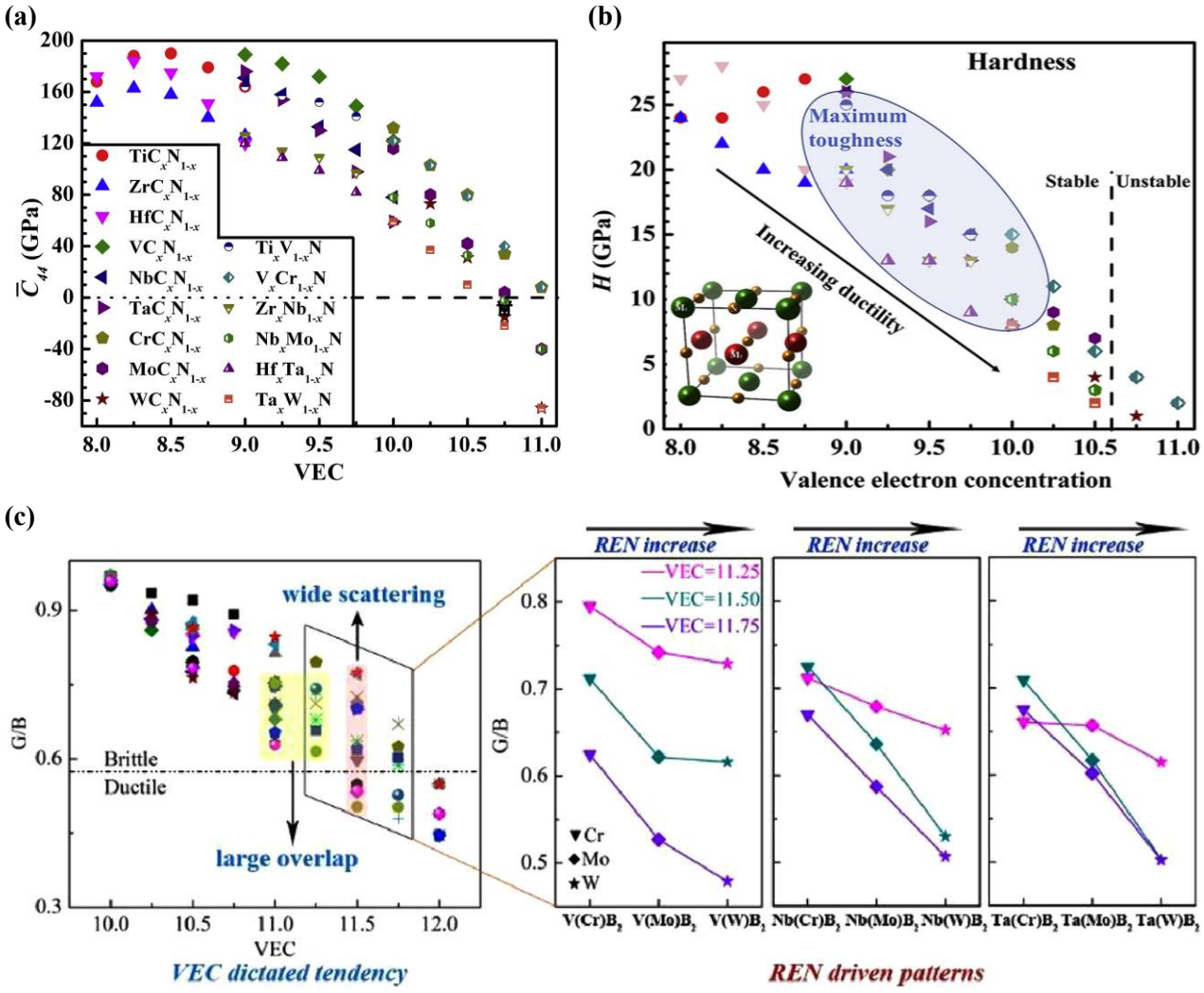

Fig. 4. Correlations between *VEC* and (a) *C44* [78], (b) hardness [78], and (c) *G*/*B* ratio [79] of ceramics with no more than three components.

4.2 Input dimensionality reduction

Some selected features may be highly correlated, which can mislead the ML model to be more sensitive to features with high correlations and less sensitive to those with low correlations. In addition, linearly dependent features contain similar or repetitive information, lowering the learning efficiency. The correlation between input features can be quantified with the Pearson coefficient, which is calculated by:

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{(x_i - \bar{x})^2 (y_i - \bar{y})^2}}, \tag{8}$$

where x_i and y_i are the i-th value of two input dimensions, and their expectations are denoted with $\bar{x}$, and $\bar{y}$, respectively. In practice, input features with r higher than 0.90 or lower than -0.90 can be considered highly dependent. If two input dimensions are found to have high dependence, one of them should be removed. Some current ML models for predicting HECs adopted this method to

remove highly dependent input descriptors [39, 72].

Another way to reduce the input dimensionality is the principal component analysis (PCA). After this procedure, the input dimensions are linearly independent as they are orthogonal to each other, and most data covariance is reserved.

5. Model selection and interpretation

5.1 Principles for model evaluation

In supervised machine learning, the dataset should be randomly split into training, validation, and test sets. Among these datasets, the training set is passed into the ML model, and then the model learns the correlation between input and output by minimizing the difference between predicted and true labels. The validation set is used to calculate the metrics for hyperparameter tuning. The test set consists of samples not encountered by the ML model in the training and fine-tuning procedure, which is used to give the generalization performance of the well-trained model. For a small dataset (especially for newly developed HECs), it can be separated into training and validation sets. However, such a split may lead to a biased model because high-quality samples may not be evenly distributed. The cross-validation (CV) method can be adopted to alleviate the bias. In this process, the dataset is split into n subsets, namely the n-fold CV. The leave-one-out approach is adopted when the number of folds equals the number of samples. Taking the 5-fold CV as an example [also shown in Fig. 5(a)], the dataset is split into five subsets. Then, 5 ML models are trained successively or parallelly, where one of 5 subsets is treated as the validation set in each model, and others are used as the training set. After training, the model performance is evaluated by averaging the metrics of all five models. It is worth noting that the increase in folds will increase the training burden. Generally, the 5- and 10-fold CV methods are widely accepted. This is exactly the method that is adopted in previous works [44, 50, 51, 53, 81, 82] in predicting the single-phase synthesizability of HECCs.

In the training process, overfitting inevitably occurs when too many details (even noise) from the training set are acquired, and the model is still unfamiliar with the validation or test set. For example, in a neural network model, overfitting occurs when the training loss decreases steadily with increasing validation loss [Fig. 5(b)]. In this sense, the model performance should be evaluated on the test (validation) set.

The generalization (or application) of a well-trained model is normally accurate in the model's

"comfort zone", where the training set dominates [Fig. 5(c)]. Contrarily, worse performance can be expected if a model is deployed to a space with sparse training samples [Fig. 5(c)]. Kaufmann et al. [51] trained the ML model based on their earlier experiments [17] without Cr element, while the trained model was generalized to Cr-containing HECCs. To show the model's reliability, the authors furtherly validated selected predictions by experiments [51]. Zhang et al. [39] trained their model to the experimental dataset, which contains metallic elements with equal atomic ratios. But more samples with non-equal metallic ratios are needed to enhance the model performance in predicting HECs far from the central region of the phase diagram. Tang et al. [44] collected more than 400 samples (only 10 HECs included) and trained their model, though the prediction of HECs exhibits larger errors.

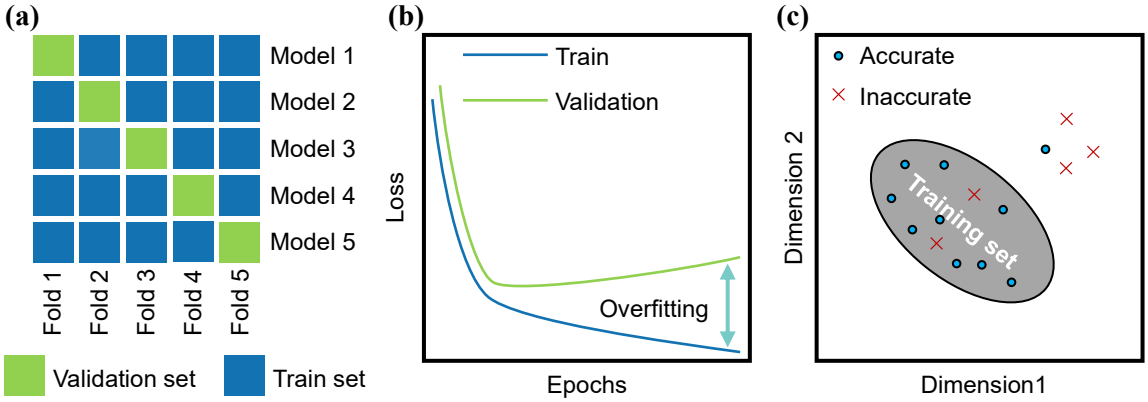

Fig. 5 Schematic representation of (a) 5-fold cross-validation, (b) overfitting, and (c) accurate prediction region.

5.2 Model interpretability

The importance of interpreting ML models lies in multiple aspects. Interpretable physical and chemical information acquired by ML models helps to guide the design and improve the performance. Moreover, such interpretation is necessary to meet the moral and legal requirements [83]. In the linear regression model, trainable parameters (coefficients) are interpretable as they deliver the importance of each input component. However, the linear model is not powerful enough to predict HECs because of the chemical complexity. Other than the linear model, it is challenging to mine the physical and chemical knowledge from trainable parameters in black-box models, such as ANN and deep decision trees. The feature importance can be understood by permuting the input variables [51] or dropping the input dimension successively [39]. For example, by dropping input features successively, Zhang et al. reported that ANN and SVM models are more sensitive to the

deviation of *VEC* and mixing entropy [39]. In contrast, the decision tree model is more sensitive to the average ionic character [51]. However, such approaches can also be misleading [83, 84]. The structural graph neural network naturally preserves crystal symmetry, while the understanding of graph variables is still challenging as graph-based models are normally deep. Thus, the explanations of behaviors made by graphs and deep neural networks for predicting HECs are still lacking [48, 49, 82].

Table 2

Machine-learning models used to design high-entropy ceramics

Model type	Training dataset size	Prediction dataset size	Label (fitting target)	Input dimension	Core Architecture	Model performance	Year	Ref
Regression	56	70	*EFA*	108-116	RF	train MAE: 3.8 $(eV/atom)^{-1}$; Cr-containing HEEC: 7.0 $(eV/atom)^{-1}$	2020	[51]
Regression	2,750	MD simulation	Energy and force	-*	ANN	Energy MAE: 9.4 meV/atom; Force MAE: 217 meV/Å	2020	[48]
Regression	Formation energy: 12,277; elastic moduli: 1,672	399,960	Formation energy and elastic moduli	-*	GNN	Energy MAE: 26 meV/atom; Elastic moduli: < 0.17 log10GPa	2021	[82]
Regression	325	436,494	elastic moduli	3	GPR	Average MAE: 20 GPa	2021	[44]
Regression	4,700	MD simulation	Energy and force	-*	ANN	Energy MAE: 9.2 meV/atom; Force MAE: 208 meV/Å	2021	[49]
Regression	63	64	*EFA*	108-116	RF	-	2021	[52]
Classification	53	90	single-phase probability	8-10	ANN, SVM	Train accuracy: 98.2%; Validation accuracy: 96.0%	2022	[50]
Classification	126	-	single-phase formability	6	KNN, LR, GNB	Test accuracy: ~94%	2022	[81]
Regression	557	9,212,112	Hardness	20	RF, GBRT, DT	$(R^2)_{avg}$: 0.907	2022	[53]
Regression	1,280	-	Carbon vacancy formation energy	Radial distribution functions	RF	Test MAE: 0.054 eV	2023	[85]

*Non-empirical code, which is represented by atomic number or one-hot code. RF: random forest; ANN: artificial neural network; SVM: support vector machine; KNN: k-nearest neighbor; LR: Logistic regression; GNB: gaussian naive Bayes; GNN: graph neural network; GPR: Gaussian process regression; GBRT: gradient boosting regression trees; DT: decision tree. MAE: mean absolute error.

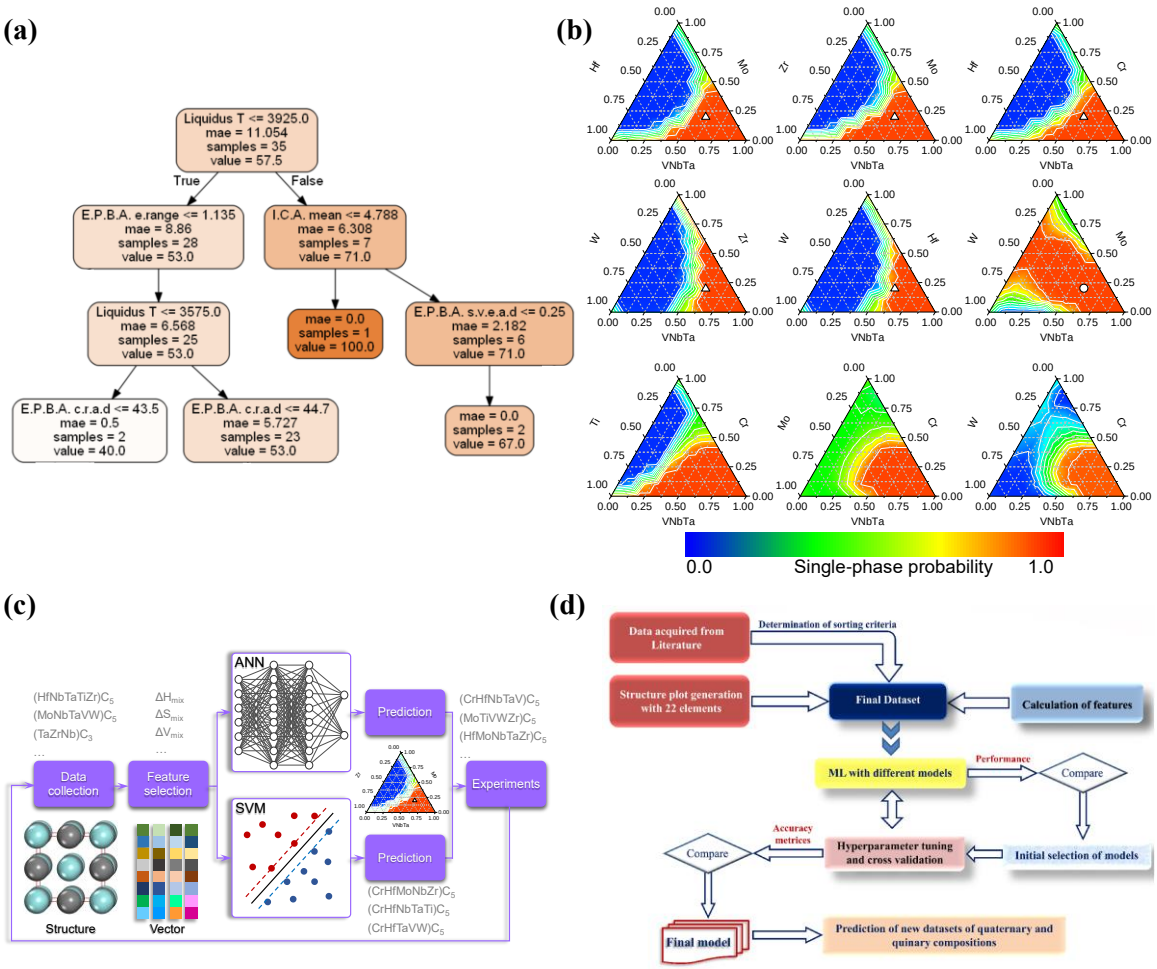

Fig. 6 Proposed ML models for predicting single-phase synthesizability of HECs. (a) Selected decision trees from random forest model [51]. (b) rational exploration of non-equiatomic HECCs [39]. (c) workflow for design single-phase HECCs [39]. (d) machine-learning model for predicting high-entropy boride ceramics (HEBCs).

The ML models can be classified as supervised and unsupervised learning according to their usage of labeled data. For HECs, as listed in Table 2, almost all the published models belong to supervised learning, where the properties, such as single-phase probability and mechanical properties, are used as labels to guide the training process. To date, many ML architectures have been successfully deployed to predict the phase formation and mechanical properties of HECs (Fig. 6), such as ANN, SVM, random forest (RF), etc.

6.1 Discovering single-phase HECs

Considering the vast composition space of HECs, the experimental exploration of HECs is quite

expensive. Only hundreds of HECs were experimentally confirmed as single-phase compounds. To solve this problem, ANN, SVM, and RF models are deployed to predict the single-phase formability of HECCs and HEBCs.

Kaufmann et al. [51] firstly introduced the ML method to HECs. They used decision trees to predict the *EFA* of HECCs, which show good agreement with DFT results. They also predicted the *EFA* values of 70 new compositions and focused on group VI elements that cannot form stable rock-salt structures in their binary form. Compared to other ML models based on limited experimental data, the performance of proposed decision trees can effectively be validated with new datasets, which are calculated by the DFT method. Additionally, their work presented that the immiscibility of binary transition-metal carbides can be conquered by the high-entropy effects, furtherly expanding the family of HECCs. Mellor et al. [52] further extended this method to HECC with more than five transition metal cations. Afterward, Zhang et al. [39] employed ANN and SVM models to predict the single-phase probability of HECCs directly. The ANN model is widely used in deep learning models, and SVM is competitive in handling small datasets, which is suitable for predicting HECs as only a small number of HECs were experimentally reported. By training the models with precursory properties only, the proposed models can rapidly explore the whole composition space without additional costly DFT calculations. Therefore, the authors further predict the phase diagram of VNbTa- and CrMoW-based senary HECCs. Based on these phase diagrams, there is still vast space that is not explored experimentally. Therefore, ML provides a facile tool for designing single-phase HECs with equal and non-equal metallic atomic ratios. Moreover, the authors proved that the conventional thermodynamic descriptors are less effective in predicting single-phase foamabilities of HECCs. The ML approach has also been applied to explore single-phase HEBCs. Akrami et al. [35] listed more than 100 borides, most of which are transitional-metal di-borides with hexagonal structures. Thanks to these newly-discovered HECBs, Mitra et al. [81] created a synthetic training dataset with structure plots and compared the dataset with available publications. Based on their empirical method, no experiments are needed, effectively circumventing the shortage of experimental data. In addition, these synthetic data from structure plots can effectively augment experimental or *ab-initio* datasets. However, the reliability of the generated dataset is difficult to be verified.

6.2 Predicting properties

Despite the single-phase formability, the direct prediction of mechanical properties, such as elastic constants and moduli, is important and relevant to the rational design of HECs with exceptional

properties.

Tang et al. [44] developed a bond-parameter-guided ML model to predict the mechanical properties of (TMCs) and HECs. The bond parameters describe the bond order, bond ionicity, and bond length of samples in the collected database. This work is distinguishable from others as the bond parameters can be directly used to correlate the mechanical properties, and thus the proposed scaling relations between mechanical properties and bond descriptors are inherently interpretable. Jaafreh et al. [86] developed several ML models to predict super-hard HEBCs. Instead of using crystal features that can only be obtained from single-phase HECs, the authors adopted compositional features. This approach is thus applicable in both single-phase and multi-phase HECs. In this work, the hardness values were collected from publications, which differs from earlier models for predicting mechanical properties by training on the calculated properties [87]. Zhao et al. [85] trained a random forest (RF) model to predict the formation energy of carbon vacancy, which is normally prevalent in HECC crystals and influential on the mechanical properties.

Intuitively, all the above models (ANN, SVM, and DT) can be generalized to other ceramic materials. However, a general model predicting the phase formability and/or elastic moduli of all types of HECs is still unavailable due to the diverse crystal and local structures associated with all HECs. Specifically, most single-phase HECCs are stable with rock-salt structures [35]. While various structural types, spinel, perovskite, fluorite, rock-salt, amorphous, orthorhombic, etc., were discovered to be stable forms of HEOCs. Therefore, ML models with only several hidden layers or simple architectures are not enough, and deep-learning models with much more trainable variables are desired to fit the complex correlations between HEC descriptors and their properties.

Recently, graph neural network has attracted extensive attention because its topological structure is naturally suitable for representing molecules and crystals [88-91]. In a crystal graph [88] (See Fig. 7), the atoms and bonds are represented by graph nodes and edges, respectively. In addition, the translation, rotation, and permutation invariances of crystal graphs are conserved to ensure an accurate and unique representation of crystals. Inspired by these merits, Zuo et al. [87] adopted a graph deep learning energy model to circumvent the high-cost DFT calculations for structural relaxation and applied the method to TMCs and ternary ceramics. Though the reported mean absolute error (MAE) of energy is low (26 meV/atom), it is still not accurate enough to derive the atomic forces from the energy by perturbating atomic positions. Thus, the authors relaxed structures by a Bayesian optimization with symmetry relaxation (BOWSR) approach. The study demonstrates

that graph neural networks show great ability in predicting energy and structural properties. However, their direct applications to HECs are challenging as such a deep-learning model requires huge amount of data.

Another methodology that can be used to predict the properties of HECs is deep potential molecular dynamics (DPMD) [92, 93]. In this approach, the atomic energy is formulated as $E = \sum_i E_i$, where E and E_i are the total energy and atomic energy, respectively. The energies result from many-body interactions within a cutoff distance (R_c). To preserve the geometrical symmetries (translation, rotation, and permutation), DPMD parses the coordinates with full (radial and angular) or radial information. After this treatment, the atomic positions (R_i) are mapped into descriptors (D_{ij}) which can be passed into neural networks directly. Atomic forces are calculated by deriving total energy (E) with respect to atomic positions. Specifically, the partial derivatives associated with atomic force and virial tensor are computed by the chain rule through backpropagation. By incorporating the error of energy, forces, and virial tensor into a loss function with user-defined weights, the deviations of model predictions from DFT-calculated targets are minimized in the training process. Such a deep-learning model is widely used to fit interatomic potentials or force fields with DFT accuracy, bridging the gap between DFT and molecule dynamics. For HECs, Dai et al. trained deep-learning potentials for $(Zr_{0.2}Hf_{0.2}Ti_{0.2}Nb_{0.2}Ta_{0.2})C$ [48] and $(Ti_{0.2}Zr_{0.2}Hf_{0.2}Nb_{0.2}Ta_{0.2})B_2$ [49] with low mean absolute error of energy (9.4 meV/atom) and force (217 meV/Å). With the well-trained interatomic potentials, the authors predicted thermal and mechanical properties of HECs, which are expensive to simulate directly with *ab-initio* calculations.

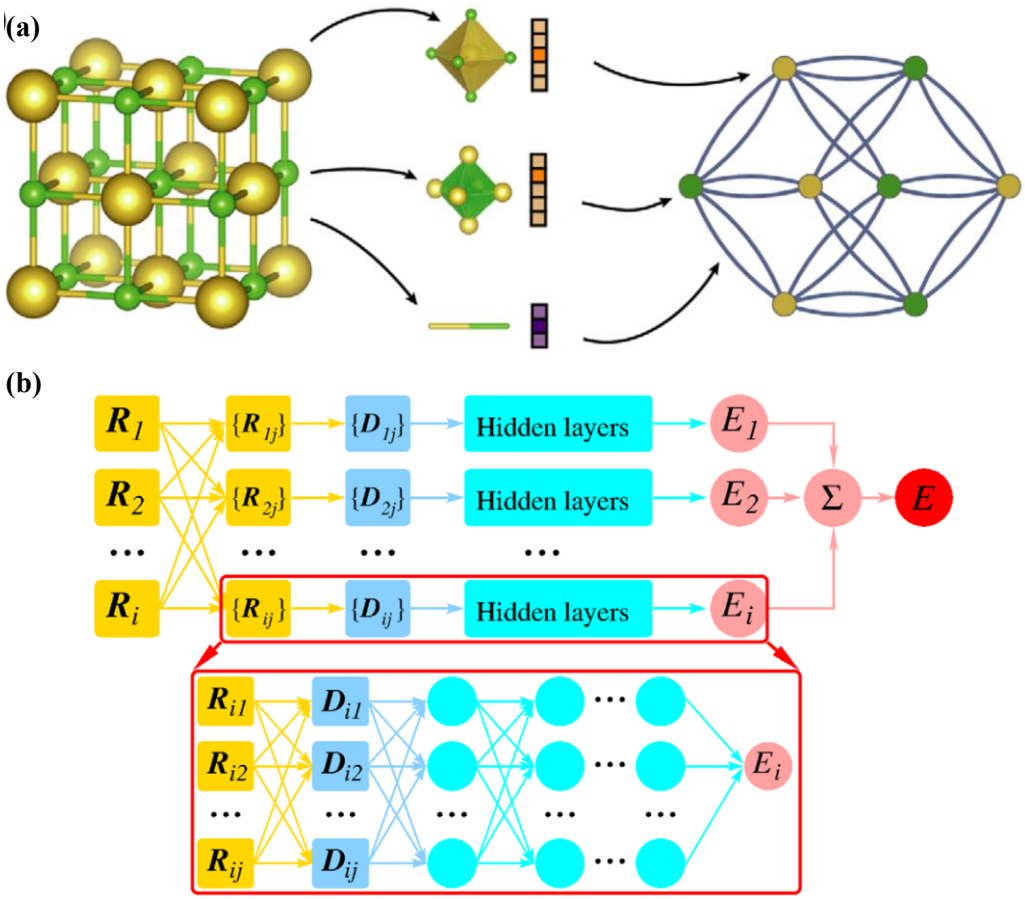

Fig. 7 Deep-learning models. (a) graph representation of a crystal [88]. (b) schematic plot of the deep potential molecular dynamics (DPMD) [93].

7. Outlook

The unique properties demonstrated in HECs render their great promise in advanced applications. One of the key challenges of developing novel HECs is the meaningful exploration within the enormous phase space. The successful applications of ML methods have significantly advanced the discovery and design of high-performance HECs. Nonetheless, this approach is still in its infancy in the HEC community, and there is still vast space for promoting the performance of ML models. In this section, we provide our vision for future directions of ML for HEC prediction.

7.1 High-quality sharable data repository

As discussed above, a database designed for HECs is yet unavailable, which acts as the bottleneck and impedes the advance of ML applications. Though HEC is still at its infant age, more than one thousand publications have been retrieved from the "Web of Science" with the keyword "high entropy ceramic" until January 2023. Building a database by manually collecting information from

all these publications is becoming an impractical task. Fortunately, some ML models (natural language process) are versatile to understand and extract key information from the literature. The natural language process has been deployed to aid the design of conventional materials with exotic properties [94], which is also a new opportunity for HECs. Additionally, the available information may have different emphases on phase stability and mechanical properties. For example, most publications for discovering HECs focus on the pure phase results, and the compounds with mixed phases may be discarded, leading to the data imbalance of ML models. In experiments, sintering temperature, preparation method, and sample porosity can all affect the structure and performance of HECs. However, these parameters are generally partially reported in the literature because the results are presented in various domains. In terms of theoretical simulations, the input configuration will influence the results significantly. For example, the supercell size adopted by the *ab-initio* method is generally small owing to high computational resources. To simulate the randomness in HECs, multiple special quasirandom structures (SQSs) are necessary to converge the system fluctuations. How this average is done thus can affect the reported values. Also, molecular dynamic simulations based on different empirical interatomic potentials may lead to contradictory conclusions. Therefore, a high-quality database that has both accurate labels and thoroughly prepared parameters is highly desirable.

7.2 Unsupervised learning

Since high-quality data from experiments and *ab-initio* simulations are scarce, unsupervised models provide new opportunities to solve this challenge. Instead of collecting the expensive labels, unsupervised learning train the model based on a dataset without property labels. Then the samples are classified into groups with clustering algorithms. After that, each group can be represented by its center. Novel materials can be discovered in the region close to those data points or group centers. Recently, such models have been used to design energy and environmental materials with ordered [95-98] and disordered structures [99-102], though they have not been adopted to predict HECs so far.

7.3 Multi-functional models

Multi-functional models show great potential as the phase stability, defect evolution, and mechanical properties are all decisive for the technical applications of HECs. However, such a model requires a sheer size of trainable variables because it needs to learn high-dimensional features that are buried in the immense phase space and complex structures. Currently, the

generalizability of proposed models is limited to similar compositions and structures. Those models for predicting single-phase synthesizability [39, 51, 81] cannot be used to predict the mechanical properties [44, 53, 82] of HECs, and vice versa. Additionally, interfaces and defects, which significantly affect phase stability and mechanical properties [40, 103, 104], are normally ignored in these models. In this sense, it is reasonable to foresee that deep learning can alleviate this high-dimensional problem because of its super adaptability to complex scenarios. Though the current deep-learning models [48, 49, 82]for HECs are used in limited applications, they can be flexible for multiple tasks if relevant data is available.

7.4 Inverse design strategy

Most ML models for material design are classified as discriminative models, which map the input features of crystal structures to their physical properties [105]. To discover high-performance materials, the well-trained model needs to scan the input space and filter the results. Hence, discriminative models cannot make inferences about new materials upon human-defined targets. In contrast, generative models overcome this challenge by learning from the distribution of a dataset and sampling new materials with probabilistic representation. For example, Ren et al. [106] proposed a general model for the inverse design of inorganic materials. In this work, solid-state crystals with user-specified formation energy, band gap, and thermoelectric power factor are generated by the generative model. Though the inverse design strategy has not been implemented in the ML models for predicting HECs, this strategy brings vast opportunities to the intelligent design of high-performance HECs.

7.5 Model interpretability

As discussed above, interpretable models explain their behaviors, building trust between humans and machines [107]. The current black-box ML models for designing HECs are normally understood with the permutation or drop-out variance of input features, which yield the relative importance or sensitivity of input dimensions. Though the interpretation of black-box models can be misleading, which is unsuitable for making high-stack decisions [83], such models still provide valuable knowledge [108]. Currently, inherently interpretable ML models are still lacking in designing advanced HECs. In material science, the commonly accepted interpretable ML method is the symbolic regression (SR), including the least absolute shrinkage and selection operator (LASSO)

and sure independence screening and sparsifying operator (SISSO). For example, the SISSO method was used to find the optimal analytical model for predicting mechanical properties [109] and stacking fault energies [110] of HEAs. The LASSO model was reported to explore the composition space of HEA-based electrocatalysts [111]. These SR-based models are employed to search mathematical expressions between inputs and outputs. Thus, predictions made by these models are inherently explainable. More importantly, SR models can generate new knowledge, providing design principles for novel materials.

7.6 ML-aided high-throughput experimentation

With the help of developed ML models, HECs with both equal and non-equal metallic compositions can be explored by a high-throughput filter. For their experimental counterpart, high-throughput preparation method can also enable the rapid design and screening of high-performance HECs. For example, Wang et al. [112] developed an ultrafast method that could synthesize and sinter ceramics in seconds. With this method, high-throughput materials screening was applied to confirm the DFT results. Though, most current ML applications for HECs that combine experiments [39, 51] are conducted in a series of steps. ML models first learn the governing information from the training dataset, followed by the prediction of new HECs. Then, experiments are performed to validate ML predictions. After that, new experimental results are augmented to the database, and models are retrained (see Fig. 1). Though these procedures are faster than conventional trial and error, human labor is still needed to perform experiments and bring results to ML models, limiting the advancement of ML models. In contrast, high-throughput experimentation rapidly evaluates ML prediction and provides feedback to ML models. Besides, reinforcement learning (RL), different from supervised and unsupervised learning, can adjust the input variables interactively according to the environmental rewards and punishments within the huge phase space of HECs. By combining high-throughput experimentation and RL, exceptional HECs can be rapidly discovered without human supervision, furtherly boosting the designing cycle of high-performance HECs. In addition, the high-throughput experimentation is also promising in combination with semi-supervised active learning algorithm [113-116], where compositions can be refined by an automated optimizer (such as Bayesian optimization).

8. Conclusions

In this article, we have introduced key challenges of developing high-performance HECs, due to their huge phase space and complex structures. We show that ML is a promising approach for HEC

prediction, both for the single-phase synthesizability and mechanical properties. For this purpose, we have thoroughly reviewed recent progress on ML applications in HECs, including database acquisition, selection of input features, model selection and evaluation, and model generalization. In the end, we outline the future directions of developing ML models for HECs and emphasize the importance of database construction, deep learning, label-free unsupervised learning, and high-throughput experimentation.

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

This work was supported by National Key R&D Program of China (No. 2022YFE0200900) and Research Grant Council of Hong Kong (No. 21200919).

References

[1] C. Oses, C. Toher, S. Curtarolo, High-entropy ceramics, Nat. Rev. Mater. 5(4) (2020) 295-309.

[2] L. Feng, W.G. Fahrenholtz, D.W. Brenner, High-entropy ultra-high-temperature borides and carbides: A new class of materials for extreme environments, Annu. Rev. Mater. Res. 51(1) (2021) 165-185.

[3] S.J. Zinkle, G.S. Was, Materials challenges in nuclear energy, Acta Mater. 61(3) (2013) 735-758.

[4] P. Hosemann, D. Frazer, M. Fratoni, A. Band, M.F. Ashby, Materials selection for nuclear applications: Challenges and opportunities, Scripta Mater. 143 (2018) 181-187.

[5] F. Wang, X.L. Yan, T.Y. Wang, Y.Q. Wu, L. Shao, M. Nastasi, Y.F. Lu, B. Cui, Irradiation damage in $(Zr_{0.25}Ta_{0.25}Nb_{0.25}Ti_{0.25})C$ high-entropy carbide ceramics, Acta Mater. 195 (2020) 739-749.

[6] Y. Osetsky, A.V. Barashev, Y. Zhang, Sluggish, chemical bias and percolation phenomena in atomic transport by vacancy and interstitial diffusion in NiFe alloys, Curr. Opin. Solid State Mater. Sci. 25(6) (2021) 100961.

[7] B.P. Uberuaga, D.A. Andersson, C.R. Stanek, Defect behavior in oxides: Insights from modern atomistic simulation methods, Curr. Opin. Solid State Mater. Sci. 17(6) (2013) 249-256.

[8] J.W. Yeh, S.K. Chen, S.J. Lin, J.Y. Gan, T.S. Chin, T.T. Shun, C.H. Tsau, S.Y. Chang, Nanostructured high-entropy alloys with multiple principal elements: Novel alloy design concepts and outcomes, Adv. Eng. Mater. 6(5) (2004) 299-303.

[9] J.W. Yeh, Recent progress in high-entropy alloys, Annales de Chimie Science des Matériaux 31(6) (2006) 633-648.
[10] B. Cantor, I.T.H. Chang, P. Knight, A.J.B. Vincent, Microstructural development in equiatomic multicomponent alloys, Mater. Sci. Eng. A 375 (2004) 213-218.
[11] D.B. Miracle, O.N. Senkov, A critical review of high entropy alloys and related concepts, Acta Mater. 122 (2017) 448-511.
[12] E.P. George, D. Raabe, R.O. Ritchie, High-entropy alloys, Nat. Rev. Mater. 4(8) (2019) 515-534.
[13] T.K. Chen, T.T. Shun, J.W. Yeh, M.S. Wong, Nanostructured nitride films of multi-element high-entropy alloys by reactive DC sputtering, Surface and Coatings Technology 188-189 (2004) 193-200.
[14] C.M. Rost, E. Sachet, T. Borman, A. Moballegh, E.C. Dickey, D. Hou, J.L. Jones, S. Curtarolo, J.P. Maria, Entropy-stabilized oxides, Nat Commun 6 (2015) 8485.
[15] J. Gild, Y. Zhang, T. Harrington, S. Jiang, T. Hu, M.C. Quinn, W.M. Mellor, N. Zhou, K. Vecchio, J. Luo, High-entropy metal diborides: A new class of high-entropy materials and a new type of ultrahigh temperature ceramics, Sci. Rep. 6(1) (2016) 37946.
[16] J.Y. Zhou, J.Y. Zhang, F. Zhang, B. Niu, L.W. Lei, W.M. Wang, High-entropy carbide: A novel class of multicomponent ceramics, Ceram. Int. 44(17) (2018) 22014-22018.
[17] P. Sarker, T. Harrington, C. Toher, C. Oses, M. Samiee, J.P. Maria, D.W. Brenner, K.S. Vecchio, S. Curtarolo, High-entropy high-hardness metal carbides discovered by entropy descriptors, Nat. Commun. 9(1) (2018) 4980.
[18] E. Castle, T. Csanadi, S. Grasso, J. Dusza, M. Reece, Processing and properties of high-entropy ultra-high temperature carbides, Sci. Rep. 8(1) (2018) 8609.
[19] Z.T. Li, Z. Wang, Z.G. Wu, B. Xu, S.J. Zhao, W.D. Zhang, N. Lin, Phase, microstructure and related mechanical properties of a series of (NbTaZr)C-Based high entropy ceramics, Ceram. Int. 47(10) (2021) 14341-14347.
[20] L. He, J. Zhang, Z.T. Li, N. Lin, B. Liu, S.J. Zhao, K. Jin, H.Y. Chen, H.G. Yan, F. Peng, Y. Ma, Z.G. Wu, Toughening (NbTaZrW)C high-entropy carbide ceramic through Mo doping, J. Am. Ceram. Soc. 105(8) (2022) 5395-5407.
[21] T. Jin, X. Sang, R.R. Unocic, R.T. Kinch, X. Liu, J. Hu, H. Liu, S. Dai, Mechanochemical-assisted synthesis of high-entropy metal nitride via a soft urea strategy, Adv. Mater. 30(23) (2018) 1707512.
[22] R.Z. Zhang, F. Gucci, H. Zhu, K. Chen, M.J. Reece, Data-driven design of ecofriendly thermoelectric high-entropy sulfides, lnorg. Chem. 57(20) (2018) 13027-13033.
[23] P. Edalati, X.-F. Shen, M. Watanabe, T. Ishihara, M. Arita, M. Fuji, K. Edalati, High-entropy oxynitride as a low-bandgap and stable photocatalyst for hydrogen production, J. Mater. Chem. A

9(26) (2021) 15076-15086.
[24] S. Akrami, P. Edalati, Y. Shundo, M. Watanabe, T. Ishihara, M. Fuji, K. Edalati, Significant CO_2 photoreduction on a high-entropy oxynitride, Chem. Eng. J. 449 (2022) 137800.
[25] O.F. Dippo, N. Mesgarzadeh, T.J. Harrington, G.D. Schrader, K.S. Vecchio, Bulk high-entropy nitrides and carbonitrides, Sci. Rep. 10(1) (2020) 21288.
[26] T.Q. Wen, B.L. Ye, M.C. Nguyen, M.D. Ma, Y.H. Chu, Thermophysical and mechanical properties of novel high-entropy metal nitride-carbides, J. Am. Ceram. Soc. 103(11) (2020) 6475-6489.
[27] L. Hu, F. Zhong, J. Zhang, S. Zhao, Y. Wang, G. Cai, T. Cheng, G. Wei, S. Jia, D. Zhang, R. Yin, Z. Chen, C. Jiang, F. Ren, High hydrogen isotopes permeation resistance in (TiVAlCrZr)O multi-component metal oxide glass coating, Acta Mater. 238 (2022) 118204.
[28] X.S. Zhang, Y.G. Li, C.X. Li, F. Yang, Z.M. Jiang, L.Y. Xue, Z.H. Shao, Z.G. Zhao, M.Y. Xie, S.W. Yu, A novel $(La_{0.2}Ce_{0.2}Gd_{0.2}Er_{0.2}Tm_{0.2})_2(WO_4)_3$ high-entropy ceramic material for thermal neutron and gamma-ray shielding, Mater. Des. 205 (2021) 109722.
[29] J. Wang, R. Shu, J.L. Chai, S.G. Rao, A. le Febvrier, H.C. Wu, Y.B. Zhu, C.F. Yao, L.H. Luo, W.P. Li, P.F. Gao, P. Eklund, Xe-ion-irradiation-induced structural transitions and elemental diffusion in high-entropy alloy and nitride thin-film multilayers, Mater. Des. 219 (2022) 110749.
[30] D. Berardan, S. Franger, A.K. Meena, N. Dragoe, Room temperature lithium superionic conductivity in high entropy oxides, J. Mater. Chem. A 4(24) (2016) 9536-9541.
[31] S.Y. Niu, Z.W. Yang, F.G. Qi, Y. Han, Z.Z. Shi, Q.W. Qiu, X.P. Han, Y. Wang, X.W. Du, Electrical discharge induced bulk-to-nanoparticle transformation: Nano high-entropy carbide as catalysts for hydrogen evolution reaction, Adv. Funct. Mater. 32(35) (2022) 2203787.
[32] Z. Lun, B. Ouyang, D.H. Kwon, Y. Ha, E.E. Foley, T.Y. Huang, Z. Cai, H. Kim, M. Balasubramanian, Y. Sun, J. Huang, Y. Tian, H. Kim, B.D. McCloskey, W. Yang, R.J. Clement, H. Ji, G. Ceder, Cation-disordered rocksalt-type high-entropy cathodes for Li-ion batteries, Nat. Mater. 20(2) (2021) 214-221.
[33] A. Sarkar, L. Velasco, D. Wang, Q. Wang, G. Talasila, L. de Biasi, C. Kubel, T. Brezesinski, S.S. Bhattacharya, H. Hahn, B. Breitung, High entropy oxides for reversible energy storage, Nat. Commun. 9(1) (2018) 3400.
[34] C. Zhao, F. Ding, Y. Lu, L. Chen, Y.S. Hu, High-entropy layered oxide cathodes for sodium-ion batteries, Angew. Chem. Int. Ed. 59(1) (2020) 264-269.
[35] S. Akrami, P. Edalati, M. Fuji, K. Edalati, High-entropy ceramics: Review of principles, production and applications, Mat. Sci. Eng. R 146 (2021) 100644.
[36] B. Liu, J.L. Zhao, Y.C. Liu, J.Q. Xi, Q. Li, H.M. Xiang, Y.C. Zhou, Application of high-throughput first-principles calculations in ceramic innovation, J. Mater. Sci. Technol 88 (2021) 143-157.

[37] Z. Wang, Z.T. Li, S. Zhao, Z.G. Wu, High-entropy carbide ceramics: a perspective review, Tungsten 3(2) (2021) 131-142.
[38] R.Z. Zhang, M.J. Reece, Review of high entropy ceramics: design, synthesis, structure and properties, J. Mater. Chem. A 7(39) (2019) 22148-22162.
[39] J. Zhang, B.A. Xu, Y.X. Xiong, S.H. Ma, Z. Wang, Z.G. Wu, S.J. Zhao, Design high-entropy carbide ceramics from machine learning, npj Comput. Mater. 8(1) (2022) 5.
[40] S.J. Zhao, Lattice distortion in high-entropy carbide ceramics from first-principles calculations, J. Am. Ceram. Soc. 104(4) (2021) 1874-1886.
[41] F. Otto, Y. Yang, H. Bei, E.P. George, Relative effects of enthalpy and entropy on the phase stability of equiatomic high-entropy alloys, Acta Mater. 61(7) (2013) 2628-2638.
[42] T.J. Harrington, J. Gild, P. Sarker, C. Toher, C.M. Rost, O.F. Dippo, C. McElfresh, K. Kaufmann, E. Marin, L. Borowski, P.E. Hopkins, J. Luo, S. Curtarolo, D.W. Brenner, K.S. Vecchio, Phase stability and mechanical properties of novel high entropy transition metal carbides, Acta Mater. 166 (2019) 271-280.
[43] D. Moskovskikh, S. Vorotilo, V. Buinevich, A. Sedegov, K. Kuskov, A. Khort, C. Shuck, M. Zhukovskyi, A. Mukasyan, Extremely hard and tough high entropy nitride ceramics, Sci. Rep. 10(1) (2020) 19874.
[44] Y.Q. Tang, D. Zhang, R.L. Liu, D.Y. Li, Designing high-entropy ceramics via incorporation of the bond-mechanical behavior correlation with the machine-learning methodology, Cell Rep. Phys. Sci. 2(11) (2021) 100640.
[45] R. Feng, P.K. Liaw, M.C. Gao, M. Widom, First-principles prediction of high-entropy-alloy stability, npj Comput. Mater. 3(1) (2017) 50.
[46] Y. Lederer, C. Toher, K.S. Vecchio, S. Curtarolo, The search for high entropy alloys: A high-throughput ab-initio approach, Acta Mater. 159 (2018) 364-383.
[47] D. Morgan, G. Pilania, A. Couet, B.P. Uberuaga, C. Sun, J. Li, Machine learning in nuclear materials research, Curr. Opin. Solid State Mater. Sci. 26(2) (2022) 100975.
[48] F.Z. Dai, B. Wen, Y.J. Sun, H.M. Xiang, Y.C. Zhou, Theoretical prediction on thermal and mechanical properties of high entropy $(Zr_{0.2}Hf_{0.2}Ti_{0.2}Nb_{0.2}Ta_{0.2})C$ by deep learning potential, J. Mater. Sci. Technol 43 (2020) 168-174.
[49] F.Z. Dai, Y. Sun, B. Wen, H. Xiang, Y. Zhou, Temperature dependent thermal and elastic properties of high entropy $(Ti_{0.2}Zr_{0.2}Hf_{0.2}Nb_{0.2}Ta_{0.2})B_2$: Molecular dynamics simulation by deep learning potential, J. Mater. Sci. Technol 72 (2021) 8-15.
[50] J. Zhang, B. Xu, Y. Xiong, S. Ma, Z. Wang, Z. Wu, S. Zhao, Design high-entropy carbide ceramics from machine learning, npj Comput. Mater. 8(1) (2022).
[51] K. Kaufmann, D. Maryanovsky, W.M. Mellor, C.Y. Zhu, A.S. Rosengarten, T.J. Harrington, C. Oses, C. Toher, S. Curtarolo, K.S. Vecchio, Discovery of high-entropy ceramics via machine

learning, npj Comput. Mater. 6(1) (2020) 42.
[52] W.M. Mellor, K. Kaufmann, O.F. Dippo, S.D. Figueroa, G.D. Schrader, K.S. Vecchio, Development of ultrahigh-entropy ceramics with tailored oxidation behavior, J. Eur. Ceram. Soc. 41(12) (2021) 5791-5800.
[53] R. Jaafreh, Y.S. Kang, J.G. Kim, K. Hamad, Machine learning guided discovery of super-hard high entropy ceramics, Mater. Lett. 306 (2022) 130899.
[54] W. Zhu, W. Huo, S. Wang, X. Wang, K. Ren, S. Tan, F. Fang, Z. Xie, J. Jiang, Phase formation prediction of high-entropy alloys: a deep learning study, Journal of Materials Research and Technology 18 (2022) 800-809.
[55] Y.G. Yan, D. Lu, K. Wang, Overview: recent studies of machine learning in phase prediction of high entropy alloys, Tungsten (2022).
[56] D. Morgan, R. Jacobs, Opportunities and challenges for machine learning in materials science, Annu. Rev. Mater. Res. 50(1) (2020) 71-103.
[57] J. Schmidt, M.R.G. Marques, S. Botti, M.A.L. Marques, Recent advances and applications of machine learning in solid-state materials science, npj Comput. Mater. 5(1) (2019) 83.
[58] E.W. Huang, W.J. Lee, S.S. Singh, P. Kumar, C.Y. Lee, T.N. Lam, H.H. Chin, B.H. Lin, P.K. Liaw, Machine-learning and high-throughput studies for high-entropy materials, Mat. Sci. Eng. R 147 (2022).
[59] L. Qiao, Y. Liu, J.C.A. Zhu, A focused review on machine learning aided high-throughput methods in high entropy alloy, J. Alloys Compd. 877 (2021) 160295.
[60] N.K. Katiyar, G. Goel, S. Goel, Emergence of machine learning in the development of high entropy alloy and their prospects in advanced engineering applications, Emergent Mater. 4(6) (2021) 1635-1648.
[61] H.D. Fu, H.T. Zhang, C.S. Wang, W. Yong, J.X. Xie, Recent progress in the machine learning-assisted rational design of alloys, Int. J. Miner., Metall. Mater. 29(4) (2022) 635-644.
[62] S. Zhao, Application of machine learning in understanding the irradiation damage mechanism of high-entropy materials, J. Nucl. Energy 559 (2022) 153462.
[63] G.L.W. Hart, T. Mueller, C. Toher, S. Curtarolo, Machine learning for alloys, Nat. Rev. Mater. 6(8) (2021) 730-755.
[64] X. Liu, J. Zhang, Z. Pei, Machine learning for high-entropy alloys: Progress, challenges and opportunities, Progress in Materials Science 131 (2023) 101018.
[65] J.J. de Pablo, B. Jones, C.L. Kovacs, V. Ozolins, A.P. Ramirez, The materials genome initiative, the interplay of experiment, theory and computation, Curr. Opin. Solid State Mater. Sci. 18(2) (2014) 99-117.
[66] J.J. de Pablo, N.E. Jackson, M.A. Webb, L.-Q. Chen, J.E. Moore, D. Morgan, R. Jacobs, T. Pollock, D.G. Schlom, E.S. Toberer, J. Analytis, I. Dabo, D.M. DeLongchamp, G.A. Fiete, G.M.

Grason, G. Hautier, Y. Mo, K. Rajan, E.J. Reed, E. Rodriguez, V. Stevanovic, J. Suntivich, K. Thornton, J.-C. Zhao, New frontiers for the materials genome initiative, npj Comput. Mater. 5(1) (2019) 41.
[67] A. Jain, S.P. Ong, G. Hautier, W. Chen, W.D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, K.A. Persson, Commentary: The Materials Project: A materials genome approach to accelerating materials innovation, APL Mater. 1(1) (2013) 011002.
[68] S. Curtarolo, W. Setyawan, S.D. Wang, J.K. Xue, K.S. Yang, R.H. Taylor, L.J. Nelson, G.L.W. Hart, S. Sanvito, M. Buongiorno-Nardelli, N. Mingo, O. Levy, AFLOWLIB.ORG: A distributed materials properties repository from high-throughput ab initio calculations, Comput. Mater. Sci. 58 (2012) 227-235.
[69] J.E. Saal, S. Kirklin, M. Aykol, B. Meredig, C. Wolverton, Materials design and discovery with high-throughput density functional theory: The open quantum materials database (OQMD), JOM 65(11) (2013) 1501-1509.
[70] S. Kirklin, J.E. Saal, B. Meredig, A. Thompson, J.W. Doak, M. Aykol, S. Ruhl, C. Wolverton, The open quantum materials database (OQMD): assessing the accuracy of DFT formation energies, npj Comput. Mater. 1(1) (2015) 15010.
[71] L. Chanussot, A. Das, S. Goyal, T. Lavril, M. Shuaibi, M. Riviere, K. Tran, J. Heras-Domingo, C. Ho, W. Hu, A. Palizhati, A. Sriram, B. Wood, J. Yoon, D. Parikh, C.L. Zitnick, Z. Ulissi, Open catalyst 2020 (OC20) dataset and community challenges, ACS Catal. 11(10) (2021) 6059-6072.
[72] R. Mitra, A. Bajpai, K. Biswas, Machine learning based approach for phase prediction in high entropy borides, Ceram. Int. 48(12) (2022) 16695-16706.
[73] W. Hume Rothery, G.W. Mabbott, K.M. Channel Evans, H.C.H. Carpenter, The freezing points, melting points, and solid solubility limits of the alloys of sliver and copper with the elements of the b sub-groups, Philos. Trans. R. Soc. London, Ser. A 233(721-730) (1997) 1-97.
[74] Y. Zhang, Y.J. Zhou, J.P. Lin, G.L. Chen, P.K. Liaw, Solid-solution phase formation rules for multi-component alloys, Adv. Eng. Mater. 10(6) (2008) 534-538.
[75] S. Guo, C.T. Liu, Phase stability in high entropy alloys: Formation of solid-solution phase or amorphous phase, Prog. Nat. Sci.: Mater. Int. 21(6) (2011) 433-446.
[76] S. Guo, C. Ng, J. Lu, C.T. Liu, Effect of valence electron concentration on stability of fcc or bcc phase in high entropy alloys, J. Appl. Phys. 109(10) (2011) 103505.
[77] S.H. Jhi, J. Ihm, S.G. Louie, M.L. Cohen, Electronic mechanism of hardness enhancement in transition-metal carbonitrides, Nature 399(6732) (1999) 132-134.
[78] K. Balasubramanian, S.V. Khare, D. Gall, Valence electron concentration as an indicator for mechanical properties in rocksalt structure nitrides, carbides and carbonitrides, Acta Mater. 152 (2018) 175-185.
[79] X.L. Gu, C. Liu, H. Guo, K. Zhang, C.F. Chen, Sorting transition-metal diborides: New

descriptor for mechanical properties, Acta Mater. 207 (2021).
[80] N. De Leon, X.X. Yu, H. Yu, C.R. Weinberger, G.B. Thompson, Bonding effects on the slip differences in the B1 monocarbides, Phys. Rev. Lett. 114(16) (2015) 165502.
[81] R. Mitra, A. Bajpai, K. Biswas, Machine learning based approach for phase prediction in high entropy borides, Ceram. Int. (2022).
[82] Y.X. Zuo, M.D. Qin, C. Chen, W.K. Ye, X.G. Li, J. Luo, S.P. Ong, Accelerating materials discovery with Bayesian optimization and graph deep learning, Mater. Today 51 (2021) 126-135.
[83] C. Rudin, Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead, Nat. Mach. Intell 1(5) (2019) 206-215.
[84] J.A. Esterhuizen, B.R. Goldsmith, S. Linic, Interpretable machine learning for knowledge generation in heterogeneous catalysis, Nat. Catal. 5(3) (2022) 175-184.
[85] X. Zhao, S. Yu, J. Zheng, M.J. Reece, R.-Z. Zhang, Machine learning of carbon vacancy formation energy in high-entropy carbides, J. Eur. Ceram. Soc. 43(4) (2023) 1315-1321.
[86] R. Jaafreh, Y.S. Kang, J.-G. Kim, K. Hamad, Machine learning guided discovery of super-hard high entropy ceramics, Mater. Lett. 306 (2022).
[87] Y. Zuo, M. Qin, C. Chen, W. Ye, X. Li, J. Luo, S.P. Ong, Accelerating materials discovery with Bayesian optimization and graph deep learning, Mater. Today (2021).
[88] T. Xie, J.C. Grossman, Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties, Phys. Rev. Lett. 120(14) (2018) 145301.
[89] C.W. Park, M. Kornbluth, J. Vandermause, C. Wolverton, B. Kozinsky, J.P. Mailoa, Accurate and scalable graph neural network force field and molecular dynamics with direct force architecture, npj Comput. Mater. 7(1) (2021) 73.
[90] K. Choudhary, B. DeCost, Atomistic line graph neural network for improved materials property predictions, npj Comput. Mater. 7(1) (2021) 185.
[91] C. Chen, W.K. Ye, Y.X. Zuo, C. Zheng, S.P. Ong, Graph networks as a universal machine learning framework for molecules and crystals, Chem. Mater. 31(9) (2019) 3564-3572.
[92] H. Wang, L.F. Zhang, J.Q. Han, W.N. E, DeePMD-kit: A deep learning package for many-body potential energy representation and molecular dynamics, Comput. Phys. Commun. 228 (2018) 178-184.
[93] L. Zhang, J. Han, H. Wang, R. Car, W. E, Deep potential molecular dynamics: A scalable model with the accuracy of quantum mechanics, Phys. Rev. Lett. 120(14) (2018) 143001.
[94] E.A. Olivetti, J.M. Cole, E. Kim, O. Kononova, G. Ceder, T.Y.-J. Han, A.M. Hiszpanski, Data-driven materials research enabled by natural language processing and information extraction, Applied Physics Reviews 7(4) (2020) 041317.
[95] Y. Zhang, X. He, Z. Chen, Q. Bai, A.M. Nolan, C.A. Roberts, D. Banerjee, T. Matsunaga, Y. Mo, C. Ling, Unsupervised discovery of solid-state lithium ion conductors, Nat. Commun. 10(1)

(2019) 5260.
[96] Z.L. Wang, J.F. Cai, Q.X. Wang, S.C. Wu, J.J. Li, Unsupervised discovery of thin-film photovoltaic materials from unlabeled data, npj Comput. Mater. 7(1) (2021) 128.
[97] A. Vasylenko, J. Gamon, B.B. Duff, V.V. Gusev, L.M. Daniels, M. Zanella, J.F. Shin, P.M. Sharp, A. Morscher, R. Chen, A.R. Neale, L.J. Hardwick, J.B. Claridge, F. Blanc, M.W. Gaultois, M.S. Dyer, M.J. Rosseinsky, Element selection for crystalline inorganic solid discovery guided by unsupervised machine learning of experimentally explored chemistry, Nat. Commun. 12(1) (2021) 5561.
[98] X. Jia, Y.S. Deng, X. Bao, H.H. Yao, S. Li, Z. Li, C. Chen, X.Y. Wang, J. Mao, F. Cao, J.H. Sui, J.W. Wu, C.P. Wang, Q. Zhang, X.J. Liu, Unsupervised machine learning for discovery of promising half-Heusler thermoelectric materials, npj Comput. Mater. 8(1) (2022) 34.
[99] W.J. Huang, P. Martin, H.L.L. Zhuang, Machine-learning phase prediction of high-entropy alloys, Acta Mater. 169 (2019) 225-236.
[100] P.C. Vilalta, S. Sheikholeslami, K.S. Ruiz, X.C. Yee, M. Koslowski, Machine learning for predicting the critical yield stress of high entropy alloys, J. Eng. Mater. Technol. 143(2) (2021).
[101] M. Wagih, C.A. Schuh, Learning grain-boundary segregation: From first principles to polycrystals, Phys. Rev. Lett. 129(4) (2022) 046102.
[102] Y. Li, B. Holmedal, B. Liu, H. Li, L. Zhuang, J. Zhang, Q. Du, J. Xie, Towards high-throughput microstructure simulation in compositionally complex alloys via machine learning, Calphad 72 (2021) 102231.
[103] M.D. Hossain, T. Borman, A. Kumar, X. Chen, A. Khosravani, S.R. Kalidindi, E.A. Paisley, M. Esters, C. Oses, C. Toher, S. Curtarolo, J.M. LeBeau, D. Brenner, J.P. Maria, Carbon stoichiometry and mechanical properties of high entropy carbides, Acta Mater. 215 (2021) 117051.
[104] S. Zhao, Defect energetics and stacking fault formation in high-entropy carbide ceramics, J. Eur. Ceram. Soc. 42(13) (2022) 5290-5302.
[105] P.B. Jørgensen, M.N. Schmidt, O. Winther, Deep Generative Models for Molecular Science, Molecular Informatics 37(1-2) (2018) 1700133.
[106] Z. Ren, S.I.P. Tian, J. Noh, F. Oviedo, G. Xing, J. Li, Q. Liang, R. Zhu, A.G. Aberle, S. Sun, X. Wang, Y. Liu, Q. Li, S. Jayavelu, K. Hippalgaonkar, Y. Jung, T. Buonassisi, An invertible crystallographic representation for general inverse design of inorganic crystals with targeted properties, Matter 5(1) (2022) 314-335.
[107] A.E.A. Allen, A. Tkatchenko, Machine learning of material properties: Predictive and interpretable multilinear models, Sci. Adv. 8(18) (2022) eabm7185.
[108] E.A. Holm, In defense of the black box, Science 364(6435) (2019) 26-27.
[109] G. Vazquez, P. Singh, D. Sauceda, R. Couperthwaite, N. Britt, K. Youssef, D.D. Johnson, R. Arroyave, Efficient machine-learning model for fast assessment of elastic properties of

high-entropy alloys, Acta Mater. 232 (2022) 117924.
[110] T.Z. Khan, T. Kirk, G. Vazquez, P. Singh, A.V. Smirnov, D.D. Johnson, K. Youssef, R. Arroyave, Towards stacking fault energy engineering in FCC high entropy alloys, Acta Mater. 224 (2022) 117472.
[111] V.A. Mints, J.K. Pedersen, A. Bagger, J. Quinson, A.S. Anker, K.M.Ø. Jensen, J. Rossmeisl, M. Arenz, Exploring the composition space of high-entropy alloy nanoparticles for the electrocatalytic H_2/CO oxidation with bayesian optimization, ACS Catal. (2022) 11263-11271.
[112] C. Wang, W. Ping, Q. Bai, H. Cui, R. Hensleigh, R. Wang, A.H. Brozena, Z. Xu, J. Dai, Y. Pei, C. Zheng, G. Pastel, J. Gao, X. Wang, H. Wang, J.-C. Zhao, B. Yang, X. Zheng, J. Luo, Y. Mo, B. Dunn, L. Hu, A general method to synthesize and sinter bulk ceramics in seconds, Science 368(6490) (2020) 521-526.
[113] K. McCullough, T. Williams, K. Mingle, P. Jamshidi, J. Lauterbach, High-throughput experimentation meets artificial intelligence: A new pathway to catalyst discovery, Phys. Chem. Chem. Phys. 22(20) (2020) 11174-11196.
[114] K. Tran, Z.W. Ulissi, Active learning across intermetallics to guide discovery of electrocatalysts for CO_2 reduction and H_2 evolution, Nat. Catal. 1(9) (2018) 696-703.
[115] C. Wen, Y. Zhang, C.X. Wang, D.Z. Xue, Y. Bai, S. Antonov, L.H. Dai, T. Lookman, Y.J. Su, Machine learning assisted design of high entropy alloys with desired property, Acta Mater. 170 (2019) 109-117.
[116] T. Lookman, P.V. Balachandran, D.Z. Xue, R.H. Yuan, Active learning in materials science with emphasis on adaptive sampling using uncertainties for targeted design, npj Comput. Mater. 5(1) (2019) 21.