Journal of Materials Science & Technology 79 (2021) 178–190

Contents lists available at ScienceDirect

Journal of Materials Science & Technology

journal homepage: www.jmst.org

Research Article

# Accelerating materials discovery using machine learning

Yongfei Juan[a], Yongbing Dai[a], Yang Yang[b], Jiao Zhang[a,*]

[a] Shanghai Key Lab of Advanced High-temperature Materials and Precision Forming, Shanghai Jiao Tong University, Shanghai, 200240, China
[b] Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, 200240, China

ARTICLE INFO

*Article history:*
Received 14 July 2020
Received in revised form 16 October 2020
Accepted 7 December 2020
Available online 24 December 2020

*Keywords:*
Materials discovery
Materials design
Materials properties prediction
Machine learning
Data-driven

ABSTRACT

The discovery of new materials is one of the driving forces to promote the development of modern society and technology innovation, the traditional materials research mainly depended on the trial-and-error method, which is time-consuming and laborious. Recently, machine learning (ML) methods have made great progress in the researches of materials science with the arrival of the big-data era, which gives a deep revolution in human society and advance science greatly. However, there exist few systematic generalization and summaries about the applications of ML methods in materials science. In this review, we first provide a brief account of the progress of researches on materials science with ML employed, the main ideas and basic procedures of this method are emphatically introduced. Then the algorithms of ML which were frequently used in the researches of materials science are classified and compared. Finally, the recent meaningful applications of ML in metal materials, battery materials, photovoltaic materials and metallic glass are reviewed.



**Contents**



* Corresponding author.
*E-mail address:* 18817307579@163.com (J. Zhang).

https://doi.org/10.1016/j.jmst.2020.12.010
1005-0302/

## 1. Introduction

The researches of materials science have past three paradigms, including experimental science, mathematical theory and simulation [1]. The first-paradigm study is totally relying on the intuitive observation experience, but no scientific quantification basis. Up till a few centuries ago, the physical models characterized by mathematical equations started to form, which providing certain theoretical foundations such as thermodynamics rules for materials research [2]. Then the third scientific paradigm is coming as the invention of computers decades ago, allowing the simulation of complex practical problems based on the theory gained in the second paradigm. At this period, materials research by big data computing emerged with two noteworthy methods of density functional theory and molecular dynamics proposed [3–5]. Nevertheless, the study processes of traditional materials science, which mostly rely on trial-and-error methods, always take 15–25 years or even longer period of time from research to application and numerous results which seemed incorrect have not been utilized in effect [6]. Meanwhile, more and more material characterization techniques are formulated due to the huge data and high dimensions in materials research [7–11]. For material calculation and simulation methods, such as first principles, molecular dynamics, phase-field theory and finite element analysis, even though the structures and performance of materials can be calculated and predicted at different scales, the models are usually specific to the given material systems. It still cannot meet the requirements of multifarious description for varied properties, so further researches of materials are greatly restricted. Nowadays, materials research has come to the fourth scientific paradigm with the promotion and popularization of artificial intelligence, these problems are tended to be cracked.

Nowadays, materials research is coming to the stage of big data-driven science also known as the fourth paradigm of materials science benefited from the mass data generated by experiments and calculation methods such as the first principles and molecular dynamics theory [12–15]. Besides, the data are growing faster than ever before with more application of high-throughput computing in materials research [16]. Applying machine learning (ML) in the prediction of material properties and explore new materials have become the hot research spots in the intersection field of materials science and computer science as the revolution of artificial intelligence [17–20]. The basic links of artificial intelligence, computing science and big data science are shown in Fig. 1. Combined with the empirical theory excited in materials science and the intelligent computing methods of computer science, making full use of the advantages of the two subjects to interpret material data more efficient is one of the major topics in this area. In the researches of materials science, the establishment of the classical model relies much on the physical mechanism such as conservation law and thermodynamics, then deriving the mathematical formula of parameter regression from the data [21–23]. In comparison, applying ML to study the regularities and rules of material data does not require any specific principles or physical insights, but only training the models with the form of flexibility and usually nonlinear from the available data. With these superiorities, ML has become a prominent means in predicting material properties, selecting materials and optimizing design [24–26]. At the same time, there are some challenges in the application of traditional ML in materials science. Traditional ML relies heavily on sample data, and the predictive model has poor interpretability due to the lack of relevant domain knowledge or known physical insight [27].

The literature search, as depicted in Fig. 2 which statistics from the ScienceDirect website, demonstrates that the researches of ML and its utilization in materials science both increased significantly in this decade. Considering the increasing status of big data and artificial intelligence in modern times, the purposes of this paper are to reveal the main strategies and basic methods of ML in the application of materials science, as well as other breakthroughs and hot points, are discussed.

## 2. Principle analysis of machine learning in materials science

### 2.1. The realization methods of machine learning

ML, as an important part of artificial intelligence, the methods primarily include supervised learning, unsupervised learning and reinforcement learning.

In supervised learning, each instance is composed of an input object (usually a vector) and an output value also known as the supervised signal. One of the main characteristics of supervised learning is that the data are labeled, including data category, data attribute and feature point location [28,29]. After train the algorithms with the labeled data, the parameters of the algorithms are modified based on the comparison results of the predicted data and the expected ones, then repeats enough times by this way until the algorithms converge to the optimal solution, and finally, a specific model with the ability of intelligent decision-making is obtained.

Contrary to supervised learning, the data in unsupervised learning are unlabeled, and the goal is to seek and deduce the potential connections of the samples [30]. The common methods of unsupervised learning are clustering and dimensionality reduction, in which, analyzing the distribution of data samples in the feature space to gather similar data into one group and separate the data with different types is essential for clustering method due to the data categories are unknown in advance. Moreover, high-dimensional data sets are not rare in ML, then some problems such as sparse data samples and distance calculations are easily arisen, which also called dimension disaster [31–33]. The main theory of dimension reduction in ML is mapping the data points from the original high-dimensional space to the low-dimensional space, and the algorithms of dimension reduction mainly involve singular value decomposition (SVD), principal component analysis (PCA), factor analysis (FA) and independent component analysis (ICA) [34–37].

Different from supervised learning and unsupervised learning, reinforcement learning gets the learning information and constantly modifies the model parameters by receiving the feedback from the environment but not require the data to be provided in advance [38]. Generally speaking, a certain positive incentive is gathered when the machine acts correctly, otherwise, it will give a negative incentive. In this case, some dynamic planning ideas are generated in the ML algorithms and the action mode which can maximize the incentive is finally chosen. This illustrates that the application of the reinforcement learning method needs less information and is easier to design, which is conducive to deal with more complicated decision problems. Besides, deep reinforcement learning, which combines reinforcement learning with deep learning, is developing rapidly as a research hotspot in the field of artificial intelligence, such as in automatic driving, natural language processing, robot and other realms [39–44].

### 2.2. Common processes of machine learning in materials science

A basic framework of materials discovery and design based on ML methods is shown in Fig. 3, in which, three main steps are mentioned: the construction of samples, the building of algorithm models, models verification and materials prediction.

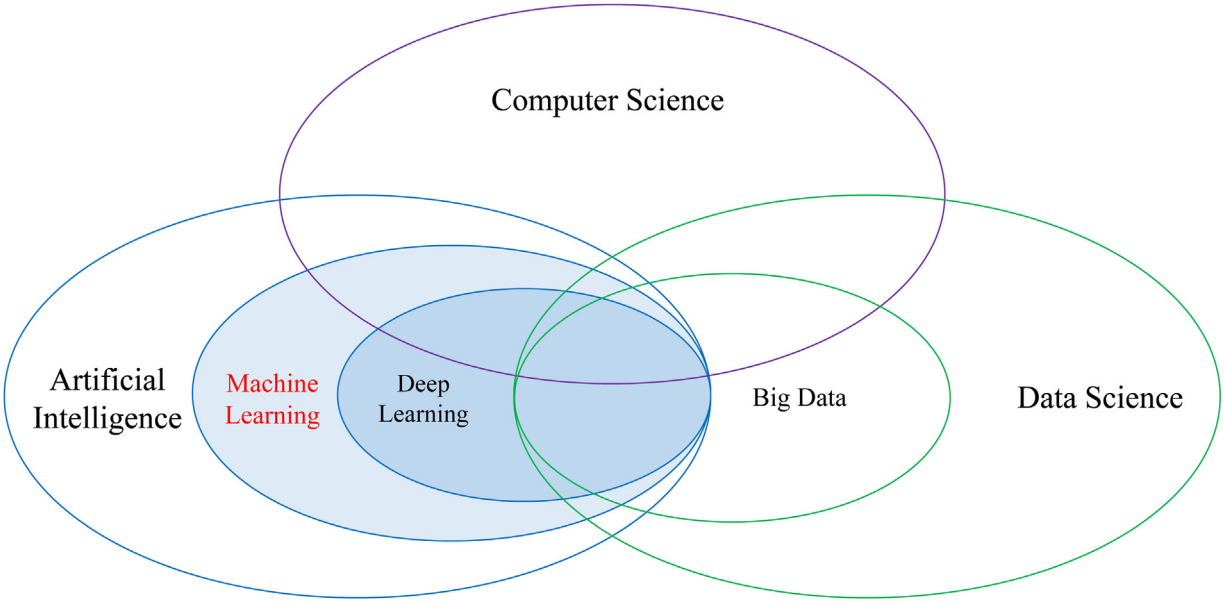


**Fig. 1.** The basic links of artificial intelligence, computing science and big data science.

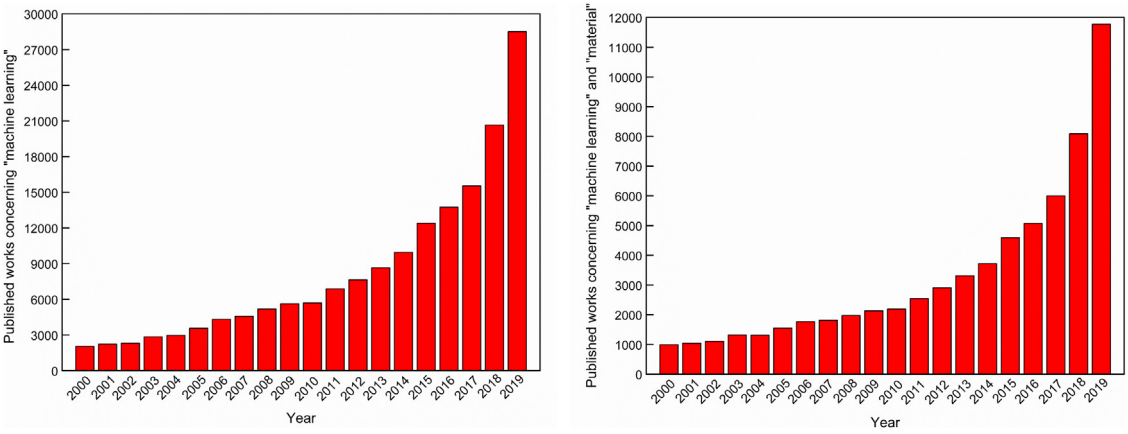


**Fig. 2.** Number of published works on “machine learning” and “machine learning” + “material”.

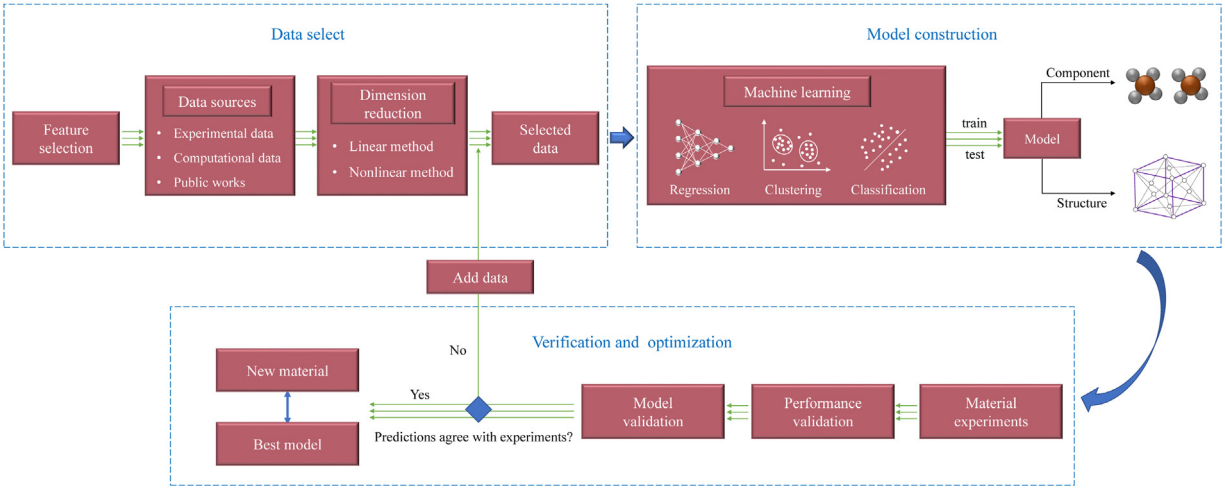


**Fig. 3.** The basic framework of materials discovery and design based on ML methods.

材料属性和相关信息的数据库



**Table 1**
The database of material properties and related informations.

| Database names | Websites | Database category |
|---|---|---|
| NIST Standard Reference Data | http://www.nist.gov/srd/dblistpcdatabases.cfm | Standard data |
| Materials Project | http://www.materialsproject.org | Calculation |
| AFLOWLIB | http://www.aflowlib.org | Calculation |
| Harvard Clean Energy Project | http://cepdb.molecularspace.org | Calculation |
| Open Quantum Materials Database | http://www.oqmd.org | Calculation |
| Citrination | http://www.citrination.com | Material data |
| CRC Handbook | http://www.hbcpnetbase.com | Material data |
| Springer Materials | http://materials.springer.com | Material data |
| Pauling File | http://www.paulingfile.com | Material data |
| Granta CES Selector | http://www.grantadesign.com/products/ces | Material data |
| Matbase | http://www.matbase.com | Material data |
| NIST Materials Data Repository | https://materialsdata.nist.gov | Material data |
| Total Materia | http://www.totalmateria.com | Material data |
| MatNavi (NIMS) | http://mits.nims.go.jp/index_en.html | Material data |
| MatWeb | http://www.matweb.com | Material data |
| Open KIM | http://www.openkim.org | Material simulation |
| Knovel | http://app.knovel.com/web/browse.v | Material engineering |
| AIST Research Information Databases | http://www.aist.go.jp/aist_e/list/database/riodb | Material characterization |
| Reaxys | http://www.elsevier.com/solutions/reaxys | Chemical data |
| Scifinder/ChemAbstracts | http://scifinder.cas.org | Chemical data |
| ChemSpider | http://www.chemspider.com | Chemical structure |
| TE Design Lab | http://www.tedesignlab.org | Thermodynamics |
| CALPHAD | http://www.opencalphad.com/index.html | Thermodynamics |
| Powder Diffraction File (PDF) | http://www.icdd.com/products/index.htm | Crystallography |
| Inorganic Crystal Structure Database | http://cds.dl.ac.uk/cds/datasets/crys/icsd/llicsd.html | Crystallography |
| Cambridge Crystallographic Data Centre | http://www.ccdc.cam.ac.uk | Crystallography |
| CrystWorks | https://cds.dl.ac.uk | Crystal data |
| Crystallography Open Database | http://www.crystallography.net | Crystal data |
| ASM Alloy Database | https://www.asminternational.org/materials-resources/onlinedatabases | Alloy material |
| CINDAS Alloys Database | http://www.cindasdata.com/products/hpad | Alloy material |
| NanoHUB | http://www.nanohub.org | Nanometer material |
| Nanomaterials Registry | http://www.nanomaterialregistry.org | Nanometer material |
| Mindat | http://www.mindat.org | Mineral |
| American Mineralogist Crystal Structure Database | http://rruff.geo.arizona.edu/AMS/amcsd.phP | Mineral |
| UCSB-MRL Thermoelectric Database | http://www.mrl.ucsb.edu:8080/datamine/thermoelectric.jsp | Thermoelectric materials |
| International Glass Database System | http://www.newglass.jp/interglad_n/gaiyo/info_e.html | Glass |
| SUNCAT | http://suncat.stanford.edu/theory/it-facilities | Catalyzer |
| Handbook of Optical Constants of Solids | / | Reference book |
| Metallurgical Thermochemistry | / | Reference book |
| Pearson's Handbook: Crystallographic Data | / | Reference book |

### 2.2.1. Sample construction

The database information corresponding to specific materials or material properties is summarized in Table 1. In the construction of samples, the ideal scenario is that fewer data have been selected but the model worked best. According to the prospective objectives such as prediction performance or discovery of new materials, the characteristic values, in materials science, are determined firstly, and then the data with high correlations are identified and selected as the input samples of ML models [45].

In this step, dimension reduction processing is the crucial part to reduce the quantity of the data, as well as improve the operating speed and the prediction accuracy of the models. Besides, more redundant information and noise are tended to be produced, such as repeated values, missing values, out-of-range outliers and labeling errors, if dimension reduction processing is not applied [46–48]. To reduce the sample capacity, Balcazar [49] proposed a fast sampling reduction method, in which, the training set was randomly sampled and the sampling results were replaced by the original training set. Different from random sampling method, Lyhyaoui [50] put forward a sample selection method based on the clustering technology that the whole training set was clustered firstly with the corresponding algorithms, and then the vectors which closed to the other clusters were selected with the comparison of gaps between the cluster center and other clusters. Feature selection and feature transformation are essentially two ways of feature reduction. Selecting appropriate descriptors or features is one of the central problems in exploring the structure activity relationship of materials using ML models. Shi et al. [51], proposed a data-driven multi-layer feature selection method. Users can input training data without manually adjusting the super parameters. The domain expert knowledge was quantified by weighted score and integrated into the selection process to eliminate the risk of crucial features being deleted.

### 2.2.2. Models building

The essential step for the application of ML in materials science is to build a proper model, which is demanded to characterize the relationships between the injection features (such as atomic size and electronegativity) and the properties of the materials precisely. In this process, the mapping models for objects and targets are built based on a set of known material data. All kinds of ML methods, regardless of simple linear and nonlinear regression or highly complex kernel ridge regression and neural network, can be employed to establish this mapping. And its difference from the traditional material computation methods is that the mapping relationship is essentially a black box, where the injection features and the output data are linked with the specific nonlinear or linear functions [52,53]. In the typical researches of materials science, there is always a complex connection between the condition factors and the target attributes, which is difficult to expound with traditional methods. In comparison, almost all ML algorithms can be translated to the optimum solutions of specific problems, where the objective function is constructed to get the model parameters of ML algorithms. Then how to construct a reasonable objective function is the critical factor to establish a suitable ML algorithm. Once the objective function is determined, the next step is to solve the opti-

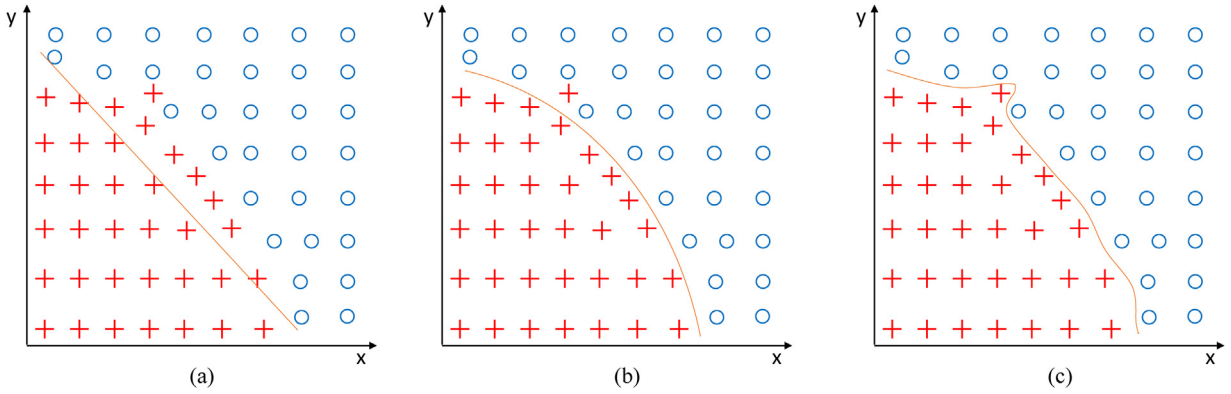


**Fig. 4.** The results of ML researches which mainly include (a) underfitting, (b) overfitting and (c) correct fitting.

mization problems, which generally has a ready-made scheme in mathematics [54–57]. That is, the relationship between condition factors and decision attributes are modeled with the ML algorithms based on the given samples, and allowing the solution of uncharted complex problems is the core superiority to design and discover materials with ML.

#### 2.2.3. Models evaluation

The main task of ML is to get generalized models. As shown in Fig. 4, the results of ML researches mainly include underfitting, overfitting and correct fitting. Underfitting means that the models cannot fit the training samples very well, and the prediction of data is not accurate enough. Overfitting denotes that the training samples are fitted relatively well with the ML models, but the prediction accuracy of the new data is rather poor. The appearance of overfitting in algorithms illustrates that the optimization models are too complex, where the fitting results of the training data are pretty good, but the generalization ability of the models is insufficient [58–60].

Models evaluation can be understood as measuring the generalization ability of the models. The commonly used methods include simple hold-out verification, k-fold verification and repeated k-fold verification with disordered data [61–63]. Hold-out verification method is to set aside a certain proportion of data as a test set. The model is trained on the remaining data, and then evaluated based on the test set. As with hold-out verification, k-fold verification also requires independent validation set for model calibration [64]. In K-fold validation, the data is divided into k partitions with the same size. For each partition i, the model is trained on the remaining k-1 partitions, and then the model is evaluated on the partition i. The final score is equal to the average of K scores. Iterated k-fold validation with shuffling can be selected when the available data is relatively small. In this method, k-fold verification is used many times and the data is scrambled before dividing the data into k partitions. The final score is the average value obtained after k-fold cross validation, but the calculation cost of this method is relatively high. Besides, there is a special cross validation method named bootstrapping [65]. One sample is randomly selected each time from the data set which containing m samples, then put back to the data set. In this way, samples are taken m times to form a new data set as the training set. This method is easy to operate, but the application of bootstrapping needs to be based on many statistical hypotheses.

## 3. Commonly used machine learning algorithms in materials science

Algorithms are the concrete calculation method of learning models in ML. ML is in fact to solve the optimization problems, and the main purpose of ML algorithms is to find the global optimal solution and ensure the efficiency of the solution processes [66]. The tasks of ML include four main tasks based on the most common application: classification, regression, clustering and probability estimation, as shown in Fig. 5. The characteristics and main applicable situations of the algorithms are summarized in Table 2. However, it should be emphasized that many algorithms can solve classification problems as well as deal with regression problems, such as random forest.

### 3.1. Support vector machine

In ML algorithms, support vector machine (SVM) possesses many special advantages in solving the problems of small sample, nonlinear and high-dimensional pattern recognition, and it can be extended to other ML problems such as function fitting [67,68]. As an important application in SVM, support vector classification (SVC), which belongs to binary classification algorithms, is probably the most powerful classifier. In SVC, the data points are divided into two groups with ($n$-1) dimensional hyperplane for the sample sets in $n$-dimensional coordinates, where hyperplane is the line to divide input variable space [69–71]. For instance, all the input points can be completely separated by hyperplane ($\omega \cdot x + b=0$) in two dimensions as shown in Fig. 6(a), and undeniably that the hyperplane for the linearly separable data is infinite, but the separated hyperplane with the largest geometric spacing is unique. In practical problems, nevertheless, the factors affected material properties are not single, and the data are usually not linearly separable in the multi-dimensional space [72–75]. That is to say, only one hyperplane is not enough to complete the required classification task. Then use kernel technique to map the data from the input space to the higher dimensional space through a specific function and search for the hyperplane in higher dimension space is a feasible solution for it, but high computational cost is easily to be brought about with this method [76]. Therefore, a kernel function is introduced, on the premise that the calculation of SVC only involves the inner-product calculation, to transform the inner-product calculation that in the high-dimensional feature space into the non-linear transformation of the inner-product operation of the data ($x$, $y$) in the low-dimensional input space [77–79]. Then the most decisive work is to search the appropriate kernel function for SVC.

The algorithm of SVM is extremely effective not only in dealing with binary classification tasks but also in linear regression, that is support vector regression (SVR), which is especially appropriate for the samples with relatively small amount of data [80–83]. As shown in Fig. 6(b), a 2 $\varepsilon$ wide interval is built with $f(x)$ as the centerline, then the training sample which falls into the interval is considered to be predicted correctly. The difference between SVR

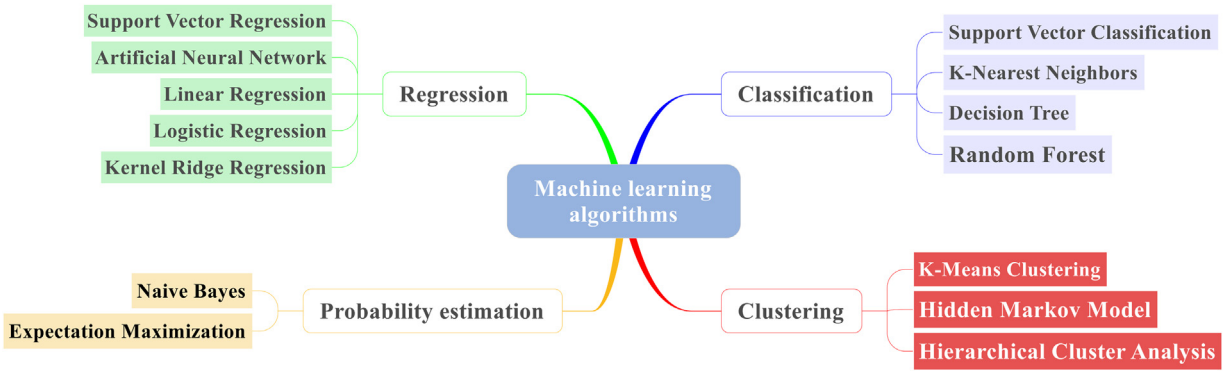


**Fig. 5.** The ML algorithms commonly employed in materials science.

**Table 2**
Applications and features of the main ML algorithms.

| Method | Category | Applicable situations and features |
|---|---|---|
| Support Vector Regression | Regression | SVR is a nonlinear algorithm which can dealing with the small-volume data and has the feature of resistance to over-fitting. |
| Artificial Neural Network | Regression | Large data are needed, the self-study and fault-tolerance abilities are relatively strong for ANN although the interpretability is weak. |
| Linear Regression | Regression | Needs strict assumptions and the data should be linearly conformable, the features of fast modeling and good interpretability are significant. |
| Logistic Regression | Regression | This model can easily be utilized in classification problems, but it cannot deal with the data contained multiple features or variables. |
| Kernel Ridge Regression | Regression | KRR can deal with the nonlinear data, but the prediction speed is lower than SVR when the volume of data is large. |
| Support Vector Classification | Classification | SVC also known as maximum margin classifier is an important classification model, especially for two-class data. |
| K-Nearest Neighbors | Classification | KNN is suitable for multi-classification model, but its calculating quantity is large and the requirement of sample balance is high. |
| Decision Tree | Classification | DT can deal with the data with missing attributes and the interpretability is good, but it does not support online learning and is easier to be over-fitting. |
| Random Forest | Classification | RF not only has the advantages that DT possessed, but also can prevent the over-fitting when there are small noises. |
| K-Means Clustering | Clustering | K-Means is a classical clustering algorithm which has the features of simple and fast, but this algorithm is quite sensitive to initial data. |
| Hierarchical Cluster Analysis | Clustering | HCA can complete the whole clustering process at one time by building the hierarchy of clusters, but the computational capacity is quite large. |
| Hidden Markov Model | Clustering | HMM is an important stochastic model of signal and it has a wide range of applications in pattern recognition. |

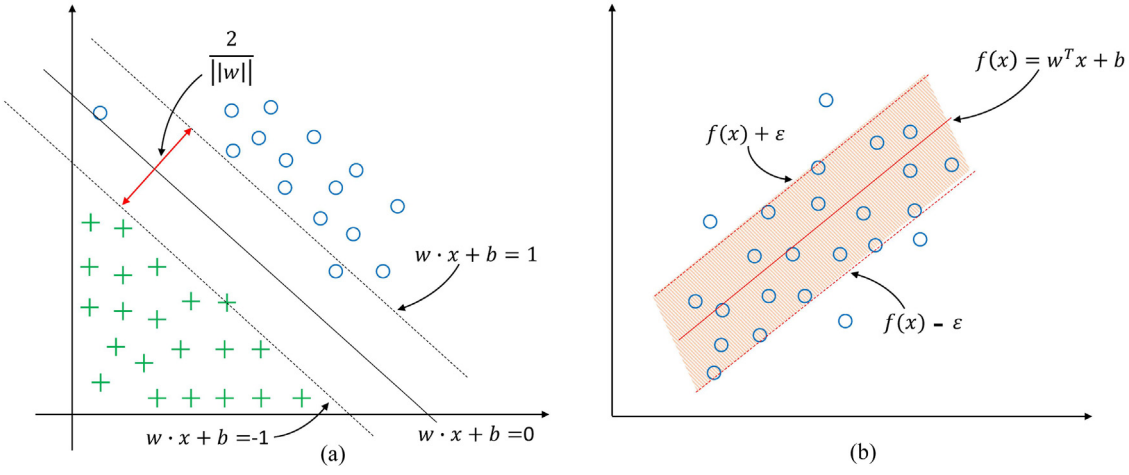


**Fig. 6.** The working principle of SVM: (a) the algorithm of SVC, (b) the algorithm of SVR.

and SVC is mainly that the sample points of SVR are ultimately to one class, and the optimal hyperplane is not to make the two or more classes of sample points separated precisely like SVC, but the total deviation of all sample points from the hyperplane is the smallest.

### 3.2. Decision tree

Decision tree (DT) is a commonly used algorithm in classification and regression, which belongs to the supervised learning. In the application of ML, the feature selections of quantitative evaluation

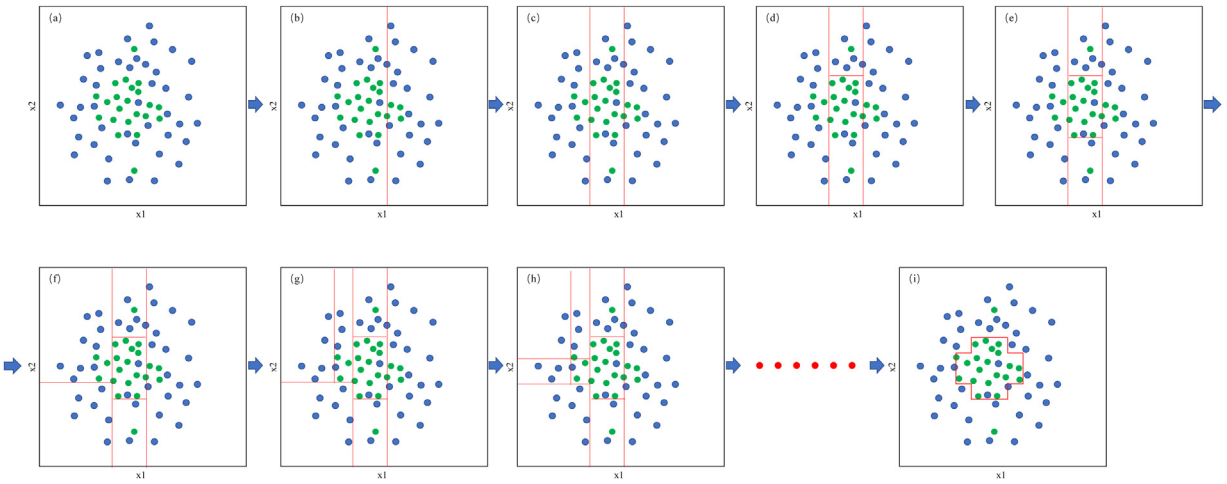


**Fig. 7.** The working principle of the algorithm of DT.

methods are different, resulting in extensive DT has been derived, such as ID3 (feature selection by information gain), C4.5 (feature selection by information gain ratio) and CART (feature selection by Gini index). In which, ID3 and C4.5 belong to model algorithms that using the tree structures to classify samples, while the CART algorithm is applicable for both classification and regression [84–87]. As shown in Fig. 7, selecting certain features to distinguish the data and then divided into different sub-nodes according to the results of the discrimination is the working principle of DT. In this algorithm, each sub-node represents a value of the features, after constantly classifying with features selection until reached the leaf node, all the data are divided into corresponding categories. Put simply, DT represents the conditional probability distribution of different categories under the specific characteristics, where different DT corresponds to the classification models with different complexity [88–90]. Furthermore, there are multiple models that do not contradict each other for the same training data sets in the application processes of DT, therefore, it is critical to select the model with the characteristics of good fitting results and strong generalization ability based on the actual situations.

DT learning is a method of top-down recursive, the basic idea is to construct a tree with the most rapid decline in entropy when taking information entropy as the measure, and the entropy is 0 at the leaf node [91,92]. It possesses the advantages of fast classification speed and readability. However, repeated training is required in this algorithm to determine the tree structures and various parameters, which is easy to cause the overfitting problems and effecting the generalization ability [93]. At this point, the methods of pruning operation or build random forest are good choices.

### 3.3. *Random forest*

Random forest (RF) algorithm, which takes DT as the unit, is built based on the idea of integrated learning with multiple trees consisted in [94–96]. Same as the algorithm of DT, RF is applicable for both regression and classification problems. In this algorithm, k features are randomly selected from the data sets [97]. Firstly, DT is established and repeated n times according to these k features. Then the prediction results of each DT with the input of random variables are stored, and the number of votes for each prediction target is calculated, finally, the prediction target with the highest votes is regarded as the final forecast result. For the problems of regression, after the prediction of the output values calculated with DT, final forecast results are determined by calculating the average value of all the DT in RF [98]. On the other side, the characteristics

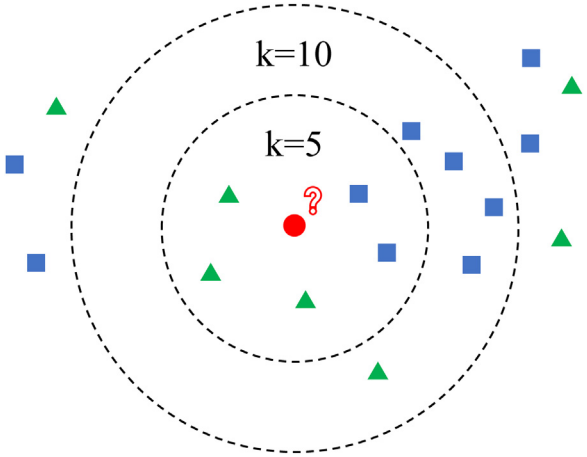


**Fig. 8.** The working principle of the algorithm of KNN.

of the latest data are predicted with every DT, and the one which selected most is considered to be the final recognition outcome of the latest data.

Aside from the abilities to solve the problems of classification and regression, the RF algorithm can deal with the classification and numerical characteristics at the same time. Besides, only at the situation that more than half of the base classifiers make an error, can get the forecasters wrong, which denotes that the stability of RF is perfect [99]. Nevertheless, RF is obviously more complex than the DT algorithm, more time and cost of calculation to train than other similar algorithms is required.

### 3.4. k-*Nearest Neighbor*

*k*-Nearest Neighbor (KNN), as shown in Fig. 8, is one of the simplest supervised classification algorithms in ML. The premise of this algorithm is to build a training data set which labeled with categories, and the distance between the test object and the data in the training set is indicated by Euclidean distance or cosine distance, in which, simple Euclidean distance is the most frequently used method [100–103]. Afterwards, find out the nearest *K* objects as neighbors of the test data, and the category of the object with the highest frequency among the *K* objects is regarded as the category of the test data. The approximation error is decreased when *k* is too small, but it can significantly increase the error of estimation. On the contrary, the approximation error will increase if *K* is set

to a large value, which might result in the generation of noise and the reduction of the classification effect. Accordingly, the cross-test (based on $k$ = 1) method is usually applied to set the value of $K$ in KNN, and the experience shows that $K$ is generally lower than the square root of the number of training samples [104].

KNN algorithm, which possesses the advantages of simple design theory, is easy to understand and realized with no estimate parameters, and this algorithm is well suitable for the classification problems of unusual data. Nevertheless, the deficiencies of this algorithm are also extremely obvious, KNN belongs to lazy algorithms, which demand a greater time to scan all the training samples to calculate the distance, so a lot of calculation and memory is required. Clearly, the choice of the data is important for all ML algorithms, KNN especially require the correctness of the data due to the fault tolerance of this algorithm to the training data is terribly poor [105–108]. Besides, the interpretability of KNN is relatively weak when compared with the algorithms of DT and RF.

### 3.5. Artificial neural network

Artificial neural network (ANN) is a computational model derived from simulating the biological neural network of human brain processing information [109,110]. In this algorithm, the neuron is abstracted as a node, then the methods of function fitting and estimation are applied to convert the input samples into the desired output data with the help of a large number of nodes [111]. The structure of ANN is shown in Fig. 9, in which, the input node receives the input data information, the processing results of the information are saved by the output node, and the intermediate procedures between the input node and the output node are called hidden nodes. All kinds of the nodes in ANN are distributed in the layers which can be connected with lines, this method of nonlinear mapping is corresponding to the synapses in the neural structures [112]. And there is an evaluation to adjust the weight of each mapping, then the storage contents of the hidden nodes are regulated. Actually, the learning processes of ANN are to optimize the whole network model by correcting the weights of the nodes in each layer with training data. The methods mainly include adjusting the weight of each synapse, modifying the neural network structures, selecting and changing the nonlinear mapping function [113–116].

The algorithm of ANN, which possesses the function of self-study and associative storage due to the similar working principle with brain, has been widely used in the researches of ML, including the applications of material research, speech recognition, signal processing, automatic control and so on [117–121]. Bassir et al. [122] proposed a new identification hybrid strategy based on genetic algorithm with parallel selection (GAPS) and ANN. ANN is designed in a way to relate the mechanical parameters of the studied material to the cost function representing the difference between the true and simulated mechanical responses. The design method of ANN is to connect the mechanical parameters of the studied materials with the cost function, which represents the difference between the real and simulated mechanical response. In the study of Koksal [123], an ANN model was established to investigate the relationship between processing parameters including chemical composition, temperature and mechanical properties such as bending strength, Young's modulus in magnesia based refractory materials. There were insignificant differences between experimental values and ANN results, which meaning that mechanical properties of refractory materials can be predicted using ANN method. Series of experiments were carried out for a varying percentage of composite fiber to characteristic loading in the work of Kazi et al. [124]. Based on those experimental data, the ANN models were trained and tested by implementing the back-propagation method. Then the models were developed to predict the load-displacement curves for better understanding the behavior of cotton fiber/polyvinyl chloride (PVC) composites.

For materials-design applications various ANNs mainly perform two tasks: classification and interpolation of data. Hundreds of different models related to ANN have been produced since the first neuron model built by McCulloch in 1943. Different ANN architectures are appropriate to particular problems. The prediction results of ANN are extremely accurate with the ability of strong robustness and high tolerance for lack of accuracy (the data may be rough or fuzzy). In the application of ANNs to a particular task, a main obstacle seems to be the lack of reliable databases. If the ANN system is combined with abounding data from reliable sources, the method should effectively help to identify the characteristics of materials [125].

## 4. The application of machine learning in materials science

The application of ML in the materials research is increasingly popular and has made a series of achievements in high entropy alloys, steel materials, superconducting materials, topological insulators, photovoltaic materials, magnetic materials, ferroelectric materials, amorphous materials, thermoelectric materials, catalytic materials, etc. The successful application of ML has been regarded as an innovative mode of materials development. According to the statistics of published papers, the recent meaningful applications of ML in metal materials, battery materials, photovoltaic materials and metallic glass were reviewed in this section.

### 4.1. The application of machine learning in metal materials

Creep rupture life is a key material parameter for service life and mechanical properties of Ni-based single crystal superalloy materials. Shi et al. [126] developed a divide-and-conquer self-adaptive (DCSA) learning method to accelerate the prediction of creep rupture life of Ni-based single crystal superalloys. The DCSA automatically classifies alloy samples with various creep mechanisms into several clusters based on comprehensive factors through K-Means clustering. Then the DCSA self-adaptively chooses the optimal ML model from the models of RF, SVR, Gaussian Process Regression (GPR), Lasso Regression (LR) and Ridge Regression (RR) based on the designed fitness function to reveal the differences in creep mechanisms of alloy samples in different clusters. The works demonstrated the effectiveness of the strategy of divide and conquer in predicting creep rupture life. High entropy alloys (HEAs), which typically consist of at least five major elements with the atomic percentage between 5% and 35 %, has been widely perceived as the new research hot in the designing field of metal materials due to their remarkable mechanical properties, outstanding wear resistance and excellent plasticity. HEAs are usually composed of simple solid solution phases like FCC and BCC, which is considered to be the most prominent feature [127–129]. Previous studies about phases prediction of HEAs mainly focused on trial-and-error experiments or thermodynamic criteria. Thus, it is an extremely significant work to predict the stable structures or the ability of phases formation of HEAs efficiently [130]. Zhang et al. [131] proposed a new system framework, where the best ML models and material descriptor were selected by genetic algorithm (GA), and the effectiveness in the prediction of phase formation in HEAs was demonstrated. Then with the application of optimized classification models, the accuracy for identifying solid-solution phases and non-solid-solution phases HEAs was as high as 88.7 %, and the formation ability of BCC, FCC and dual-phase HEAs were further identified with the accuracy of 91.3 %. And the active learning method was used to improve the accuracy of the classifier iteratively. This method serves as a general algorithm for various material problems including clas-

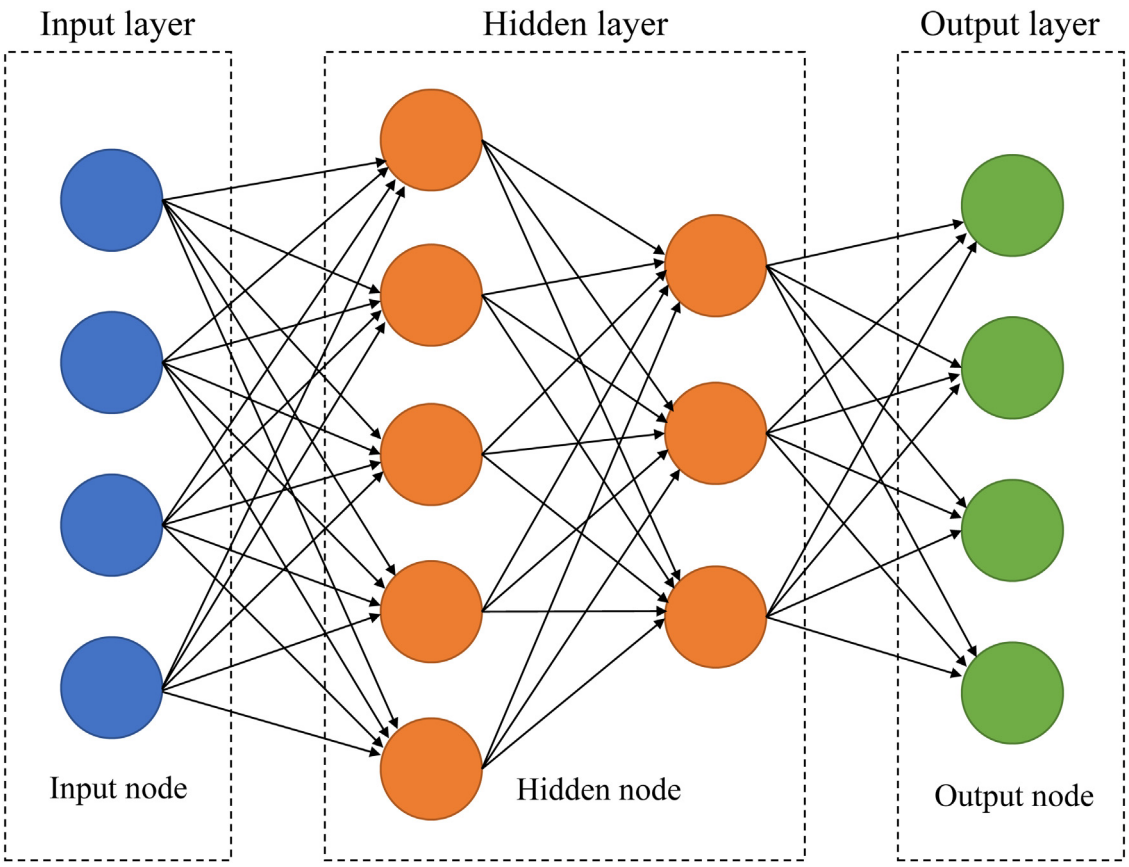


**Fig. 9.** The working principle of the algorithm of ANN.

sification and optimization of targeted properties to select the materials descriptor and the ML models. And one of the key factors to influence the performance of ML models in materials science researches is the set of material descriptors. In the study of Dai et al. [132], a new method based on the feature engineering and the least-squares to predict the formation of HEA phases was proposed. It selected the low dimension descriptor subset from the high dimension descriptor space which composed of several basic functions. The results showed that the performance of the nonlinear descriptor constructed by the linear algorithm was better than that of the original descriptor. But it should be noted that there is still much work to do in this study, a larger dataset may be helpful to further improve the prediction accuracy.

In addition to the application in the researches of new materials, the studies of ML have also made some progress in the field of traditional metal materials, such as steel materials. Wang et al. [133] constructed a new ML tool, which can realize 2D/3D microstructure analysis, the direct analysis of property predictions and the properties-to-microstructure inverse analysis were conducted, that is, Materials Genome Integration System Phase (MIPHA). Then focuses on the analysis of data obtained from MIPHA, the rMIPHA ML program based on R script was proposed. The purpose of this study was to provide a new approach of data-driven materials design for material researchers, so as to speed up the processes of material discovery. As is known, the diagram of time-temperature transition (TTT) plays an important tool in the study of stainless-steel microstructure. This means that it is of great practical significance to predict TTT diagram precisely and quickly. In the work of Huang et al. [134], a combination of ML algorithms, including BP artificial neural network, Random Committee, RF and Bagging, was employed for the prediction of TTT diagram with relevant descriptors containing the alloying elements, austenitizing temperature and holding time. The results proved that the combination of ML methods has high predictive accuracy on the TTT diagrams of stainless steel with a high correlation coefficient and a low error value. In the study of Wang et al. [135], ML was applied to build the universal models to predict the yield strength and total elongation of RAFM steels. After the database with a wide range of compositions and treatment processes was established, the highly correlated features were selected with the feature engineering methods. Then RF algorithm was trained with the selected features. The results showed that RF regressors with feature engineering guided possess the advantages of universality and accuracy for the prediction of yield strength and total elongation.

### 4.2. The application of machine learning in battery materials

The performance of the lithium-ion batteries is highly sensitive to the selection and design of the battery materials, and the relationship between the materials and the battery performance is not easy to determine due to the wide range and complexity of design variables [136]. In which, Lithium-sulfur (Li-S) battery, as one of the most important applications in the field of lithium-ion chemistry benefited from the low cost and the high theoretical specific energy of 2567 Wh $kg^{-1}$, faces various challenges including the rapid capacity degradation and the limited cycle life [137–139]. In the recent work of Kilic et al. [140], a comprehensive analysis about the effects of key factors including the peak discharge capacity and the cycle life on the battery performance was conducted with ML. The results concluded that the application of structured carbon such as porous carbon or carbon nanotubes in the encapsulated cathode has brought superior battery performance, the types and quantities of encapsulation materials were critical for high capacities and enhanced cycle life. They pointed out that the development

of encapsulation cathodes without other adhesives or conductive materials, and the design of new electrolytes which can be successfully realized at a low *E*/*S* ratio perhaps the most promising way.

In other studies, Shandiz et al. [141] proved that the crystal structure system has a significant impact on the physical and chemical properties of lithium-ion silicate cathodes, denoted that the prediction of crystal system must be a meaningful work to estimate the properties of cathodes in batteries. Based on the feature evaluation in the statistical model, which is built with ANN, SVM, RF and other ML algorithms to predict the three main crystal systems including monoclinic, rhombic and triclinic of silica-based cathodes, the close links between the crystal system and other physical properties of the cathode were confirmed. In which, RF and extremely randomized trees gave the highest overall average accuracy among the algorithms. According to the evaluation of the feature's importance in the extremely randomized trees, the volume and the sites number of the crystal showed the highest effect in determining the types of crystal data set.

### 4.3. The application of machine learning in photovoltaic materials

In the third-generation photovoltaic technology, perovskite materials, which initially applied to photovoltaic power generation in 2009, has aroused much research enthusiasm benefited from outstanding optical and electrical properties, as well as the favors of easily synthesis, low cost and rich raw materials [142–146]. Nevertheless, the heterojunction structures of perovskite battery are not stable, and the performance of the battery can be significantly reduced once the heterojunction structures are destroyed. Moreover, a lot of toxicity is produced in the fabrication process of perovskite solar cells, which also greatly restricts the spreads of the notable materials. Based on this, Wu et al. [147] proposed a target-driven method which combined ML algorithms with density functional theory (DFT) calculations to speed up the discovery of hidden hybrid organic-inorganic perovskites (HOIPs) for photovoltaic applications from 230,808 HOIPs candidates. In which, three ML models including SVR gradient lifting regression (GBR) and kernel ridge regression (KRR) were used to predict the band gap of HOIPs candidates, then 132 stable and nontoxic HOIP were verified by DFT calculation with proper band gap of solar cells. In this study, not only a series of stable and non-toxic HOIPs were discovered, but a new HOIPs database also has been constructed, which was conducive to the experimental synthesis and the design of functional materials.

In the other work, a dataset containing long-term stability data for 404 organ-lead halide perovskite cells was constructed from 181 published papers and analyzed with the ML tools including DT and association rule mining [148]. Then the effects of deposition methods, cell manufacturing materials and storage conditions on cell stability were investigated. The results demonstrated that mixed cation perovskites, multi-spin coating as one-step deposition, DMF + DMSO as precursor solution and chlorobenzene as anti-solvent were contributed to the stability of regular cells. Furthermore, the cells stored under low humidity were found to be more stable as expected, where the degradation was slightly faster for inverted cells. Till now, there is no standard test or report protocol for the stability study of perovskite solar cells, which greatly limits the application of ML models in perovskite battery.

### 4.4. The application of machine learning in metallic glass

Metallic glass, as a new amorphous material, has been widely researched owing to its special mechanical, excellent physical and chemical properties [149–152]. However, it is pretty tough for traditional methods to clearly understand the structures and characters of this material especially for the short-range orders which composed of central atoms and the nearest neighbors. In which, glass-forming ability is one of the crucial matters that must be put into consideration, and a single factor to explain the glass-forming energy of amorphous materials is not enough. Facing these issues, the method of Point-Pattern Matching (PPM) to solve the rotation problems of the short-range order unit was proposed by Banadaki et al. [153]. The principle of this method was to arrange two groups of three-dimensional points into the same direction and position as much as possible with an approximate rigid graph matching technique so as to solve the problem of mild disorder. After that, the short-range order units were divided into 30 categories using the ML clustering algorithm of density space clustering (HDBSCAN) based on the similarity between the short-range order units, that is, the metallic glass system was a disordered system composed of these short-range order units 30 categories connected with each other. This study laid a foundation for the building of a clear structure-property relationship in metallic glass.

In the study of Xiong et al. [154], ML approaches were applied to enable the construction of a predictive model to describe the glass-forming ability and elastic moduli of bulk metallic glasses (BMGs) based on a dataset contained 6471 alloys. In which, Glass-forming ability, critical casting diameter, shear moduli, and bulk moduli were predicted by the RF algorithm. The analysis results illustrated that the shear and bulk moduli of BMGs were negatively correlated with the average atomic volume of the constituent alloys, where the mixing entropy can enhance the shear moduli and the average Pauling electronegativity has an influence on the bulk moduli of BMGs. The study found that BMGs should be composed of the elements with a significant difference in work function and fusion heat, and the boiling temperature of the alloys cannot be high. Besides, the glass-forming ability was enhanced with the presence of sub group elements. Three conditions that favor the formation of BMGs were proposed: high mixing entropy, high average thermal conductivity and appropriate negative mixing enthalpy for approximately −28 kJ/mol.

## 5. Conclusions and outlook

The fourth paradigm of science, data-driven science, has made a lot of success with the arrival of big data and totally changed the philosophy of materials research. Due to the flexibility, accuracy and good generalization ability of ML, a completely different research perspective from the traditional experimental and computational simulation methods has brought. Nevertheless, as a new research method in the field of materials science, ML still faces some problems and constraints.

The main factor limiting the application of ML method in materials science research and application is the lack of effective data sets. The quality of the data obtained under different experimental conditions is different, and it is extremely tough to unify the data obtained under different experimental conditions or match the calculated results with the experimental measured values. In addition, whether the models established by ML method possess practical physical and chemical significance must be further explored. Since the process of ML is a black box for mathematical operations, the hidden association and rules among the calculated data still need to be verified by extensive facts. Therefore, ML technology can only carry out certain exploratory work, which provides new ideas and methods for material science research, but still cannot replace the traditional experimental research.

In the next development process, the worthwhile research directions of ML in materials science may focus on the following aspects.

(1) Creating standardized databases: ML is a data-driven method, which depends on data strongly. Compared with image recognition and other fields which often have millions of data, material science research often leads to over fitting of ML model in the current situation of limited data. Therefore, it is necessary to establish standardized databases and allow researchers to supplement according to the unified standards.

One of the promising methods to increase the amount of material data is to obtain theoretical data through high-throughput computing. High throughput computing is more inclined to computing, that is, to complete the specified work according to the set rules. High-throughput computing lacks the ability of extension and generalization, which is different from ML. Combining the advantages of high-throughput technical and ML is expected to further improve the research of new materials. Moreover, we can develop the method of intelligent reading literature, and obtain an abundant of experimental and theoretical data from publications. Cole et al. [155,156] developed a pipeline for fully integrating data extracted from the scientific literature into ML tools for property prediction and materials discovery. Natural language processing (NLP) and machine-learning techniques were used to reconstruct the phase diagrams of well-known magnetic and superconducting compounds.

(2) Creating interpretable descriptors: it is badly required for data-driven materials science to explore the descriptors with physical interpretability and make the black box models of statistical ML more readable and easier to understand. It can not only help people design new materials, but also make people understand the potential physical laws behind the properties of materials, and provide theoretical basis for further design of materials. Bartel et al. [157,158] presented an one-dimensional tolerance factor, $\tau$, which is physically interpretable and correctly predicts 92 % of compounds as perovskite or nonperovskite for an experimental dataset of 576 ABX3 materials based on SISSO (sure independence screening and sparsifying operator). It is believed that with the application of ML in material research, more and more new descriptors with physical interpretability will be proposed to provide guidance for rational design of materials.
(3) Machine learning for small databases: machine learning methods usually need a lot of data to learn effectively. However, in many fields of material science, experimental data are sparse and difficult to prepare. Faced with this situation, some new machine learning algorithms are needed to improve the learning efficiency and accuracy. In the case of limited data, data augmentation techniques, such as style transfer and meta learning [159,160], can be used to generate similar but different training samples to increase the capacity of the data set. The development of new technologies such as adjusting neural network and imitative learning makes this process possible [161–164], which have great potential in polymer materials and microstructure prediction of materials.

Perhaps shortly, ML can play a pivotal part in the explorations of the underlying physical rules behind the material properties, and even cause a subversive revolution in materials research, not only to help people design new materials.

## 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.

## Acknowledgement

This work was financially supported by the National Natural Science Foundation of China (No. 51627802).

## References

[1] B.O. Gregory, Science 288 (2000) 993–998.
[2] Q. Luo, Y.L. Guo, B. Liu, Y.J. Feng, J.Y. Zhang, Q. Li, K.C. Chou, J. Mater. Sci. Technol. 44 (2020) 171–190.
[3] P.G. Boyd, Y. Lee, B. Smit, Nat. Rev. Mater. 2 (2017) 17037.
[4] X. Wu, F.Y. Kang, W.H. Duan, J. Li, Prog. Nat. Sci.: Mater. Int. 29 (2019) 247–255.
[5] M.X. Zhang, C.L. Wang, A.L. Luo, Z.H. Liu, X.S. Zhang,Appl. Therm. Eng. 166 (2020), 114639.
[6] J.L. Hou, M. Chen, Y.F. Zhou, L. Bian, F.Q. Dong, Y.H. Tang, Y.X. Nie, H.P. Zhang,Appl. Surf. Sci. 512 (2020), 145642.
[7] A. Kunwar, L.L. An, J.H. Liu, S.Y. Shang, P. R◻aback, H.T. Ma, X.G. Song, J. Mater. Sci. Technol. 50 (2020) 115–127.
[8] K. Momeni, Y.Z. Ji, Y.X. Wang, S. Paul, S. Neshani, D. Yilmaz, Y.K. Shin, D. Zhang, J.W. Jiang, H.S. Park, S. Sinnott, A. Duin, V. Crespi, L. Qing, Chen, NPJ Comput. Mater. 6 (2020) 22.
[9] R. Wang, J.S. Wang, T. Dong, G.S. Ouyang,Constr. Build. Mater. 240 (2020), 117935.
[10] Y.Q. Jiang, J. Li, Y.F. Juan, Z.J. Lu, W.L. Jia, J. Alloys. Compd. 775 (2019) 1–14.
[11] P.P. Minh, N.D. Du,Compos. Part B Eng. 175 (2019), 107086.
[12] D. Aleksendric, C.B. David, B. Vasic, Tribol. Int. 43 (2010) 2092–2099.
[13] Q. Zhang, D.P. Chang, X.Y. Zhai, W.C. Lu, Chemometr. Intell. Lab. Syst. 177 (2018) 26–34.
[14] Q. Qian, Y.J. Wang, S. Zhao,Comput. Mater. Sci. 169 (2019), 109086.
[15] J.H. Zhang, X.X. Li, D.S. Xu, R. Yang, Prog. Nat. Sci.: Mater. Int. 29 (2019) 295–304.
[16] W.H. Zhu, Y.L. Xu, J.Y. Ni, G.N. Hu, X.M. Wang, W. Zhang, Mater. Sci. Eng., B. 252 (2020) 114474.
[17] S. Jonathan, R.G.M. Mário, B. Silvana, A.L.M. Miguel, NPJ Comput. Mater. 5 (2019) 83.
[18] C. Stefan, T. Alexandre, E.S. Huziel, P. Igor, T.S. Kristof, K.R. Müller, Sci. Adv. 3 (2016) 1603015–1603022.
[19] W.B. Sun, Y.J. Zheng, K. Yang, Q. Zhang, A.S. Akeel, Z. Wu, Y.Y. Sun, L. Feng, D.Y. Chen, Z.Y. Xiao, S.R. Lu, Y. Li, K. Sun, Sci. Adv. 5 (2019) 4275.
[20] P.B. Albert, D. Sandip, P. Carl oelking, B. Noam, R.K. James, C. Gábor, C. Michele, Sci. Adv. 3 (2017) 1701816–1701824.
[21] M. Manik, C.D. Steven, L.K. Benjamin, G.B. Wolfgang,Electrochim. Acta 323 (2019), 134797.
[22] K. Saliya, B. Panicaud, C. Labergère, Finite Elem. Anal. Des. 164 (2019) 79–97.
[23] A.A. Adetokunbo, E. Koffi, J.B. Douglas, Int. J. Plast. 123 (2019) 101–120.
[24] S.Q. Shi, J. Gao, Y. Liu, Y. Zhao, Q. Wu, W.W. Ju, C.Y. Ouyang, R.J. Xiao,Chin. Phys. B 25 (2016), 018212.
[25] L. Enzo, A. Carlos, Z. Andrew, L. Andrew, L. Victor, S.G. Patrick, Sci. Adv. 4 (2018) 4004–4014.
[26] D. Jia, H.T. Duan, S.P. Zhan, Y.L. Jin, B.X. Cheng, J. Li, Sci. Rep. 9 (2019) 20277–20288.
[27] Y. Liu, B. Guo, X.X. Zou, Y.J. Li, S.Q. Shi, Energy Storage Mater. 31 (2020) 434–450.
[28] P.R. Regonia, C.M. Pelicanoa, R. Tani, A. Ishizumi, H. Yanagi, K. Ikeda,Optik 207 (2020), 164469.
[29] E.J. Nicholas, S.B. Alec, W.A. Lucas, A.W. Michael, V. Vishwanath, J. Juan, Sci. Adv. 5 (2019) 3.
[30] Y. Zhang, X.F. He, Z.Q. Chen, Q. Bai, A.M. Nolan, C.A. Roberts, D. Banerjee, T. Matsunaga, Y.F. Mo, C. Ling, Nat. Commun. 10 (2019) 5260–5267.
[31] A. Luedtke, M. Carone, N. Simon, O. Sofrygin, Sci. Adv. 6 (2019) 2140–2160.
[32] O.A. Oluseun, H.W. David, R.M. Peter,Reliab. Eng. Syst. Safety 195 (2020), 106706.
[33] R. Cynthia, Int. J. Mach. Intell. Sens. Signal Process. 5 (2019) 206–215.
[34] K. Yiannis, G.M. Konstantinos, Neurocomputing 295 (2018) 29–45.
[35] G.E. Hinton, R.R. Salakhutdinov, Science 313 (2006) 504–507.
[36] B. Martin, L. Jens, A. Stuhlsatz, T. Zielke,Graph. Models 108 (2020), 101060.
[37] X.S. Andrew, N. Antje, G. Sanguinettic, J. Neurosci, Methods 228 (2014) 1–14.
[38] E. Gossetta, C. Toher, C. Oses, O. Isayev, F. Legraind, F. Rose, E. Zurek, J. Carrete, N. Mingo, A. Tropsha, S. Curtarolo, Comput. Mater. Sci. (2018) 134–145.
[39] M. Paolanti, L. Romeo, M. Martini, A.O. Mancini, E. Frontoni, P. Zingaretti, Robot. Auton. Syst. 118 (2019) 179–188.
[40] M.M. Organero, R.R. Blaquez, L.S. Fernández, Com. Envir. Urb. Sys. 68 (2018) 1–8.
[41] J.C.T. Amy, V.T. Charles, J.L. Wu, W.C.W. Jack,Int. J. Adv. Sci. Eng. Inf. Technol. 43 (2020), 101027.
[42] M.N. Rastgoo, B. Nakisa, F. Maire, A. Rakotonirainy, V. Chandran,Expert Syst. Appl. 138 (2019), 112793.
[43] P. Hähnel, J. Marecek, J. Monteil, F. O'Donncha,J. Comput. Phys. 408 (2020), 109278.
[44] L. Romeo, J. Loncarski, M. Paolanti, G. Bocchini, A. Mancini, E. Frontoni,Expert Syst. Appl. 140 (2020), 112869.
[45] Y. Liu, T.L. Zhao, W.W. Ju, S.Q. Shi, J. Materiomics 3 (2017) 159–177.

[46] N.P.L. Joseph, L.G. Thomas, T.H. Mike, F. Ian,J. Non-Cryst. Solids 533 (2020), 119852.
[47] A.H. Pablo, A.R. Gonzalo, Eng. Appl. Artif. Intel. 79 (2019) 13–22.
[48] M. Saeed, K. Javed, H.A. Babri, Neurocomputing 119 (2013) 366–374.
[49] J. Balcazar, Y. Dai, O. Watanabe, Int. Conf. Algorithmic. Learn. Theory 2225 (2001) 119–134.
[50] A. Lyhyaoui, M. Martınez, I. Mora, M. Vázquez, J.L. Sancho, A.R. Figueiras-Vidal, S. Member, IEEE. Trans. Neur. Net 10 (1999) 1474–1481.
[51] Y. Liu, J.M. Wu, M. Avdeev, S.Q. Shi,Adv. Theory Simul. 3 (2020), 1900215.
[52] J.C.S. Kadupitiya, F. Sun, G. Fox, V. Jadhao,J. Comput. Sci-Neth. 42 (2020), 101107.
[53] R. Jacobsa, T. Mayeshiba, B. Afflerbach, L. Miles, M. Williams, M. Turner, R. Finkel, D. Morgan,Comput. Mater. Sci. 176 (2020), 109544.
[54] K.K. Steven, G. Jake, M. Ryan, D.S. Taylor,Comput. Mater. Sci. 174 (2020), 109498.
[55] L. Himanen, O.J.J. Marc, V.M. Eiaki, F.C. Filippo, S.R. Yashasvi, Z.G. David, P. Rinke, S.F. Adam,Comput. Phys. Commun. 247 (2020), 106949.
[56] Y.L. Yan, T. Mattisson, P. Moldenhauer, J.A. Edward, T.C. Peter,Chem. Eng. J. 387 (2020), 124072.
[57] Y.K. Zhou, S. Zheng, G.Q. Zhang,Build. Environ. 174 (2020), 106786.
[58] I. Kotenko, I. Saenko, A. Branitskiy, Mater. Today: Proc. 11 (2019) 380–385.
[59] H.T. Zhao, I.E. Collins, W.J. Ren, W.T. Li, C.H. Pang, C.H. Zheng, X. Gao, T. Wu,Appl. Energy 254 (2019), 113651.
[60] W. Chen, H.R. Pourghasemi, A. Kornejady, N. Zhang, Geoderma 305 (2017) 314–327.
[61] F. Zablith, H.O. Ibrahim, Appl. Math. Model. 71 (2019) 569–583.
[62] B. Buisson, D. Lakehal,Nucl. Eng. Des. 354 (2019), 110197.
[63] C. Fan, F. Xiao, C. Yan, C. Liu, Z. Li, J. Wang, Appl. Energy 235 (2019) 1551–1560.
[64] H. Ling, C.X. Qian, W. Kang, C.Y. Liang, H.C. Chen, Constr. Build. Mater. 206 (2019) 355–363.
[65] M. Hasnia, M.S. Aguir, M.Z. Babai, Z. Jemai, Int. J. Prod. Econ. 216 (2019) 145–153.
[66] S.P. Ong, Comput. Mater. Sci. 161 (2019) 143–150.
[67] V.U. Lev, Neurocomputing 331 (2019) 18–32.
[68] G. Battineni, N. Chintalapudi, F. Amenta,Infor. Medi. Unlo. 16 (2019), 100200.
[69] C.L. Julio, Alves, B.H. Claudete, J.P. Ronei, Spectrochim. Acta A. 117 (2014) 389–396.
[70] A. Gavriilidis, J. Velten, S. Tilgner, A. Kummert, J. Franklin. I 355 (2018) 2009–2021.
[71] J.Z. Wang, J.M. Hu, Energy 93 (2015) 41–56.
[72] M. Marjanović, M. Kovačević, B. Bajat, V. Voženílek, Eng. Geol. 123 (2011) 225–234.
[73] G.Q. Wu, R.B. Zheng, Y.J. Tian, D.L. Liu, Neural Netw. 122 (2020) 24–39.
[74] M. Abbaszadeh, A. Hezarkhani, S.S. Mohammadi, Chem. Erde 73 (2013) 545–554.
[75] M. Abdar, V. Makarenkov, Measurement 146 (2019) 557–570.
[76] N.A. Hitam, A.R. Ismail, F. Saeed, Proc. Comput. Sci. 163 (2019) 427–433.
[77] J. Zhang, J.D. Richardson, T.D. Benjamin, Sci. Rep. 10 (2020) 5937.
[78] T.I. Dhamecha, A. Noore, R. Singh, M. Vats, Pattern Recogn. 95 (2019) 173–190.
[79] A. Beghi, L. Cecchinato, C. Corazzol, M. Rampazzo, F. Simmini, G.A. Susto, Ifac Proc. Vol. 47 (2014) 1953–1958.
[80] K.S. Anusha, R. Ramanathan, M. Jayakumar, Eng. Sci. Technol. Int. J. 23 (2020) 483–493.
[81] S.D. Harsh, D. Deb, M.G. Josep, Renewable Sustain. Energy Rev. 108 (2019) 369–379.
[82] W. Fan, F.Q. Si, S.J. Ren, C. Yu, Y.F. Cui, P. Wang,Chemometr. Intell. Lab. Syst. 195 (2019), 103870.
[83] Y. Liu, J.M. Wu, G. Yang, T.L. Zhao, S.Q. Shi, Sci. Bull. (Beijing) 64 (2019) 1195–1203.
[84] A.B. Justicia, J.D. Ferrer, S. Martínez, D. Sánchez, Knowledge Based Syst. 194 (2020) 105532.
[85] M.P. Romero, Y.M. Chang, L.A. Brunton, J. Parry, A. Prosser, P. Upton, E. Reesa, O. Tearne, M. Arnold, K. Stevens, J.A. Drewe,Prev. Vet. Med. 175 (2020), 104860.
[86] H.F. Lu, X. Ma,Chemosphere 249 (2020), 126169.
[87] H.N.R. Wagner, H. Kökeb, S. Dähne, S. Niemann, C. Hühne, R. Khakimova, Compos. Struct. 220 (2019) 45–63.
[88] M. Rezapour, A.M. Molan, K. Ksaibati, Int. J. Transp. Sci. Technol. 9 (2020) 89–99.
[89] Z.T. Liu, M. Wu, W.H. Cao, J.W. Mao, J.P. Xu, G.Z. Tan, Neurocomputing 273 (2018) 271–280.
[90] T. Lajnef, S. Chaibi, P. Ruby, P.E. Aguera, J.B. Eichenlaub, M. Samet, A. Kachouri, K. Jerbi, J. Neurosci. Methods 250 (2015) 94–105.
[91] J. Schmidt, R.G.M. Mário, S. Botti, A.L.M. Miguel, NPJ Comput. Mater. 5 (2019) 83.
[92] D. Warner, E. Vasseur, M.L. Daniel, R. Lacroix,Comput. Electron. Agr. 169 (2020), 105193.
[93] M. Ebrahimi, M.D. Manijeh, E. Ebrahimie, R.P. Kiro,Comput. Biol. Med. 114 (2019), 103456.
[94] Y. Ao, H.Q. Li, L.P. Zhu, S. Ali, Z.G. Yang, J. Pet. Sci. Eng. 174 (2019) 776–789.
[95] N. Schnitzler, P.S. Ross, E. Gloaguen,J. Geochem. Explor. 205 (2019), 106344.
[96] S. Bhattacharya, S. Mishra, J. Pet. Sci. Eng. 170 (2018) 1005–1017.
[97] T. Kawasaki, M. Kidoh, T. Kido, D. Sueta, S. Fujimoto, K.K. Kumamaru, T. Uetani, Y. Tanabe, Orig. Invest. 27 (2020) 1700–1708.
[98] L. Lebanov, L. Tedone, A. Ghiasvand, B. Paull,Talanta 208 (2020), 120471.
[99] I.S. Stafford, M. Kellermann, E. Mossotto, R.M. Beattie, B.D. MacArthur, S. Ennis, Dig. Med. 3 (2020) 30.
[100] M. Xia, W.T. Lu, J. Yang, Y. Ma, W. Yao, Z.C. Zheng, Neurocomputing 160 (2015) 238–249.
[101] W. Li, Y.M. Chen, Y.P. Song,Knowledge Based Syst. 195 (2020), 105606.
[102] A.W. David,J. Pet. Sci. Eng. 184 (2020), 106587.
[103] G.S. Ow, A.K. Vladimir, Sci. Rep. 10 (2017) 36493.
[104] T. Denoeux, O. Kanjanatarakul, S. Sriboonchitta, Int. J. Approx. Reason. 113 (2019) 287–302.
[105] P. Skryjomski, B. Krawczyk, A. Cano, Neurocomputing 354 (2019) 10–19.
[106] Ö.F. Ertugrula, M.E. Tagluk, Appl. Soft Comput. 55 (2017) 480–490.
[107] T. Sathish, S. Rangarajan, A. Muthuram, R.P. Kumar, Mater.Today: Proc. 21 (2020) 108–112.
[108] Y.P. Zhou, M.H. Huang, M. Pecht,J. Clean. Prod. 249 (2020), 119409.
[109] J.F. Barbosa, A.F.O.C. José, R.C.S. Freire, M.P.D. Jesusd,Int. J. Fatigue 135 (2020), 105527.
[110] B. Zhao, T.Y. Yu, W.F. Ding, X.Y. Li, H.H. Su, Prog. Nat. Sci: Mater. Int. 28 (2018) 315–324.
[111] T. Sabiston, K. Inal, P.L. Sullivan,Compos. Sci. Technol. 190 (2020), 108034.
[112] G.P. Xiao, Z.K. Zhu, Tribol. Int. 43 (2010) 218–227.
[113] M.İnal.S. Sahin, Y. Sahin, IFAC 51 (2018) 277–281.
[114] M.N.A.P. Claudio, R. Schirru, J.G. Kelcio, L.C. José, Ann. Nucl. Energy 105 (2017) 219–225.
[115] A.B. Hassan, M.A. Elaziz, H.E. Ammar, A.S. Ezzat, M. Elhadary, D. Wu, Y.S. Liu, Alex. Eng. J. 58 (2019) 1077–1087.
[116] D. Yadav, D. Chhabra, R.K. Garg, A. Ahlawat, A. Phogat 21 (2020) 1583–1591.
[117] B.G. Dipta, B.K. Bijaya, J.W. Wang,J. Nucl. Phys. Mater. Sci. Radiat. Appl. 530 (2020), 151957.
[118] L. Mennel, J. Symonowicz, S. Wachter, K.P. Dmitry, J.M. Aday, T. Mueller, Nature 579 (2020) 62–66.
[119] P.C. Verpoort, P. MacDonald, G.J. Conduit, Comput. Mater. Sci. 147 (2018) 176–185.
[120] Y. Liu, T.L. Zhao, G. Yang, W.W. Ju, S.Q. Shi, Comput. Mater. Sci. 140 (2017) 315–321.
[121] L.P. Lingamdinne, J. Singh, J.S. Choi, Y.Y. Chang, J.K. Yang, R.R. Karri, J.R. Koduru, J. Mol. Liq. 265 (2018) 416–427.
[122] D.H. Bassir, S. Guessasma, L. Boubakar, Compos. Struct. 88 (2009) 262–270.
[123] N.S. Koksal, Comput. Mater. Sci. 47 (2009) 86–92.
[124] M.K. Kazi, F. Eljack, E. Mahdi,Compos. Struct. 254 (2020), 112885.
[125] J. Kasperkiewicz, J. Mater, Proc. IEEE Int. Symp. Signal Proc. Inf. Tech. 106 (2000) 74–79.
[126] Y. Liu, J.M. Wu, Z.C. Wang, X.G. Lu, M. Avdeev, S.Q. Shi, C.Y. Wang, T. Yu, Acta Mater. 195 (2020) 454–467.
[127] Y.F. Juan, J. Li, Y.Q. Jiang, W.L. Jia, Z.J. Lu, Appl. Surf. Sci. 465 (2019) 700–714.
[128] B. Yin, F. Maresca, W.A. Curtin, Acta Mater. 188 (2020) 486–491.
[129] Y.F. Juan, J. Zhang, Y.B. Dai, Q. Dong, Y.F. Han, Acta Metall. Sin. (Engl. Lett.). 33 (2020) 1064–1076.
[130] Y.F. Juan, J. Li, Y.Q. Jiang, W.L. Jia, Z.J. Lu, Mater. Technol. 53 (2019) 751–758.
[131] Y. Zhang, C. Wen, C.X. Wang, S. Antonov, D. Xue, Y. Bai, Y.J. Su, Acta Mater. 185 (2020) 528–539.
[132] D.B. Dai, T. Xu, X. Wei, G.T. Ding, Y. Xu, J.C. Zhang, H.R. Zhang,Comput. Mater. Sci. 175 (2020), 109618.
[133] Z.L. Wang, Y. Adachi, Mater. Sci. Eng. A 744 (2019) 661–670.
[134] X.Y. Huang, H. Wang, W.H. Xue, S. Xiang, H.L. Huang, L. Meng, G. Ma, A. Ullah, G.Z. Zhang,Comput. Mater. Sci. 171 (2020), 109282.
[135] C.C. Wang, C.G. Shen, Q. Cui, C. Zhang, W. Xu,J. Nucl. Mater. 529 (2020), 151823.
[136] K. Liu, Y.Y. Liu, D.C. Lin, A. Pei, Y. Cui, Sci. Adv. 4 (2018) 6.
[137] M.D. Walle, M.Y. Zhang, K. Zeng, Y.J. Li, Y.N. Liu,Appl. Surf. Sci. 497 (2019), 143773.
[138] Y.Z. Zuo, P.J. Ren, M. Zhao, W.M. Su, Y.B. Chen, Y.F. Tang, Y.F. Chen,J. Alloys. Compd. 819 (2020), 152995.
[139] N. Kang, Y.X. Lin, L. Yang, D.P. Lu, J. Xiao, Y. Qi, M. Cai, Nat. Commun. 10 (2019) 4597.
[140] A. Kilic, Ç. Odabaş, R. Yildirim, D. Eroglu,Chem. Eng. J. 390 (2020), 124117.
[141] M.A. Shandiz, R. Gauvin, Comput. Mater. Sci. 17 (2016) 270–278.
[142] H.Y. Liu, J.C. Cheng, H.Z. Dong, J.G. Feng, B.L. Pang, Z.Y. Tian, S. Mad, F. Xia, C. Zhang, L.F. Dong,Comput. Mater. Sci. 177 (2020), 109614.
[143] B. Chen, S.W. Baek, Y. Hou, E. Aydin, M.D. Bastiani, B. Scheffel, A. Proppe, Z. Huang, M.Y. Wei, Y.K. Wang, E.H. Jung, G.A. Thomas, E.V. Kerschaver, Nat. Commun. 11 (2020) 1257.
[144] H.B. Alexander, P.K. Erwin, J.C. Alfonso, L.M.R. Jennifer, Acta Mater. 178 (2019) 163–172.
[145] L. Zhang, B. Wu, Appl. Sur. Sci. 483 (2019) 1052–1057.
[146] H.B. Alexander, P.K. Erwin, J.C. Alfonso, L.M.R. Jennifer, Acta Mater. 178 (2019) 163–172.
[147] T.M. Wu, J. Wang,Nano Energy 66 (2019), 104070.
[148] Ç. Odabas, R. Yıldırım,Sol. Energy Mater. Sol. Cells 205 (2020), 110814.
[149] T.J. Paul, A. Singh, C.L. Kenneth, J. Ilavsky, P.H. Sandip, Sci. Rep. 10 (2020) 2033.
[150] C.H. Douglas, Biomater. Sci. 329 (2010) 1294–1295.
[151] X.X. Yue, J. Brechtl, F.J. Wang, Z.X. Chang, K.L. Peter, C. Fana,Mater. Des. 191 (2020), 108660.
[152] O. Baulin, T. Douillard, D. Fabrègue, M. Perez, J.M. Pelletier, M. Bugnet,Mater. Des. 168 (2019), 107660.

[153] J.M. Jason, A.D. Banadaki, S. Patala, M.V. Paul, Acta Mater. 175 (2019) 35–45.
[154] J. Xiong, S.Q. Shi, T.Y. Zhang,Mater. Des. 187 (2020), 108378.
[155] C.S. Matthew, M.C. Jacqueline, J. Chem. Inf. Model. 56 (2016) 1894–1904.
[156] J.C. Callum, M.C. Jacqueline, NPJ Comput. Mater. 18 (2020) 6.
[157] J.B. Christopher, L.M. Samantha, M.D. Ann, R.R. John, W. Tumas, W.W. Alan, L. Stephan, S. Vladan, B.M. Charles, M.H. Aaron, Nat. Commun. 9 (2018) 4168.
[158] J.B. Christopher, S. Christopher, R.G. Bryan, O. Runhai, B.M. Charles, M.G. Luca, S. Matthias, Sci. Adv. 5 (2019) 0693.
[159] C.W. Lee, J.S. Jang, S.H. Lee, Y.S. Kim, H.J. Jo, Y. Kim, Sci. Rep. 10 (2020) 13694.
[160] J.A. Gabriel, G.M. Rafael, M.M. Saulo, P.F.L. AndréC, F.C.C. Gabriel, B.J. Sylvio, Pattern Recogn. Lett. 128 (2019) 480–487.
[161] G.Y. Zuo, K.X. Chen, J.H. Lu, X.S. Huang, Neurocomputing 388 (2020) 60–69.
[162] Y. Liu, W. Zhang, S.W. Pan, Y.J. Li, Y.T. Chen, Comput. Commun. 150 (2020) 346–356.
[163] W.Z. Qiao, X.J. Bi,Biomed. Signal Process. Control 57 (2020), 101745.
[164] B. Boloukian, F.S. Esfahani, Neural Netw. 121 (2020) 186–207.