Fundamental Research 5 (2025) 624–639

Contents lists available at ScienceDirect

Fundamental Research

journal homepage: http://www.keaipublishing.com/en/journals/fundamental-research/

Review

Intelligent design and synthesis of energy catalytic materials

Linkai Han, Zhonghua Xiang*

State Key Laboratory of Organic-Inorganic Composites, Beijing University of Chemical Technology, Beijing 100029, China

ARTICLE INFO

Article history:
Received 24 May 2023
Received in revised form 15 October 2023
Accepted 19 October 2023
Available online 10 December 2023

Keywords:
Artificial intelligence
Catalytic materials
Machine learning
Retrosynthesis analysis
Intelligent laboratory

ABSTRACT

Efficient energy conversion and storage are crucial for the sustainable development and growth of renewable energy sources. However, the limited varieties of traditional energy catalytic materials cannot match the fast-expansion requirement of raising various clean energy for industrial applications. Thus, accelerating the design and synthesis of high-performance catalysts is necessary for the application of energy equipment. Recently, with artificial intelligence (AI) technology being advanced by leaps and bounds, it is feasible to efficiently and precisely screen materials and optimize synthesis conditions in a huge unknown space. Here, we introduce and review AI techniques used in the development of catalytic materials in detail. We describe the workflow for designing and synthesizing new materials using machine learning (ML) and robotics. We summarize the sources of data collection, the intelligent algorithms commonly used to build ML models, and the laboratory modules for the intelligent synthesis of materials. We provide the illustrations of predicting the properties of catalytic materials with ML assistance in different material types. In addition, we present the potential strategies for finding material synthesis pathways, and advances in robotics to accelerate high-performance catalytic materials synthesis in the review. Finally, the summary, challenges, and potential directions in the development of AI-assisted catalytic materials are presented and discussed.

1. Introduction

With the expansion of energy demand and the increasing emphasis on environmental protection, the industrial system of green energy has been developed [1–2], which puts forward the demand for large-scale application of energy conversion and storage equipment such as fuel cells, electrolyzers, and energy storage batteries [3–6]. However, traditional energy catalytic materials are usually based on precious metals, and the catalytic performance remains to be improved [7–12]. The current development of energy conversion and storage equipment requires higher activity and lower cost of energy catalytic materials to increase the output of clean energy such as hydrogen energy and reduce energy consumption [13–16]. Therefore, it is necessary and challenging to develop new catalytic materials with high activity and low cost. The traditional research and development pattern is usually conducted by trial and error in the lab [17–18], which requires researchers to find potential target materials and synthesis paths based on prior knowledge, and then carry out experiments to optimize the synthesis conditions. The long development cycle and uncertain performance hinder the discovery and application of new catalytic materials. Density functional theory (DFT) calculation has been widely used to verify or predict the theoretical performance of catalytic materials and explore the reaction mechanism [19–21]. However, DFT calculation is still helpless against predicting the unknown chemical space. More importantly, the computational power cost of DFT calculation increases significantly with the increase of real environmental factors considered [19,22]. So, more strategies are needed to speed up the discovery of new materials.

Artificial intelligence (AI) is a new technical science that studies and develops theories, methods, techniques and application systems used to simulate, extend and expand human intelligence [23]. Machine learning (ML) and intelligent robot technology in AI have been widely used in recent years [24–26]. ML builds predictive models based on a rich variety of algorithms to predict the unknown space, which has been applied in catalytic materials [24,27-29], material characterization [30–33] and other fields [34–35]. Intelligent robotics, coupled with automated systems, can replace researchers in the laboratory, continuously synthesizing and screening high-performance catalytic materials [25,36]. In addition, AI-assisted retrosynthesis exhibits the exciting potential in screening the synthetic pathways of organic [37] and drug molecules [38], which provides a valuable strategy for energy catalytic materials. The application of AI in materials science has changed traditional research patterns and provided exciting possibilities for the discovery of new catalytic materials.

Here, we review the applications of AI in intelligent design and synthesis of energy catalytic materials. We summarize and introduce AI techniques applied to materials science. The workflow of ML is described

* Corresponding author.
E-mail address: xiangzh@mail.buct.edu.cn (Z. Xiang).

https://doi.org/10.1016/j.fmre.2023.10.012

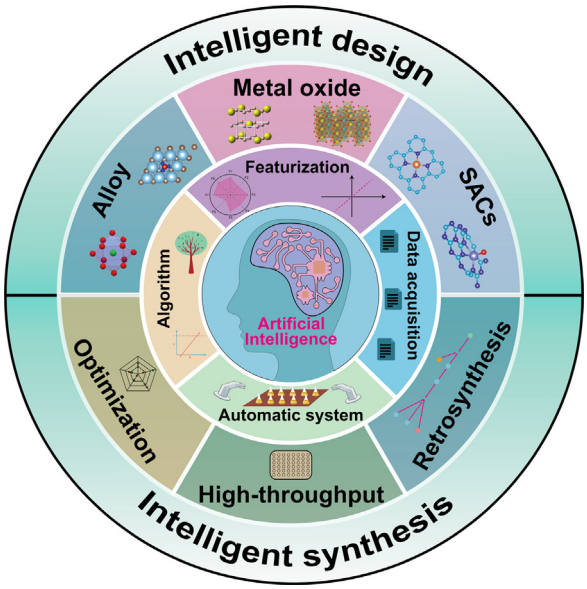

Fig. 1. Overview of the application of artificial intelligence technology in the design and synthesis of energy catalytic materials.

in detail, and the sources of data collection and algorithms used to build ML models are summarized. The strategy of replacing humans by intelligent robots to realize automatic synthesis is also introduced in detail. We divided the commonly used energy catalytic materials into metal oxides, alloys, atomic-level dispersion catalysts and other materials, and introduced how to realize the intelligent design and screening of materials by illustrations. In addition, we also present the application of the computer-aided retrosynthesis analysis method to find synthesis path backward. Remarkably, this review introduces the intelligent weighting of factors in the synthesis of catalytic materials by illustrations. Finally, this review summarizes the current problems that still need to be solved for the AI-assisted discovery of catalytic materials and the future development direction. This review aims to provide guidance for materials scientists working with computer scientists to intelligently design and synthesize catalytic materials.

2. Artificial intelligence technology

AI is a theory and technology that uses digital computers or machines controlled by computers to simulate and expand human intelligence, perceive the environment, acquire knowledge and use knowledge to obtain the best results [39]. The core of AI is to build intelligent artificial systems. AI is a knowledge project that uses machines to perform a series of actions that mimic humans. Typically, chemistry is an experiment-based discipline that requires scientists to do a lot of experiments. However, long and costly trial and error in the vast chemical space have difficulty in meeting the demand for new material discovery in much energy equipment. Therefore, with the assistance of AI technology, such as ML, deep learning, and robotics (Fig. 1), the speed and accuracy of energy catalytic materials screening and synthesis have been greatly improved [26].

2.1. Machine learning models for intelligent design of materials

ML method belongs to AI technology, which studies selecting appropriate algorithms from data, automatically inductive logic or rules, and makes predictions based on the inductive results and new data [28]. Compared with the traditional DFT calculation and manual experimental trial and error in chemistry, ML greatly reduces the cost and cycle of material exploration in the discovery of new materials [29]. The workflow of developing new catalytic materials using ML mainly includes data collection and key feature extraction, algorithm selection and model training, model evaluation and feedback.

2.1.1. Data collection and feature analysis

Accurate collection of valid data and initial analysis is essential for ML to discover new materials. High-quality data are very important for rapid optimization of ML models and accurate prediction of target parameters [31,40]. In addition, feature analysis is an important part of ML for discovering new materials. The feature parameters are usually the properties that have a great influence on the performance of materials [41]. The establishment of the feature parameters data set requires professional scientific skills and intuition as the basis.

As shown in Fig. 2, for chemical materials, the sources of data collection are usually common databases, DFT calculations, and experiments. Databases are generally comprehensive data sets that are collected and collated by academic institutions, including physical properties, chemical properties, electronic structures, and other properties of materials. Although collecting data from a database has the advantage of comprehensiveness and directness, databases usually contain a large amount of invalid data on target performance of materials. Therefore, it is necessary to collect a data set dedicated to target performance orientation. Data collection through DFT calculation and experiment requires a large amount of upfront work investment, but high-quality and specialized data sets can be established quickly. In addition, data can be collected from reported work, which can reduce the upfront cost of building data set. However, due to the disunity of material characterization equipment and material performance testing conditions, the credibility of the collected data needs to be further confirmed [42].

Once the data is collected, it needs to be converted into a text based in mathematical form for ML model [43], which can be named as features. These features usually refer to key properties of the material (such as the type of metal, the number of d-electrons, electronegativity, electron affinity and a series of parameters that can be read directly) [24]. In addition, new features can be formed by combining specific property parameters, which are called descriptors. The descriptor set must provide valid and unique information about the material structure, composition, and physicochemical properties. In addition, the number of descriptors should be as appropriate as possible to avoid overfitting and compromising the ability to predict new materials. It is very important to find the core features for the feature analysis. By using the established feature parameters and performance parameters for preliminary fitting, the core influencing factors of target parameters can be quickly found, and then a more specialized database can be established [32]. Data collection and featurization are the first steps for ML to discover new materials. The suitability of data and feature parameters profoundly affects the quality of ML models.

For energy catalytic materials, there are differences in the feature extraction process of different materials. When researchers extract the features of materials, they usually need to consider the characteristics of the materials. For atomically dispersed materials, the active sites involved in the reaction are usually single or several atoms, so the extraction of features is often focused on atoms that have a significant effect on the reaction. Therefore, the extracted features are usually d orbital information of metal atoms or p orbital information of non-metal atoms, as well as atomic charge information, which can have a direct effect on the reaction species. For metal oxides and alloy materials, the extraction of characteristics usually focuses on metal-metal bonds and metal-oxygen bonds, as well as the electronic structure of coordinated unsaturated metal atoms. In addition, the feature extraction of materials in different application scenarios is also different, which usually requires researchers to use their own experience to analyze possible features to collect key features for the establishment of predictive models.

A complete ML model building process usually requires the identification of target prediction systems, data collection and preprocessing, feature extraction, and the selection of appropriate algorithms. Usually,

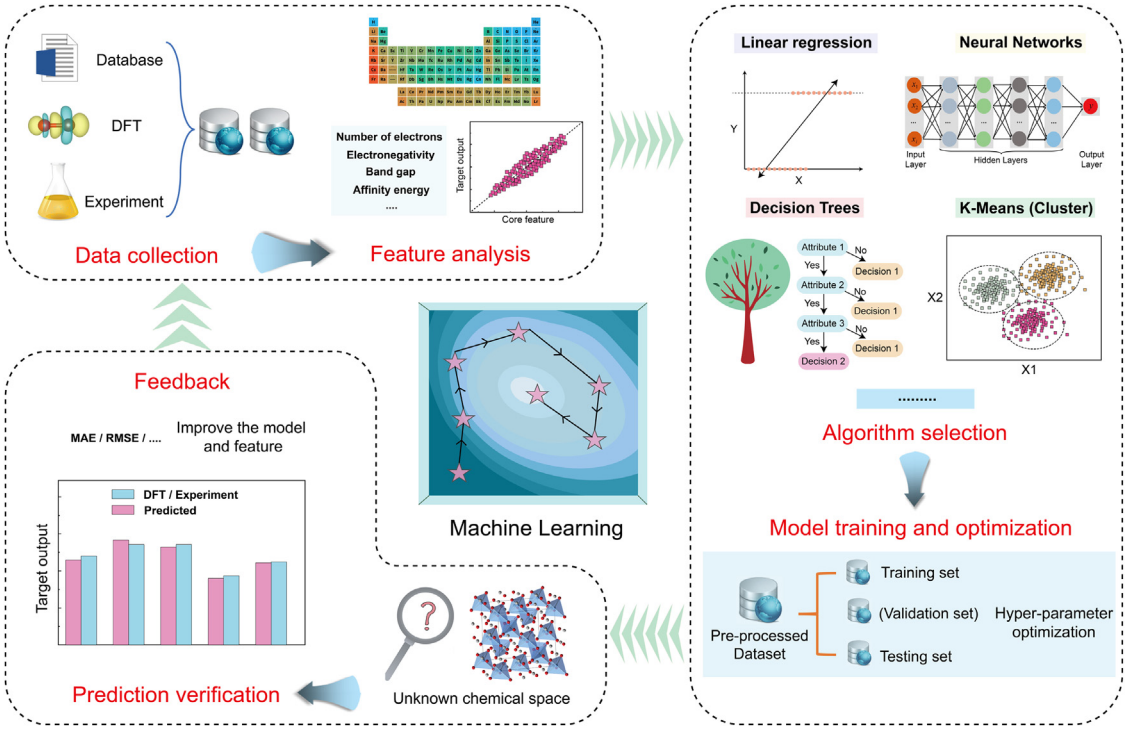

Fig. 2. Workflow diagram of ML discovering new materials. Data collection is first completed and then cleaned to analyze the effective features between the data and the target output. Select the appropriate algorithm to establish the training model. The model is used to predict the unknown chemical space, and the prediction results are obtained and verified. Constantly optimize workflow through feedback to discover new materials accurately and quickly.

after the target prediction system is determined, the selection of algorithms is a process of continuous optimization and comparison. Therefore, for a classical ML model building, after determining the target system, data collection and preprocessing will produce a multi-dimensional database, usually containing valid and invalid data. After feature extraction, the effective data affecting the prediction results can be screened out as much as possible, which establishes the database for ML model. The above process is crucial in building ML models. Typically, after completing these steps, the ML model is considered to have been initially built. On this basis, different algorithms are selected and compared for prediction.

2.1.2. Algorithm selection and model training optimization

After the database is built, ML algorithm can be selected for model training. The choice of the algorithm also has a great influence on the prediction results of the model [44–45]. Once the data set is selected and established, the ML model can be trained using a variety of linear and nonlinear methods. If the data is not labeled, unsupervised learning is used, and if the data is labeled, supervised learning is used [24,28]. At present, in the field of chemical materials, ML algorithms mainly include linear regression, neural network, decision tree, cluster, etc. [28]. Linear regression algorithm is the simplest algorithm for ML modeling, which is suitable for large and small data sets. Linear regression algorithm is often used in feature importance analysis and descriptor selection in the design of energy catalytic materials. In addition, linear regression, as a convenient and efficient algorithm, is also used in the prediction of reaction yield [46–47] and other applications [48–49]. Decision tree algorithms generate a tree-like decision path by using rules formulated by the researcher to quickly determine the type of material. The decision tree model can discover new knowledge and process multidimensional data. They are quick and easy to train, produce high accuracy rates, and are easier to interpret than some ML methods. Decision tree algorithms are often powerless in the face of complex problems. However, as a general algorithm, decision tree algorithm still has a high priority in the field of material design. Because the field of materials design is still highly specialized, researchers can often quickly identify material systems and potential properties [50]. In addition, the decision tree algorithm can also be developed into a random forest algorithm, which can be regarded as a collection of multiple decision tree algorithms, which also has the potential to solve complex problems [51–53]. Neural network (NN) is a biologically inspired algorithm like neurons in the functional brain. A neural network consists of a series of connected input/output units, each of which can produce a weighted parameter. In the learning phase, the weight generated by the neural network can reduce the error between the predicted value and the target until the error is acceptable or not optimized. For large training data sets, neural networks can achieve long training times and the best work. Compared with linear regression, neural network algorithm has obvious advantages in dealing with complex problems and large databases. When the core characteristics of the target material are not obvious, NN algorithms can be used to achieve effective prediction [54]. In addition, NN algorithms also have some potential in other areas of materials science, such as the processing of complex reaction paths in molecular dynamics simulations [55–57] and the efficient design of material synthesis paths in inverse synthesis analysis [54,58-59].

Clustering is one of the unsupervised learning methods, which divides data into different subsets (the elements in the subsets are similar) to determine their similarity. Cluster analysis helps to discover the types of data, and then efficiently analyze the relationship between input and output parameters. Because the conventional material design is usually limited to a label (same type of material, such as metal oxide, alloy, etc.), the clustering algorithm cannot be used well. In the face of data sets that can be divided into multiple labels (i.e., databases with different material systems), clustering algorithms can effectively achieve material prediction [60]. In addition, clustering algorithms in the spectral analysis of materials can greatly simplify the analysis of the obtained data, and only a few representative spectra need to be interpreted [61]. Therefore, compared with the supervised learning algorithm, the clustering algorithm has obvious advantages in the field of material spectral analysis and multi-type material design.

The data set used to build the ML model is usually divided into two or three subsets, including the training set, the test set, and the prediction set, among which the training set and the prediction set are necessary [62]. The selected algorithm is used to train the training set, the test set is used to predict, and the parameters are constantly optimized according to the error of the predicted results.

2.1.3. Algorithm selection and model training optimization

After the ML model is established, the prediction data set of unknown chemical space is established for prediction and feedback, which completes the workflow of discovering new materials. For the discovery of new materials, the predictive performance of unknown data is an important part of evaluating ML models [63]. The obtained prediction data not only provide valuable results for new materials, but also provide correction space for the whole prediction process of the model. Root mean square error (RMSE, Eq. 1) and mean average error (MAE, Eq. 2) are commonly used to evaluate model quality [40]. The specific calculation method is as follows.

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{m}\left(\hat{y}_i - y_i\right)^2}{m}} \tag{1}$$

$$\mathrm{MAE} = \frac{\sum_{i=1}^{m}\left|y_i - \hat{y}_i\right|}{m} \tag{2}$$

where the m is the size of the data set, y_i and $\tilde{y}_i$ is the data classified into measurement parameters and prediction parameters, respectively. Evaluation of the ML model can test the quality of the data collected, the accuracy of the characterization engineering, and the appropriateness of the model algorithm selection. Thus, based on the model evaluation results, the workflow for the discovery of new materials can be constantly fed back and optimized. It's worth noting that there are many factors affecting the predictive ability of ML models. The quality of collected data, extraction of core features, selection of algorithms, optimization parameters, etc., will have an impact on the entire workflow.

2.2. Intelligent synthesis technology of materials

Whether it is predicted by ML or screened by DFT calculation, the potential catalytic materials need to be prepared in the laboratory. However, the predicted target material is an idealized model, usually a simple molecular or atomic structure constructed by theoretical scientists. These structures may have multiple synthesis routes or cannot be synthesized due to thermodynamic instability [8]. In addition, there are still many unknown difficulties in the initial attempt of the synthesis process and the optimization of the synthesis conditions, which usually takes a lot of time for scientists. Therefore, accurate and rapid preparation of target materials is very important and a serious challenge for the discovery and application of new materials. Intelligent synthesis helps researchers to reduce economic and time losses in the synthesis process. Therefore, intelligent synthesis requires the ability to find the optimal solution from a variety of potential synthetic pathways and be able to give guidance to the optimization of synthesis conditions [33,38,64]. So, intelligent synthesis is mainly divided into two parts, the selection of synthetic route and synthesis condition optimization (Fig. 3a).

2.2.1. Synthetic path intelligent screening based on ML

In 1969, Prof. E. J. Corey [65] proposed retrosynthetic analysis method, which provides a powerful thinking tool for chemists to complete the synthesis of complex organic molecules. This route design method takes the scientist about 30 to 60 min to design a possible synthesis route, and the efficiency of the synthesis route depends greatly on the expertise of scientist. In recent years, with the rapid development of AI technology, many researchers try to combine AI with retrosynthesis analysis. With the help of AI, the backward synthesis of materials has become efficient and easy for researchers [34]. Currently, AI-assisted chemosynthesis is usually able to: predict the synthesis route of unreported molecules [66–67]; find new routes for reported molecules [68–69]; generate multiple synthetic routes [70]; find more suitable reaction conditions [30], etc.

A comprehensive planner was developed to implement intelligent filtering of composite routes (Fig. 3b). These programs work by proposing discrete intermediate compounds that are used as the basis for the synthesis of the target materials. In the module, the target material and the potential precursor need to disconnect, by predicting and feeding back a possible precursor structure. Iterating to find all possible precursor chemicals. This process will produce multiple reaction synthesis paths, then find the optimal synthesis route by comparison.

2.2.2. Intelligent synthesis laboratory based on intelligent robot technology

After the synthesis route of the target material is determined, it should be synthesized by experiment. However, the selection of a material synthesis route usually gives multiple possible paths, and there are multiple reaction conditions in one synthesis path. In traditional synthesis, finding the optimal reaction conditions is often costly and time consuming, forcing scientists to repeat the work in the laboratory. Therefore, it is very necessary to design intelligent laboratory for intelligent synthesis. Although the design cost of an intelligent laboratory is high, efficient output and simple maintenance can not only work non-stop but also free scientists for more high-quality work. Design automation laboratory is the basis of intelligent synthesis [36]. An intelligent laboratory usually includes intelligent robot systems, chemical material areas, reaction operation platforms, etc. These modules are integrated into a unified system and can easily start or stop the reaction operation. These automated systems can be called high-throughput operating platforms, enabling single, multi-channel experiments. The experimental operation is done by an intelligent robot, which can work non-stop after the program is completed.

Although the high-throughput operating platform enables multi-channel experiments, the system is still not smart enough to automatically synthesize the best-performing materials. The addition of an intelligent algorithm solves this problem well. Intelligent algorithms (such as Bayesian optimization algorithm) are added to the high-throughput platform to give the robot intelligence: automatically collect performance test data and reaction condition data, and then the optimal response conditions are obtained by continuous comparison. The data collected by the continuous work of the robot is analyzed and a visual operating system is formed, which is convenient for researchers to obtain the synthesis trajectory of the target material and the optimization process of reaction conditions (Fig. 3c). In route selection and material synthesis of target materials, intelligence algorithms bring the whole process to life, so that scientists do not have to do repeated work in the laboratory all the time, and only need to set simple data can be obtained.

New materials can be screened and synthesized rapidly by ML algorithms and robotics. The whole process is as follows: the theoretical structure model of the target material can be screened by ML; intelligent screening of possible reaction paths of target materials; the robot intelligently synthesizes the target material; test the material performance and give feedback.

2.3. Intelligent characterization of materials

Generally, the intelligent screening and synthesis of new materials is aimed at the synthesis of high-performance catalytic materials, while the structural properties and characterization of the synthesized catalytic materials are not too much considered. However, to better reveal the relationship between the structure and properties of the materials, it is necessary to conduct structural characterization of the synthesized material. This can not only verify whether the structure of the material predicted by ML is synthesized, but also verify the result of the retrosynthesis of the material. On this basis, many high-throughput characterization techniques have been developed [71]. These high-throughput characterization devices can often detect the structural properties of multi-

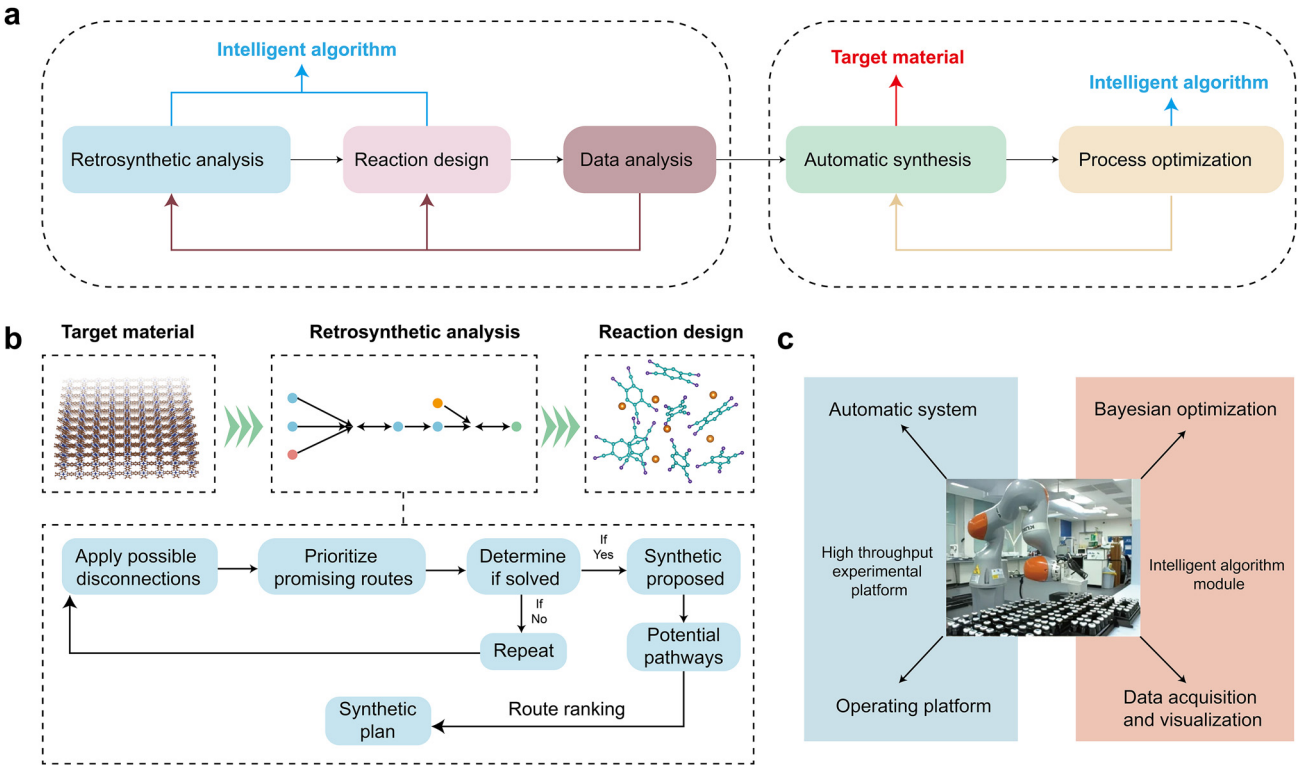

Fig. 3. Workflow diagram of Intelligent synthesis technology. (a) Schematic diagram of intelligent material synthesis route. (b) Schematic of material synthesis route screening realized by retrosynthesis analysis. (c) Schematic of robot synthesis for laboratory automation. Laboratory figure reprinted from ref. [36].

ple materials at once (e.g. X-ray Diffraction, X-ray Photoelectron Spectroscopy, Fourier Transform infrared spectroscopy, etc.). In addition, deep learning technology also has great advantages in the microscopic characterization of materials [33]. Based on deep learning, rapid identification and characterization of the macroscopic structure of materials can be realized, the category and thickness of materials can be identified directly by optical microscope images, and the physical properties of unknown materials can be further predicted.

3. Intelligent design of energy catalytic materials

Catalysis is a complex, multi-dimensional and dynamic process, which usually involves multiphase reaction [72–73]. At present, the design of catalytic materials mainly depends on the experience of researchers, which usually costs a great deal of time and effort, and still cannot meet the demand for efficient catalysts for various catalytic reactions, such as oxygen reduction reaction (ORR), oxygen evolution reaction (OER), and hydrogen evolution reaction (HER). In recent years, with the aid of ML method, the traditional process of discovery and selection of energy catalytic materials has been fundamentally changed [28]. Through data collection and modeling, catalytic materials that meet various needs can be efficiently predicted. Although the accuracy of predictions is affected by multiple factors and requires long periods of optimization of ML models, the process goes from aimless trial and error in the lab to purposeful progress on the computer. Catalytic materials usually have different theoretical performance indexes in different applications, such as theoretical overpotential in electrocatalysis field [74], adsorption energy of intermediates associated with overpotential [51], band gap in photocatalysis field [75], etc. These performance evaluation parameters are usually the simulated energy barrier of a catalytic reaction or the properties of the material itself, which need to be obtained by calculation [22,76]. This section describes the application examples of the ML method in the screening and discovery of catalytic materials, as well as the material application properties of different materials (metal oxides, alloys, atomic dispersion materials).

3.1. Metal oxide material

Metal oxide is a binary compound composed of oxygen and another metallic chemical element. In electrocatalytic reactions, most of the metal oxides in commercial use are gold-based materials [10,77], such as IrO_2 and RuO_2. The high price of noble metal base oxide and the huge improvement of catalytic activity encourage scientists to look for alternatives. The researchers conducted ML method by doping noble metal oxides with non-noble metals or selecting non-noble metal base oxides as the basis. Firstly, the surface structure is constructed by constructing an idealized model and doping one or more non-precious metals. Then, DFT is used to calculate structural properties (such as band gap, state density, charge distribution, etc.) or to simulate the adsorption process of reactant molecules on the surface to obtain adsorption energy data (ΔG_O, ΔG_{OH}, ΔG_H, ΔG_{NH}, etc.) to complete the data collection process. The data is then characterized to find the features most likely to influence the structural properties or catalytic activity of the structure. After the algorithm is selected and the model is built, the structural properties or catalytic activity of materials that have not yet been calculated or synthesized are predicted. In this section, we review the application of AI to different aspects of metal oxide materials, such as material screening and data analysis.

In material screening, the structural framework of metal oxides and the complexity and diversity of doped metals provide a suitable stage for ML. Xie et al. [78] introduced Lewis acid strength, which is strongly related to the ORR kinetic reaction rate of perovskite oxides at high temperature, as a descriptor and verified the validity of eight different regression models (Fig. 4a). Among all regression methods, the MSE values of the ML model-based training set and the test set are 0.009 and 0.013 Ω cm^2, respectively, achieving the best fitting effect among all regression models. The authors found that the intensity of Lewis acid at sites contributed the most to the effect of performance. The results indicate a strong correlation between ionic Lewis acid strength and the inherent ORR activity at high cathode temperatures, which has been verified in the electrochemical characterization (Fig. 4c). Four kinds of

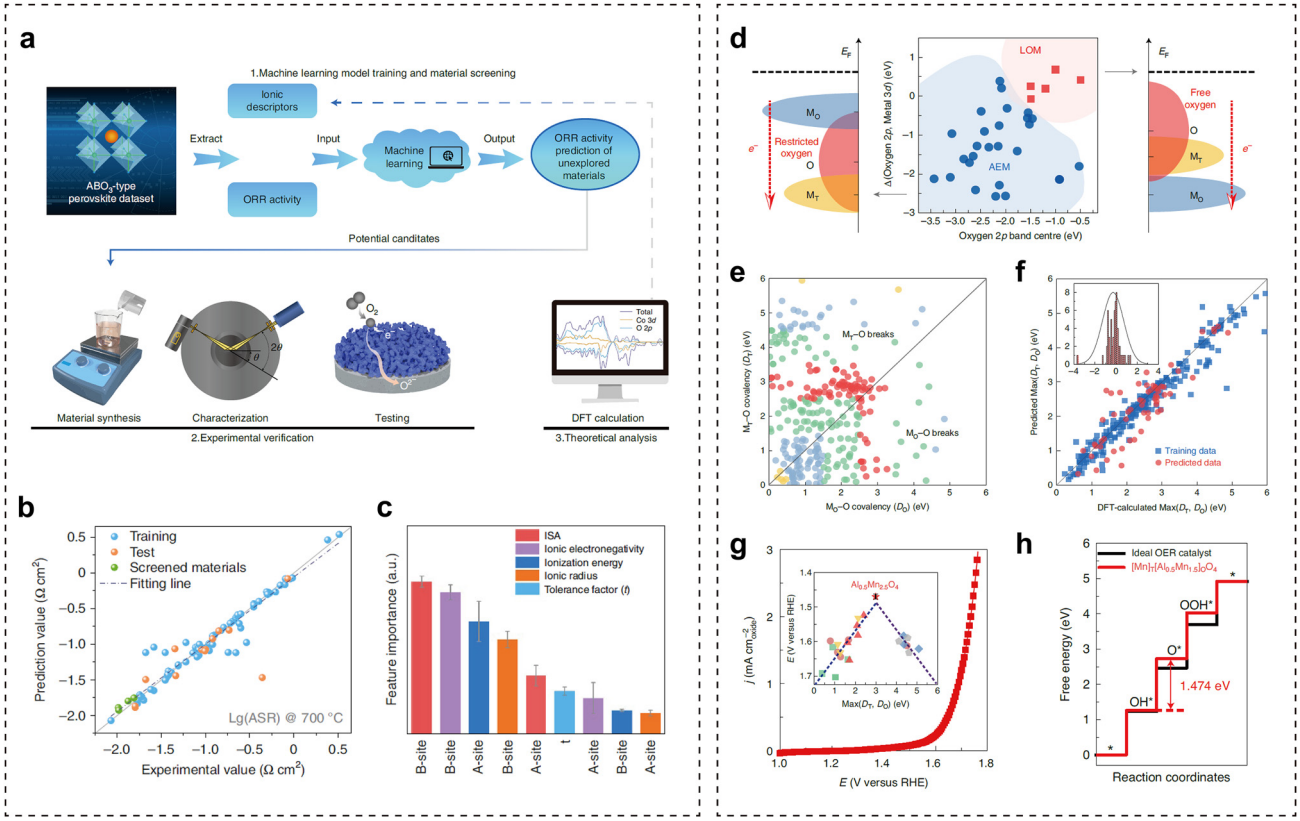

Fig. 4. ML assisted screening of metal oxide materials [51,78]. (a) ML assisted DFT and experiments to achieve efficient screening. (b) Comparison of experimental values and ML predicted values. (c) Feature importance of the ANN model. (d) An illustration of how the oxygen 2p band center, and the relative band centers between oxygen 2p and active metal 3d, co-regulate the reaction mechanism of OER on spinel oxides. (e) The calculated electronic dataset of more than 300 spinel oxides. (f) Comparison of predicted and DFT calculated values. (g) Linear sweep voltammetry curves of the as-synthesized $Al_{0.5}Mn_{2.5}O_4$ oxide. (h) A DFT-calculated OER free energy diagram.

perovskite cathodes were selected from 6871 kinds of perovskite oxides automatically generated and predicted by the machine and synthesized successfully (Fig. 4b). Compared with the high-throughput DFT calculation method, this method can predict the material properties by training the regression model only through molecular formula without establishing molecular model, which breaks the technical barrier of low efficiency of cathode material development. Machine learning is considered a powerful tool for the material design process because it can simplify many factors and direct them to the desired target material. However, for materials science, a deeper understanding of the relationship between material structure and activity is still necessary. Researchers can still use ML to efficiently analyze data to obtain structure-activity relationships.

Recently, Xu et al. have found through computational and experimental studies assisted by ML that the OER activity of spinel oxides is essentially determined by covalent competition between tetrahedral and octahedral position points (Fig. 4d) [51]. The authors found that covalent competition between tetrahedral and octahedral position points determines cationic site exposure, and therefore activity (Fig. 4e). Driven by this finding, the authors calculated data sets for more than 300 spinel oxides through theoretical calculations and used them in ML models to screen for covalent competition of spinel oxides with an average absolute error of 0.05 eV (Fig. 4f). It was predicted that $[Mn]_T$ $[Al_{0.5}Mn_{1.5}]_OO_4$ was a highly active OER catalyst (Fig. 4g-h), and its excellent activity was confirmed by subsequent experimental and DFT results. In this study, based on the use of ML to screen materials, the findings of DFT calculations on the relationship between material structural properties and activity were validated using ML. This work provides a comprehensive understanding of the electrocatalytic OER of spinel oxides and proposes the mechanism of the use of spinel oxides in water oxidation, which can be extended to other applications.

In addition to electrocatalysis, metal oxide materials screened with the aid of ML also have good application prospects in other catalytic fields. Recently, Wang et al. developed a targeted drive method based on ML techniques and DFT calculations to find stable lead-free organic-inorganic hybrid perovskites (HOIPs) [75]. Researchers trained ML models from 212 reported HOIPs bandgap values. Then, six orthogonal lead-free HOIPs with appropriate solar band gap and room temperature thermal stability were successfully selected from 5158 undeveloped potential HOIPs, and two of them showed direct band gap and excellent environmental stability in visible areas. Although this work did not verify the synthesis of the selected materials, the researchers made great efforts to extract the core characteristics of photocatalytic materials, which provides support for the design and synthesis of more photocatalytic materials. Considering the development of ML-screened photocatalytic materials in applications, Sun et al. investigated the effect of doping on the separation and migration capacity (CST) of photogenerated charge in semiconductors and the final water oxidation process using Fe_2O_3 photoelectrodes [52]. The structure database was constructed by doping 17 different concentrations of metal ions, and the basic parameters of 10 elements were screened. The influence of the valence state of the doped element, the bond energy between the doped metal and oxygen, and the ionic radius of the doped element on the CST of the metal oxide was further verified in the CuO photocathode system.

After discussing ML techniques in these studies, we found that researchers focused more on the use of ML for material screening and verification of DFT calculations. As we have emphasized, the development of materials science is inseparable from scientists' understanding of the relationship between the structural properties of materials and activity.

Featurization is an important part of the process of screening materials using ML, it is established between scientists on the material structure and the target activity based on the understanding. Extraction of descriptor and core features affects the accuracy of prediction results. Smit et al. trained a set of ML models to automatically assign oxidation states to metal ions in a metal-organic framework (MOF) [64]. This study focuses on the prediction of oxidation states of metal centers in MOF, especially mixed valence MOF and flexible MOF. In addition, the study provides an application for the assignment of oxidation states to metal centers in MOF on a material cloud. This research provides a more concise and accurate idea for predicting the properties of materials themselves. ML is used to assign values and record data for properties such as valence states of material structures.

3.2. Alloy material

Alloy materials, usually a combination of two or more metal atoms, are widely used in catalytic reactions, gas adsorption, electrical conductivity and other engineering applications. It is very important to determine the exposed surface of the alloy when the catalytic reaction occurs on the surface of the alloy. Alloys consist of a variety of metal combinations and constitute complex structures, which makes it difficult to explain the synergies between metals [52-53,79]. Although quantum chemical calculations have shown some potential in revealing the relationship between the structural composition and properties of alloys, high computational and time costs limit the development of alloy catalytic materials. Fortunately, the ability of machine learning to deal with complex problems has great potential in alloy material systems. A basic ML model can predict material properties, which speeds up the discovery of high-performance alloy materials.

Recently, Sargent et al. developed a high Faraday efficiency copper-aluminum electrocatalyst using theoretical computing and ML [53]. The researchers constructed 244 different Cu-containing intermetallic crystals and identified 12,229 surfaces and 228,969 adsorption sites. The CO adsorption energy at different sites was calculated by DFT to form a data set for ML (Fig. 5a-b). ML prediction results show that Cu-Al is the most promising material with high activity and selectivity for CO_2 reduction (Fig. 5c). Calculations and experiments show that Cu-Al alloy provides multiple sites to achieve the best binding with CO, to effectively reduce carbon dioxide (Fig. 5d). The Faraday efficiency exceeds 80% when the current density is 400 mA/cm^2, which is the highest Faraday efficiency obtained so far (Fig. 5e-f). The data used in this work came from a combination of databases and DFT calculations, providing a common and efficient data collection method. Rigorous and efficient data collection supports the screening and synthesis of high-performance new materials. For CO_2 reduction, in addition to material design and screening, it is also important to clarify the reaction mechanism of CO_2 reduction in the product production mechanism. Also, for electrochemical CO_2 reduction reactions, Wang et al. [80] reported the discovery and optimization of additives using ML in the preparation of copper (Cu) catalysts for electrochemical CO_2 reduction. The copper catalyst was prepared by electrochemical deposition with copper salt as raw material, and different metal salts and organic molecules were added as additives. After three repeated experimental tests, ML analysis, prediction and redesign, it was determined that tin salt was an important additive for obtaining CO and HCOOH, and fatty alcohol was an important additive for promoting C^{2+} formation. The further characterization of catalysts prepared with different additives indicates that fatty alcohols may promote the formation of Cu_2O cube during electrodeposition. In addition, alloy catalysts selected by ML also show exciting potential performance in oxygen reduction reactions. Han et al. [81] combined first-principles DFT with ML technology and based on neural network potential algorithm, systematically calculated, simulated and screened the composition, element distribution and ORR properties of terpolymer PtFeCu nanoparticles. The corresponding PtFeCu nanoparticles with different compositions were synthesized by a simple and rapid ultrasonic method on a commercial scale. Firstly, the structural type of the ternary alloy catalyst is determined, and the location of the metal is discussed in detail. After experimental synthesis and electrochemical evaluation, the authors demonstrated that $PtFe_{high}Cu_{low}$ has the highest site activity (776 μA cm$^{-2}{}_{Pt}$) and mass activity (0.67 m^2 g$^{-1}{}_{Pt}$).

Similarly, while reviewing the discovery of new materials by ML-assisted screening, we also discuss the different applications of ML in alloy materials. Recently, Vivek B. Shenoy et al. [82] trained a graph neural network to predict the adsorption energy of catalyst/adsorption system based on the influence of tensile effect of alloy catalyst on adsorption energy of reactive species (Fig. 5g). The Cu-based binary alloy catalyst in the Open Catalyst Project was used as the data collection source to calculate the adsorption energy under different tensile conditions (Fig. 5h). The established ML model successfully predicted the adsorption energy of 85% of the unknown test data. Taking ammonia synthesis as an example, potential catalysts for Cu-S alloy under tensile strain were selected (Fig. 5i-j). In this work, the influence of the structural tensile properties of the alloy materials on the adsorption of different species was deeply discussed by using ML. Although this study did not involve an experimental component, establishing the relationship between the structure of the alloy material and the theoretical adsorption of different species is crucial for the material involvement. For alloy materials, ML can also optimize the characterization engineering, and speed up the selection and determination of core descriptors. Linic et al. [83] use unsupervised ML principal component analysis to provide a shortcut to developing precise and interpretable electron-structure descriptors of catalytic properties of materials. The authors reconstructed the obtained descriptors of the electronic structure of the material to account for the effects of the electronic structure captured by each principal component descriptor, as well as the local changes in the geometric structure of a location that these effects map to. The method is proved by looking for principal component descriptors for chemisorption on the surface of transition metal alloys and is compared with the results of chemisorption descriptors on physical bases to verify the accuracy of the method.

3.3. Atomically dispersed material

Atomically dispersed catalyst refers to the catalyst in which single atoms are dispersed on the support [84–85]. For atomically dispersed metal catalysts, the inhomogeneity of the support results in different chemical environments for metal atoms, so the catalytic activity of different metal atoms may be different. Support for atomically dispersed catalysts is diverse [86] (e.g. metals, metal oxides, carbon materials, polymer materials, etc.). The structural properties of atomic-level dispersion catalysts determine that the interaction between the metal and the coordination atoms (C, N, O, S, P) on the support is very important for the stability and activity of single metal atoms. In recent years, based on catalysts dispersed by a single type of metal atoms, researchers have proposed diatomic catalysts. DACs have higher metal loading capacity and more complex and flexible active sites [17,87], to obtain better catalytic performance and provide more opportunities for electrocatalysis. It is difficult to find atomic-level dispersion catalysts with higher activity and stability by traditional methods of theoretical calculation and experimental synthesis due to the wide variety of coordination atoms and bimetal combinations of support. Based on this, the addition of ML can be suitable to solve this kind of problem [5,43,88].

In recent years, atomic-scale dispersion catalysts have attracted much attention in the field of electrocatalysis. For some important electrode reactions such as ORR, OER, it is urgent to design and select high performance electrocatalysts. To develop an efficient OER catalyst, Li et al. proposed a single-atom catalyst design scheme based on first-principles computing and topological ML [88]. The team used the DFT method to calculate the OER properties of 15 metal atoms at single-vacancy and double-vacancy defects respectively and obtained the theoretical overpotential by calculating the adsorption energy of oxygen

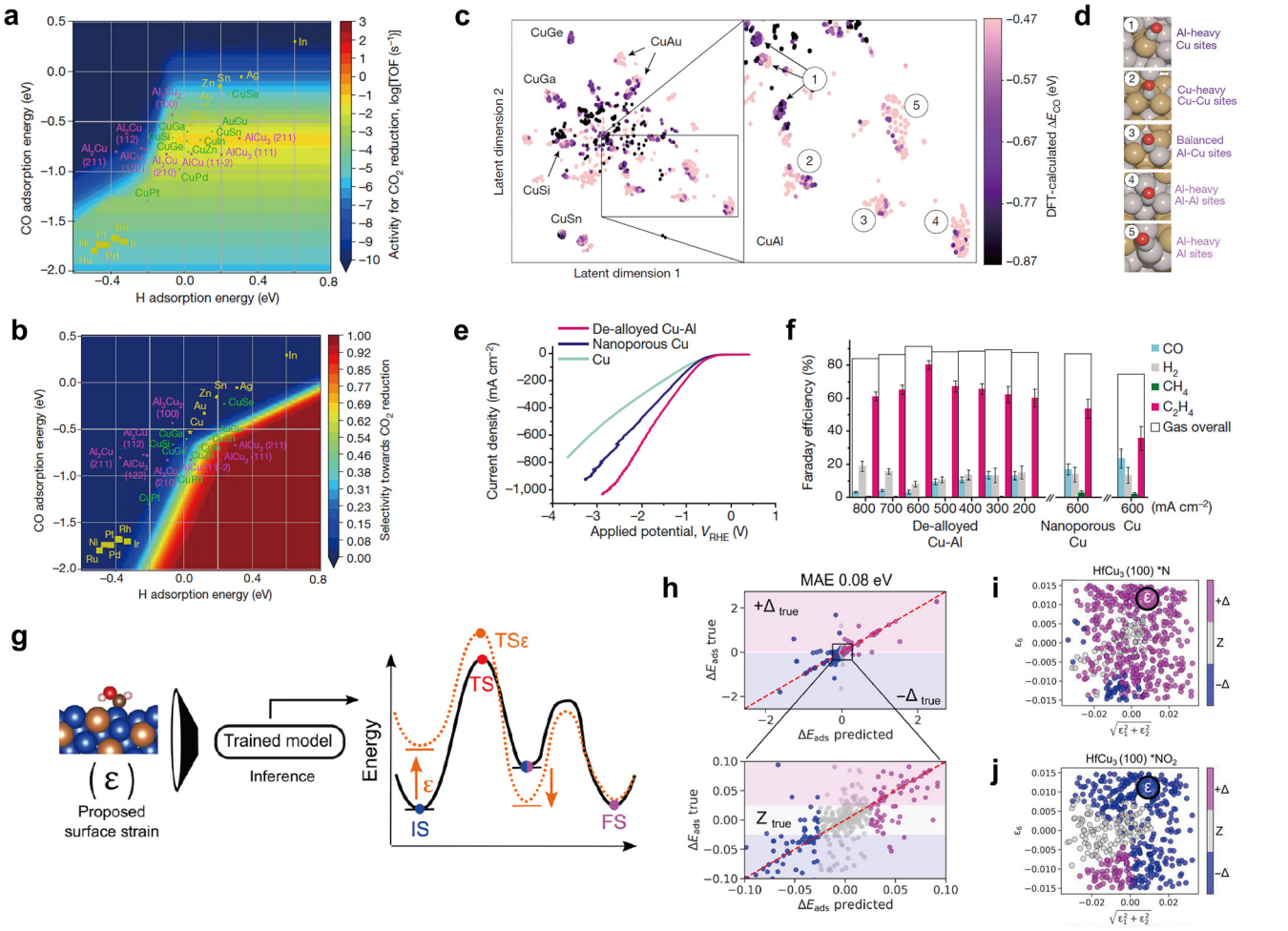

Fig. 5. ML assisted screening of alloy materials [53,82]. (a) Activity volcano plot for CO_2 reduction. (b) Selectivity volcano plot for CO_2 reduction. (c) Theoretical calculated adsorption energy value. (d) Representative coordination sites for each of the clusters labelled in the t-SNE diagram. (e) C_2H_4 production current density versus potential with de-alloyed Cu-Al, nanoporous Cu and evaporated Cu catalysts. (f) Faradaic efficiencies for gaseous products with de-alloyed Cu-Al catalysts. (g) Strain space and surface-adsorbate are combined for ML. (h) Comparison of predicted and true adsorption energy. (i) Surface strain phase diagram resulting from model inference for Cu_8S_4(201):*NH and (j) Cu_4S_2(110):*NH.

species (Fig. 6a). Then, based on the topological learning algorithm, the topological structures around metal atoms were analyzed, the node information and the link information between metal atoms and the substrate were extracted, and a small amount of DFT calculation data were combined to train the prediction model and predict the OER catalytic performance of other transition metals on carbon substrates with different structures (Fig. 6b). To select monatomic catalysts more efficiently, the team proposed a volcanic-type curve description method (Fig. 6c-e), and the screening rate of catalyst is increased by about 130,000 times (Fig. 6f). To design a high-performance dual-function oxygen evolution/reduction reaction (OER/ORR) catalyst, Huang et al. [89] based on C_2N structure, combined with DFT calculation and ML, found that the adsorption energy of a single oxygen atom has a volcanic relationship with the catalytic activity, and combined it with the normalized Fermi abundance, to form a new electronic structure descriptor. In addition, Ding et al. [62] based on a single transition metal AlP system, first used DFT calculations to verify the electrocatalytic performance of bifocal oxygen and found that replacing two P atoms with two N atoms improves catalytic activity. The authors further used the ML method based on gradient lifting regression model to investigate other potential sources of catalytic activity. The results show that the d electron number, atomic radius and atomic charge transfer of transition metal are also the main descriptors related to adsorption behavior. In terms of catalyst development for lithium-sulfur batteries, Li et al. [43] systematically studied the adsorption mode of polysulfide based on ML methods calculated by high-flux DFT calculation, and screened thousands of transition metal monatomic catalysts supported by nitrogen-doped carbon materials (Fig. 6g). Based on a classifier trained by the convolutional neural network of crystal graphs, the authors successfully distinguish the sorbent with S-S bond breaking from other types of sorbent. The regression model trained by ML also has a good ability to predict the adsorption energy (Fig. 6h), with an average absolute error of 0.14 eV, and predicts a series of catalysts with excellent performance (Fig. 6i).

Because the structure of the active center of atomic-level dispersion catalyst is relatively clear, the relationship between material structure and activity can be analyzed by DFT calculation. Therefore, for atomically dispersed catalysts, ML is more used to help researchers deal with complex metal combinations for material screening. In addition, it is precise because the active structure of atomically dispersed materials is relatively clear, and ML can help researchers analyze the dynamic process of materials to obtain a clearer reaction mechanism, which will become a new direction for the development of atomically dispersed materials assisted by ML.

To better compare the AI-assisted screening and design of different types of materials, we summarized data sources, core features, and AI functions (Table 1), etc. From the size of the database and the scale of screening, AI can process a huge material data system and directly find suitable materials. In addition, AI also provides strong support in the

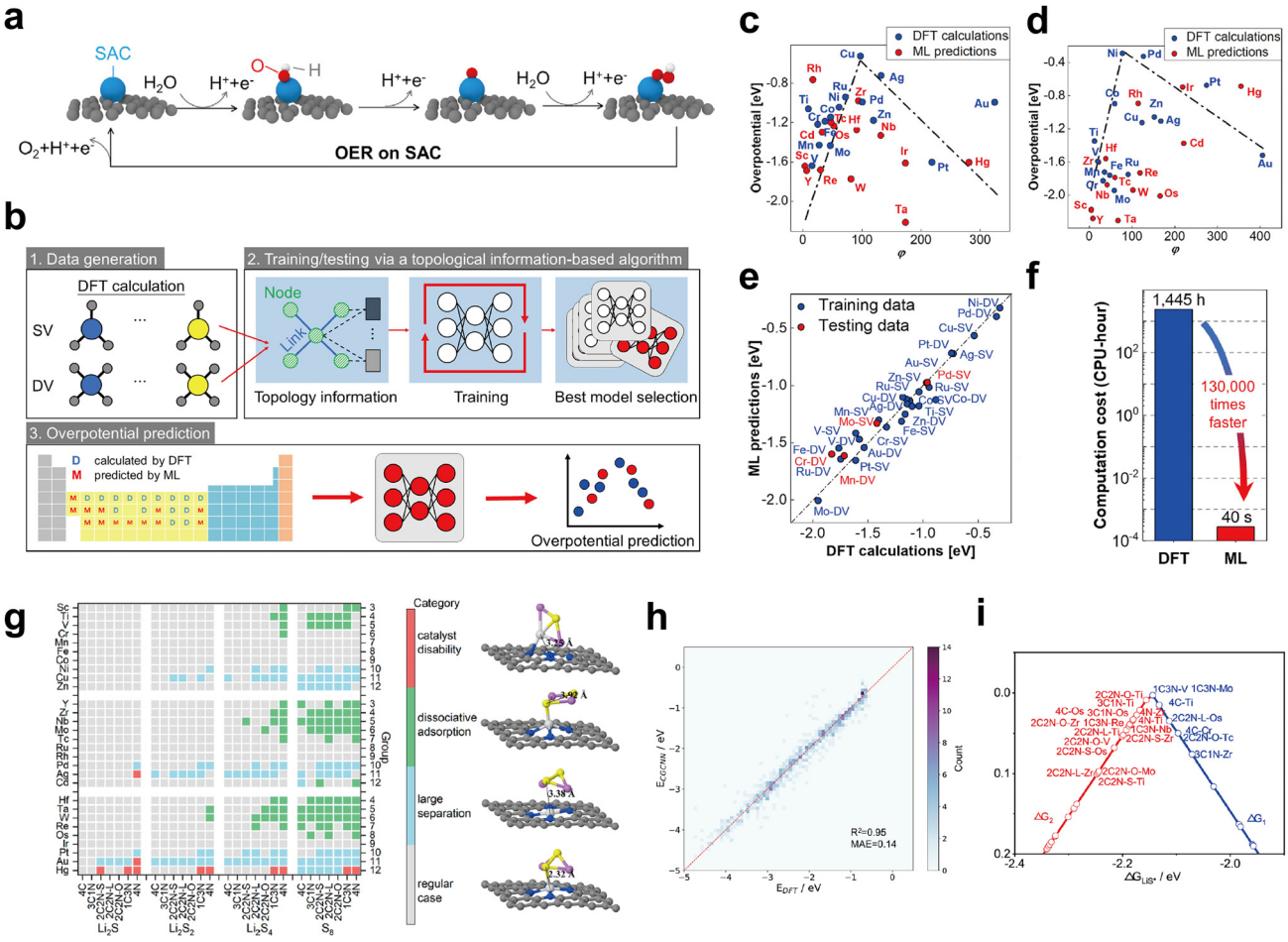

Fig. 6. ML assisted screening of atomically dispersed materials [43,88]. (a) The reaction scheme with the intermediates in the OER process of a SAC on an SV site. (b) Workflow diagram of ML assisted prediction of theoretical overpotential. (c) The volcano relationships between η^{OER} and descriptor φ on SV site. (d) The volcano relationships between η^{OER} and descriptor φ on DV site. (e) Comparison of calculated and predicted DFT values. (f) Comparison of the average computation costs for predicting the OER catalytic activity of a transition metal SAC via pure DFT calculations and ML prediction. (g) Categories of adsorption configurations. (h) Two dimensional histograms of DFT calculated and ML predicted adsorption energy of LiPSs. (i) volcano plots for catalysts with an overpotential lower than 0.1 V.

Table 1
Representative AI-assisted new material discovery.

Material categories	Algorithm	Sd	Data sources	Ss	Core feature	N	Application	Function	Ref
Metal oxide	Neural network	86	Literature	6871	Ionic Lewis acid strength	4	ORR	Screening material	[78]
	Random forest	>300	DFT	>300	p,d band center	1	OER	Screening material; Data analysis	[51]
	gradient boosting regression	212	DFT	5158	Tolerance factor; Total number of ionic charge	6	banggap	Screening material	[75]
	Random forest	~370	Database; experiment	~370	Chemical states; Concentration of dopants	2	Photo-electrochemical Water Splitting	Data analysis	[52]
Alloy	Random forest	4000	Database; DFT	228,969	Metal ratio; Adsorption site	1	CO_2RR	Screening material	[53]
	Neural network	~500	Database; DFT	~3000	Strain tensor	/	Adsorption energy	Data analysis	[82]
	Decision tree	112	experiment	2000	Local PH	/	CO_2RR	Component screening; Data analysis	[80]
	Neural network	44,884	DFT	396,862	Metal ratio	3	ORR	Component screening	[81]
SACs	Adam	30	DFT	30	Number of d electrons	/	OER	Data analysis	[88]
	Neural network	>800	DFT	>1000	Metal ratio	5	L–S Battery	Screening material	[43]
	Clustering	/	DFT	/	Valence state	3	SACs stability	Screening material	[90]

Note: Sd, Ss, N in the table refer to the database size, screening scale, and the number of selected target materials, respectively.

data analysis of the relationship between the structure and properties of materials.

3.4. Other materials

The traditional design methods of metal composite materials usually involve many parameters, and the ideal synthesis conditions can be obtained through repeated trials. In addition, the electronic structures of metal complexes are complex and often have strong correlation effects, and DFT cannot accurately predict the properties of strongly correlated systems, making DFT-based screening less credible. Recently, Liu et al. [79] and others have for the first time successfully applied classification algorithm in ML to guide the synthesis of two-dimensional materials by chemical vapor deposition (CVD) and regression algorithm to guide the hydrothermal synthesis of sulfur-nitrogen-doped blue fluorescent quantum dots with high fluorescence yield. Based on the optimized model, a carbon quantum dot solution with a yield of 55.5% was successfully synthesized. Kulik et al. [70] demonstrated a low-cost method to improve the reliability of DFT high-throughput computational screening. Based on the diagnostic parameters of the strong association effects of 5000 transition-state metal complexes, a ML model was trained to quickly identify the strong association effects of electrons in large-scale (more than 100,000 molecules) high-throughput screening. This method has a promising application in the design and development of functional molecules such as catalysts.

Here, we only give the intelligent design of some inorganic catalytic materials and metal complex materials with the aid of ML. There are still many other types of energy catalytic materials, such as non-metal materials, but most of the current research using ML for material design focuses on metal-based materials. Compared with non-metal materials, metal-based materials have clear active centers in energy catalytic reactions, which is convenient for feature extraction. In addition, the diversity of metal doping types increases the complexity of material design, which can be well dealt with and solved by ML.

4. Intelligent synthesis of materials

Synthetic chemistry techniques are widely used in the field of synthetic materials and synthetic drugs, but the field is slowly becoming labor-intensive due to the variety of chemical reaction conditions. In the process of analysis, a lot of experience accumulation is needed to open the correct synthesis path. However, too many synthesis steps, low yield or high cost make it difficult to realize the commercialization of many analysis results. Synthetic chemistry techniques are widely used in the field of synthetic materials and synthetic drugs, but the field is slowly becoming labor-intensive due to the variety of chemical reaction conditions. In the process of analysis, a lot of experience accumulation is needed to open the correct synthesis path. However, too many synthesis steps, low yield or high cost make it difficult to realize the commercialization of many analysis results. With the development of computer science, AI-assisted chemical synthesis has the potential to solve these problems. At present, the role of AI in the whole field of chemical synthesis application mainly includes two aspects. The first is the design of synthetic route, including assisting researchers to reverse synthesis prediction and route screening. Then, the automated intelligent synthesis in the laboratory requires the intelligent robot platform to solve the problem of long optimization of reaction conditions. In the synthesis of chemical materials such as catalytic materials, the collaboration of multiple AI technologies is usually required. For example, ML, and recommendation system are used to analyze multiple synthesis routes according to the structure of existing chemicals, predict the chemical properties of materials, simulate the spectrum and optimize the algorithm of routes. Then the robot technology is used to realize the automatic intelligent preparation of materials in the high throughput automated synthesis system.

4.1. Synthetic path screening and reaction design

At present, the retrosynthesis analysis based on AI can give a complete synthesis route. Its basic rules are like that of the researchers thinking and selecting the reaction types and reactants involved in the target material. But building a large rule base would be extremely time-consuming and laborious and would have to rely on the expertise of chemists. *Allchemy* [38], a reaction software platform developed by Grzybowski et al., searches tens of thousands of networks for synthetic routes leading to about 300 important drugs and pesticides in the study of chemical waste conversion to drugs and ranks these compounds according to an algorithm (Fig. 7a). Finally, the feasibility of some of the key synthesis methods is verified by experiments. Finally, the author verified the feasibility of some key synthesis methods through experiments (Fig. 7b). Professor Grzybowski cooperated with other researchers to improve the chemical software “Chematica” so that it could design the total synthesis route of complex natural products [68]. The synthesis methods of three natural products designed by this program were successfully verified in the laboratory.

Since the selected target structure is usually an idealized model, to synthesize the target material, it is necessary not only to know the synthesis steps, but also to have a general grasp of the reaction conditions of each step. The design of reaction conditions based on computer aid usually requires some valid data and descriptors to describe the reaction, and then output the reaction yield and selectivity. The workflow of response prediction is often like that of ML for screening new material, in that data is collected to establish core features or descriptors, and then appropriate algorithms are selected for prediction. To achieve prediction of chemical reactions, Glorius et al. [91] developed a ML model for predicting chemical reactivity based on molecular structure (Fig. 7c). The authors propose that all physical parameters can ultimately be traced back to the two-dimensional Lewis structure of the compound and posit that the Lewis structure can be used as an ideal input for ML-based prediction of chemical reactivity. Then they build a structure-based ML model based on molecular fingerprints to predict organic chemical properties and reactivity. More than 2900 small organic molecules were selected from the chemical library. The results showed that the multi-fingerprint signature model could accurately predict the HOMO-LUMO gap, and the average R^2 of more than 10 random cross-validation steps was 0.89 (Fig. 7d). This result confirmed their hypothesis that the molecular properties could ultimately be traced back to the 2D-Lewis structure. Finally, a model application study was conducted on the reported data base. The results show that the proposed multi-fingerprint feature model has good correlation in reaction prediction, which is significantly higher than that of the single thermal coding model (R^2: 0.76 and 0.59, respectively), which proves the versatility of the multi-fingerprint feature method and its potential in learning chemical structure and predicting chemical reaction activity (Fig. 7e-f).

Although there are few reports about finding the synthesis path of catalytic materials by retrosynthesis method, it provides a new idea for finding the synthesis path of catalytic materials more quickly. The logic and regularity of synthesis of most catalytic materials are weak, but some general synthesis rules can still be used. In addition, the synthesis of catalytic materials is currently mainly conducted by scientists who summarize the literature and then design experiments. This process can be replaced by a computer, with basic logic and rule input, allowing AI to help scientific reading literature experimental synthesis methods, and finally give potential synthesis routes.

4.2. Automated robot synthesis and reaction condition optimization

At present, the objects of chemical research are complicated and high-dimensional, but the traditional research methods are mainly “exhaustive” and “trial and error”. The discovery and synthesis of new materials need to face huge unknown chemical space. For the synthesis

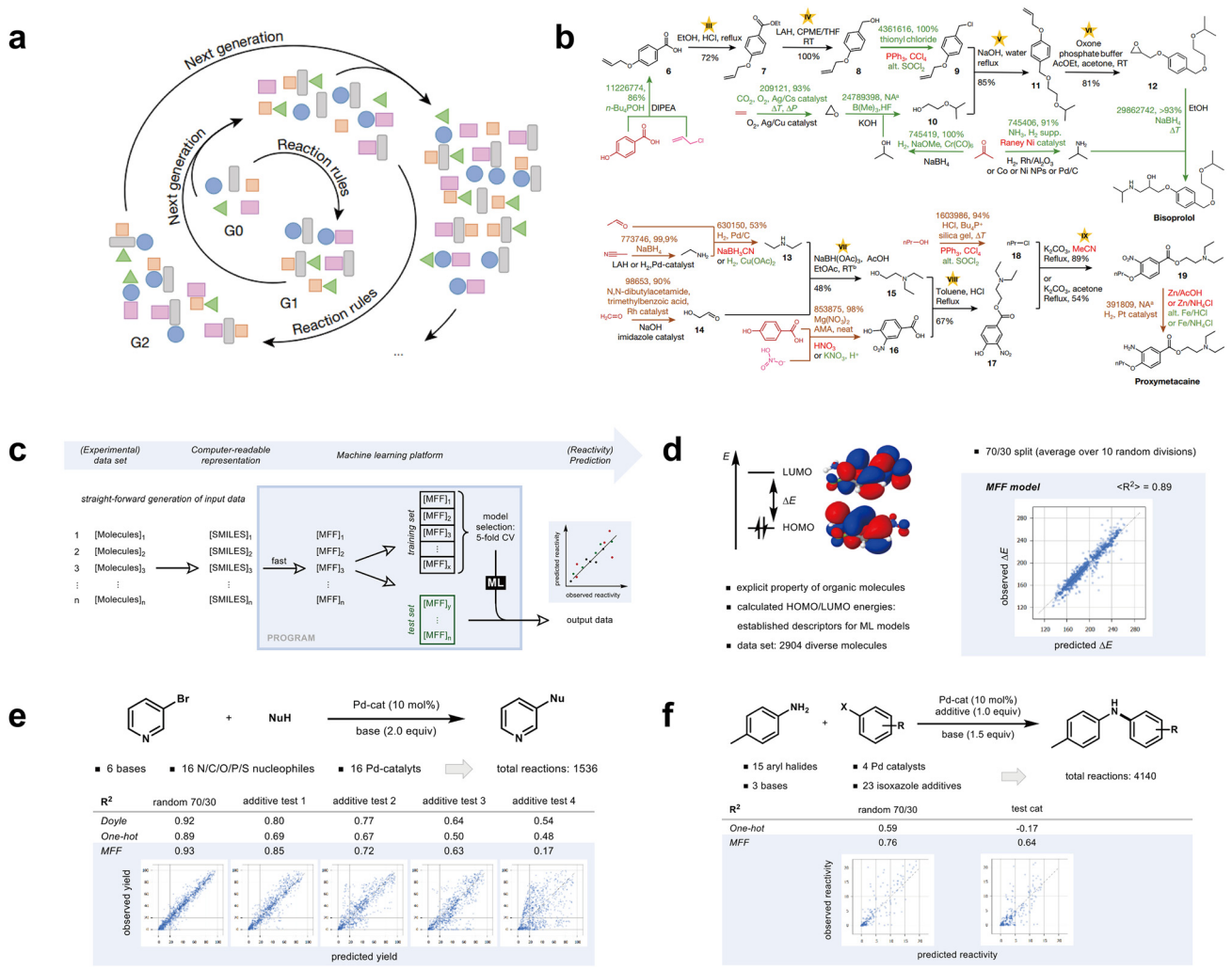

Fig. 7. Reverse screening and reaction prediction of material synthesis pathways [38,91]. (a) Scheme of the iterative, forward-synthesis algorithm. (b) Allchemy-designed, waste-to-drug syntheses of bisoprolol and proxymetacaine. (c) Standard workflow for the prediction model based on multiple fingerprint features. (d) Calculated HOMO–LUMO gaps as an explicit molecular property and performance evaluation of the multiple fingerprint features model. (e) and (f) Comparison of the original model, a one-hot encoded model as statistical probe and the multiple fingerprint features model (R^2 given as correlation measure).

of materials, it is still a challenge to optimize the synthesis conditions and reveal the relationship between process and properties, although the synthesis path can be found by using computer-aided retrosynthesis analysis to find the most likely path. Some exciting progress has been made in high-throughput synthesis, which saves time by running multiple experiments simultaneously, but still requires difficult analysis of experimental results. To synthesize new materials more efficiently, Cooper et al. [36] have successfully developed an AI robot chemist with humanoid features that can work by itself in a standard laboratory, using various laboratory instruments just like humans (Fig. 8a). In the first test, the robot independently completed 668 experiments in 8 days (Fig. 8d) by building an adaptive automated experimental environment (Fig. 8b-c) and developing a new chemical catalyst. In addition, the author realizes the optimization of synthetic variables through the algorithm. The intelligent synthesis system can analyze variables of 10 dimensions based on the results of the previous experiment and can determine the best experiment to conduct next from more than 100 million candidate chemical experiments in the laboratory (Fig. 8e). Cronin et al. [92] have proposed an autonomous chemical synthesis robot to explore, discover, and optimize nanostructures driven by real-time spectral feedback, theory, and ML algorithms. In an open exploration of multistep synthesis of gold nanoparticles (AuNPs) by visible characterization, five classes of nanoparticles were identified in about 1000 experiments in the chemical space. In addition, the authors have achieved yields of up to 95% by combining experimental and spectral simulations to optimize nanostructures with the desired optical properties.

Although many studies have reported intelligent synthesis platforms for new materials based on their respective research systems, these intelligent platforms can only be used for some special material classes. It is still necessary for the researcher to face a large amount of literature for the initial collection of material information. Luo and Jiang et al. [25] have developed a data-intelligent- driven whole-process robotic chemist by developing and integrating mobile robots, chemical workstations, intelligent operating systems, and scientific databases (Fig. 8f). The intelligent platform designed by the author can automatically search and read literature, analyze literature data, propose scientific hypotheses and formulate experimental schemes. The authors used a Bayesian optimization program to find the optimal catalyst from 550,000 possible metal ratios (Fig. 8g). Compared with materials discovered by traditional laboratory trial-and-error methods, the intelligently synthesized materials show excellent properties (Fig. 8h). Compared with other intelligent synthesis platforms, the author adds literature reading and analysis modules, making the whole workflow more intelligent.

At present, AI technology, especially robot technology, is developing rapidly. How to build an intelligent platform suitable for different kinds

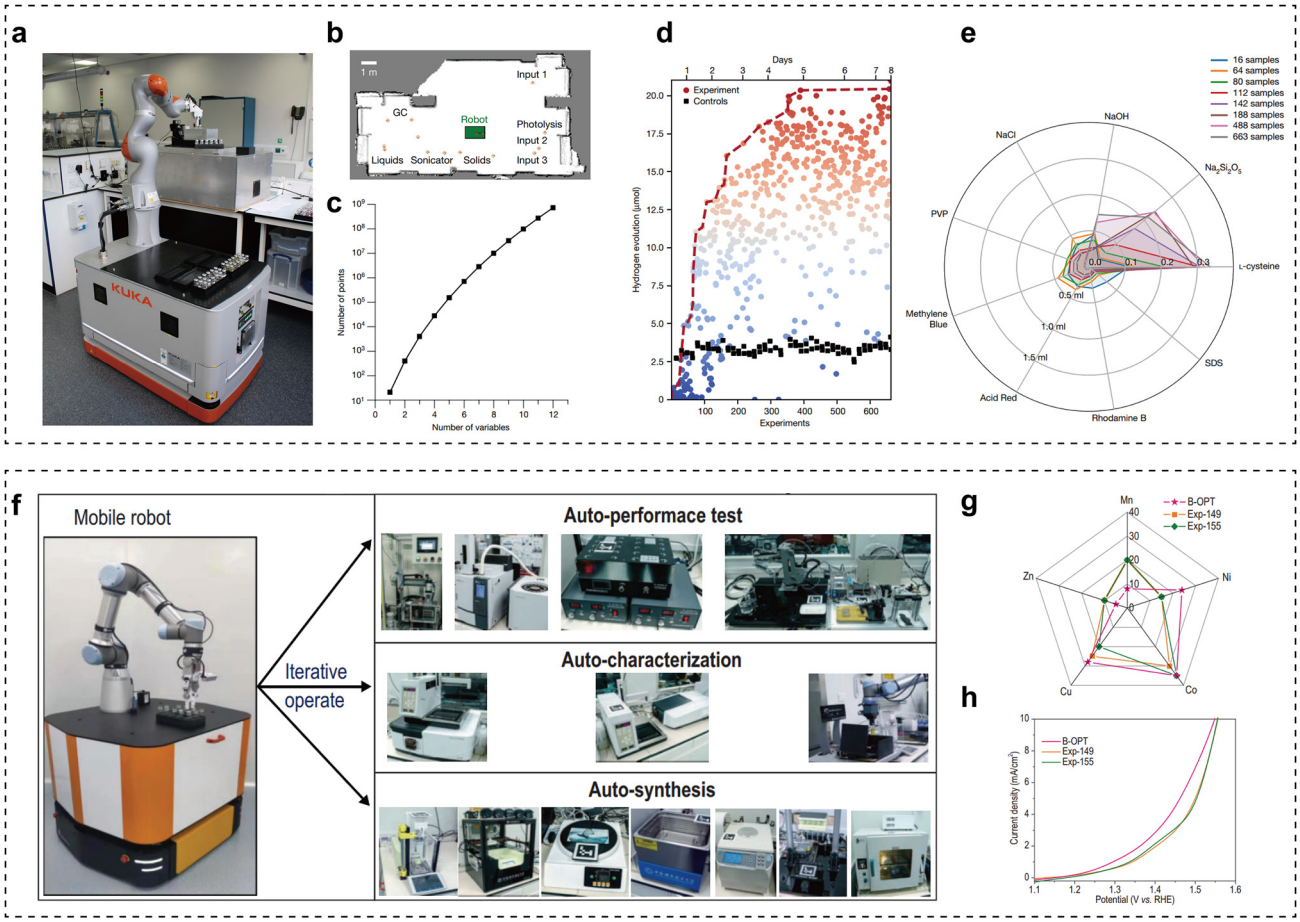

Fig. 8. High throughput synthesis platform and intelligent robot technology [25,36]. (a) Diagram of an intelligent robot at work. (b) Map of the laboratory generated by laser scanning showing positions of the eight stations. (c) Plot showing the size of the simplex, or the search space, created with a discretization of 19 concentrations for each liquid and 21 concentration levels for the solid catalyst. (d) Plot showing hydrogen evolution achieved per experiment in an autonomous search that extended over 8 days. (e) Radar plot showing the evolution of the average sampling of the search space in milliliters. (f) The workflow of the AI-Chemist and the functions of each module. (g) Kiviat diagram of composition ratios and (h) polarization curves of the optimal sample suggested by the Bayesian model and the best samples by Try–Error experiments.

of chemical materials around intelligent robots is still a huge challenge. How to make the intelligent synthesis platform more convenient and simpler still needs the cooperation of computer and chemistry experts.

5. Intelligent characterization and reaction kinetics of materials

For materials science, the design and synthesis of materials are important, while the characterization of the structural properties of synthesized materials and the observation of the kinetic processes of materials participating in catalytic reactions are also worthy of attention. Unfortunately, these studies are often limited by the development of material characterization devices and reaction observation techniques. Although we discuss the application of high-throughput characterization devices in the intelligent synthesis of materials in Section 4, we emphasize the ability of high-throughput characterization devices to handle multiple batches of samples and the importance of high-throughput characterization devices in building intelligent laboratories. However, these characterization devices are often unable to give more refined structures and microscopic reaction processes. After selecting the target materials through intelligent design and synthesis, it is still important and challenging to characterize the reaction process and fine structure of the target materials [55,93-95]. Recently, the development of AI technology has provided more potential strategies and possibilities for this research direction [57,61,96]. In this section, we review the efforts made by researchers using ML in AI technology to solve intelligent In-situ characterization of materials and simulation of reaction dynamics.

5.1. In-situ characterization of material based on ML

Recently, many researchers have begun to apply ML technology in the field of intelligent characterization of materials, to help researchers better use existing characterization equipment to observe the fine structure of materials and establish the relationship between material structure and activity. Zhao et al. [94] conducted data mining based on 1300 reported articles on gold nanorod synthesis and screened 48 solvents and 61 surfactants that could potentially adjust the size and morphology of double perovskites (Fig. 9a,c). The researchers first synthesized high-throughput materials, using in situ optical characterization to obtain RGB values and ectopic characterization (transmission electron microscopy, scanning electron microscopy) to obtain large and small sample data. Finally, through machine learning, the researchers obtained the relationship model between the structure guide agent and the absorption spectrum, and the relationship between the absorption spectrum and the nanocrystalline size (Fig. 9b). It is noteworthy that based on the large sample color information (RGB) of two materials, the relationship between color information and nanocrystalline size is constructed in this study, which can be used as another index for rapid identification of nanocrystalline size (Fig. 9d). In this study, ML is combined with in-situ characterization of materials and based on the in-

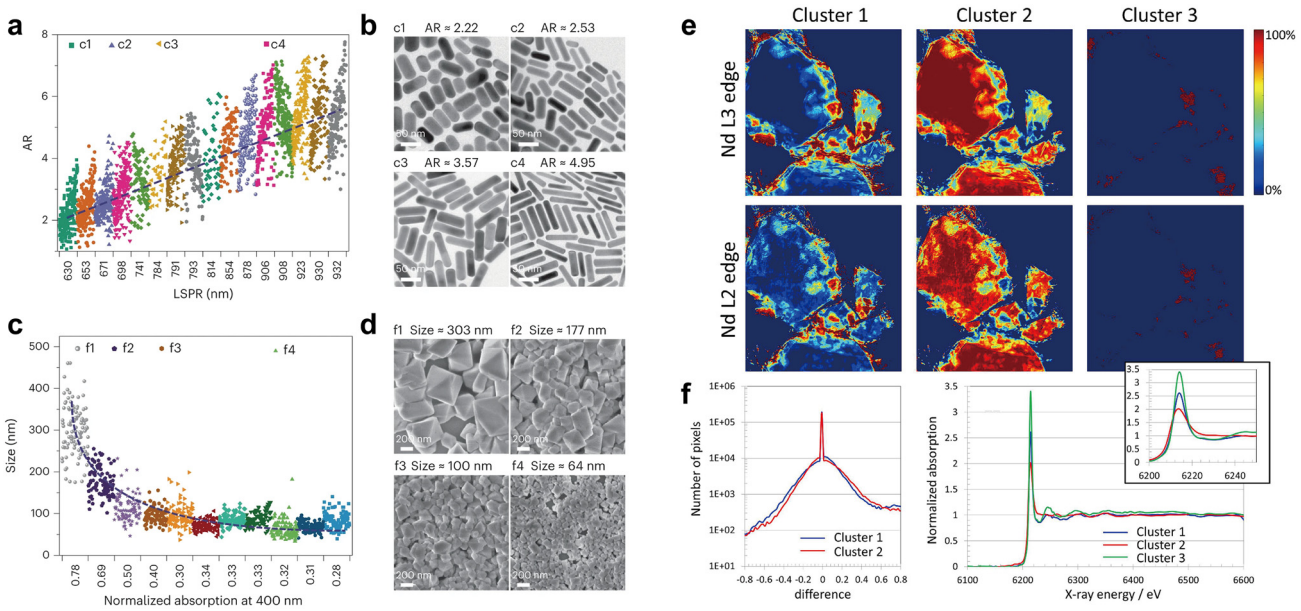

Fig. 9. ML-assisted in situ characterization of materials [93–94]. (a) Identified linear relationship between the LSPR and AR (the results for c1–c4 are measured from b). (b) TEM images for AR validation. (c) Identified relationship between the normalized absorption at 400 nm and morphological size (the results for f1–f4 are measured from d). (d) SEM images for size validation. (e) Cluster maps generated by DBSCAN based on the Nd L_3 (Top)and L_2 (down) edge data. (f) The workflow of the AI-Chemist and the functions of each module. (g,h) the averaged spectra for the three identified clusters with the near edge region magnified in the corresponding inset.

situ characterization results of materials, a model is built using machine learning to predict the crystal structure of new materials based on the observed data of material characterization (RGB values). In addition, ML can combine in situ characterization data with other characteristics to build a database for prediction of high-performance catalytic materials. Yu et al. [93], by integrating optical simulation, data modeling, automated synthesis, in-situ characterization, machine learning and other technical methods, to achieve the optimal synthesis and high-throughput screening of silver nanocrystals under the action of multiple factors, break through the limitations of traditional colloidal optical materials in controlled synthesis and color application, and accelerate the application of target materials in visual detection of mercury ions. In this study, high-throughput wet chemical synthesis and in-situ optical characterization (spectrum and color) were conducted for key parameters in automated synthesis (stabilizer, shape control agent, and reducing agent), and large sample data were effectively classified and screened. Combined with machine learning algorithms, a complex model between multi-factor synthesis parameters and material shape distribution was established. Based on controlled synthesis (effective sample number >1200), the mathematical model between material color (RGB value) and grain size is further constructed, which can be used as a visual index and empirical database for rapid identification of silver nanocrystalline size.

In the above two studies, the researchers correlated the RGB values of the observed results with the structure, features and activity of the materials for the in-situ image characterization results. These two studies provide a good guide for the combination of image characterization and ML, and the criticality of RGB values is also emphasized. In addition, ML also provides help in accelerating the spectral analysis of materials. Compared with in situ image characterization (SEM,TEM), spectral analysis generates a large amount of text data, which can be well handled by ML.

In the past decade, X-ray absorption spectroscopy has been widely used in the structural characterization of energy catalytic materials to explain the evolution of material structure and active species in the reaction. However, XAS characterization of materials usually requires complex preparation and time consumed to obtain a stable signal [96], so some researchers have applied ML to the spectral analysis of XAS to help researchers quickly obtain stable and reliable material structure information. Zheng et al. [95] constructed a database of about 800,000 K-edge XANES spectra and developed an ML tool for automatically identifying XANES data. This tool evaluates the difference between two spectra by defining a distance or similarity function. Although it is still difficult for this tool to deal with outliers, experimental noise, and systematic errors such as data errors, it still offers great potential for XAS analysis of materials. In this algorithm, the ML model learns the spectral data similarity of different known material combinations and preprocesses the data to achieve a high degree of agreement between the material theory and the experimental test results. Although this study did not apply the tool to the prediction of unknown materials, the tools it provided greatly stimulated the enthusiasm of the researchers for the intelligence of XAS analysis. In XAS analysis of unknown materials, Liu et al. [61] adopted clustering algorithm to simplify the signal acquisition process of XAS and carried out intelligent analysis on different test areas of unified materials to obtain the overall XAS data level of materials. In addition, the process of analyzing the resulting data using this method is greatly simplified, and through data monitoring, parts with significantly different structural compositions in different regions of the tested material are immediately visualized, helping researchers to understand the overall structure and fine structure of the synthesized material. As shown in Fig. 9e, the spectra collected by the researchers can be divided into three parts, and the spectral data of materials in the same region are similar (Fig. 9f). Obviously, the XAS data of different clusters are different, which indicates that the structure of the material itself will be different due to environmental differences, which is caused by the different atomic bond distances in different clusters.

After discussion, we find that the current research using ML-assisted techniques in material spectral analysis mostly focuses on data processing and integration. Based on the spectral data collected from experiments or databases, the researchers then use ML to analyze the data to help the researchers quickly gain a preliminary understanding of the spectral and structural information of the material. Although many efforts have been made in this area, researchers still face challenges in collecting data and implementing intelligent spectral analysis using ML due to inconsistent data standards, different testing conditions, and cumbersome experimental testing processes.

5.2. Simulation of material reaction kinetics based on ML

The dynamic simulation of materials is very important in the synthesis of materials and the application of materials. Revealing the dynamic mechanism of materials can help researchers to better design useful new materials. Typically, simulating the dynamics of materials through first principles usually requires high computational power and time costs, and the dynamics of materials are often complex. Therefore, it is necessary to use ML to accelerate the dynamic simulation process of materials.

A. E. Roitberg et al. [57] trained on quantum computing through deep neural networks (NN) to obtain an accurate and transferable organic molecule. By proposing the ANI method, the researchers constructed the ANI-1 model based on the constructed database. This model predicts the total energy of an organic molecule containing four types of atoms: hydrogen, carbon, nitrogen and oxygen. Through a series of case studies, the authors demonstrate that ANI-1 has higher chemical accuracy on larger molecular systems than DFT calculations. Although this study did not investigate the reaction dynamics of the exact reaction system or material system, the energy prediction model proposed by this study provides a more efficient algorithm for molecular dynamics simulation, which helps researchers to quickly and efficiently predict the calculated energy required for the kinetic simulation process. Gong et al. [55], to solve the problem that the active site of oxide-derived copper catalyst to generate multiple carbon (C_{2+}) products in the process of CO_2 electroreduction is still unclear, used neural networks (NN) to assist DFT calculation and simulation to describe the real catalyst surface model. The molecular dynamics of the material under the reaction potential were simulated by ML model, and more than 150 surface sites were detected, which proved that the planar-square and convex-square sites were the reaction sites. In this study, ML was used to simplify the complex process of CO_2 reduction reaction and provide a clear explanation of the active site of the reaction. This study shows that ML has great advantages in the simulation of reaction dynamics. Compared with traditional DFT calculation, complex microscopic reaction processes can also be clearly demonstrated and explained with the aid of AI.

Although the current research in materials science in the field of energy catalysis focuses on material performance, and researchers often ignore the microscopic reaction process and in-situ fine structure characterization of materials, it is still important to efficiently simulate the dynamic process of materials and quickly predict the in-situ characterization data. Based on AI, researchers are no longer limited to the current development of material characterization equipment and high computing power costs but can still achieve accurate prediction of material characterization data and achieve accurate interpretation of dynamic simulations.

6. Conclusion and perspective

AI has shown powerful functions in many fields. With the aid of AI technology, the discovery and synthesis of new catalytic materials have undergone qualitative changes compared with traditional methods. Based on AI technology, researchers reduce a lot of repetitive work and free up time for more advanced thinking. However, there are still opportunities and challenges for AI-assisted discovery and synthesis of new materials, thus we summarize the following possible directions for intelligent design and synthetic materials.

6.1. Effective data collection and database establishment

Although data collection sources of catalytic materials are abundant at present, such as database, DFT calculation, experimental collection by researchers, literature reference, etc., the efficiency of data collection and data quality still cannot meet the requirements at the same time. The data collected by the researchers through DFT calculation or experiment have strong reliability for the predicted results, but it requires high cost and long time, which affects the speed of new material discovery. On the other hand, the credibility of data collected through database or literature is reduced due to inconsistent standards, different reaction conditions, and differences in test equipment. Data collection is therefore an important process for researchers who use ML to discover new materials. How to balance data credibility and efficiency needs to be carefully considered by researchers. For researchers, preliminary data analysis is often effective for collecting data from databases. When a large amount of potentially available data has been obtained from a database, a preliminary analysis of the data to weed out invalid data often increases data reliability and improves prediction accuracy. It is recommended that researchers obtain data from DFT calculations when making predictions about small systems or systems that have already been reported. These systems often already have highly correlated features identified, so the importance of data reliability becomes critical for experiments and applications. Although data from databases is often plagued by data quality issues, building databases is still a challenging but rewarding path for the design of AI-assisted new materials. Therefore, how to establish a general material property database with excellent data quality is an unavoidable and important problem for scientific research institutions.

6.2. Investigation of the influence of materials electronic structure on catalytic performance

For ML models used for screening catalytic materials, the reliability of the predicted results is affected not only by the data quality, but also by the data featurization process. Since there are many factors affecting the reaction performance, it is very important to find the key factors affecting the catalytic performance quickly and accurately for the establishment of ML model when dealing with a new structural system of materials. However, the featurization process requires researchers to have a rich and comprehensive knowledge of chemistry and physics as a basis. Recently, for some catalytic reactions that can be modeled on surface adsorption, many researchers have established general descriptors to correlate the relationship between the electronic/geometric structure of materials and the theoretical properties, which lays a foundation for the establishment of ML models. Therefore, for a new catalytic reaction or material system, before establishing ML model, more experts with rich chemical or physical knowledge background are needed to simulate and analyze the material or reaction, find out the core factors and establish descriptors.

6.3. Intelligent screening of synthetic paths for catalytic materials

At present, research on synthetic path screening of materials by AI-assisted retrosynthesis analysis method mainly focuses on organic materials and pharmaceutical materials, and there are few reports on synthetic path screening of catalytic materials by intelligent method. For the screening and synthesis of catalytic materials, after the target materials are screened by ML, researchers usually need to review a lot of literature and summarize possible synthesis routes, and finally try experiments, which cannot always get satisfactory results. Compared with organic synthesis, the synthesis route of catalytic materials is more diversified, and there is no general synthesis rule for backward analysis. Therefore, researchers can set up intelligent platforms for targeted catalytic materials, automatically read literature and summarize possible synthesis routes, which will save researchers time in searching for material synthesis methods.

6.4. Intelligent prediction of the kinetics of catalytic reaction processes

It is difficult to simulate the kinetic process and the change of catalyst in the catalytic reaction due to the complex influencing factors. Researchers usually intelligently predict the change or kinetic process of the catalyst during the catalytic process through some In-situ techniques

or DFT simulations. However, the In-situ characterization technique requires the collaboration of multiple devices to collect reaction data in real time. The complex process of data collection is difficult to meet the need of revealing the catalytic reaction mechanism. DFT simulations are usually based on idealized models, gradually adding important factors such as solvent system, electric field, magnetic field, etc. Moreover, the more considerations that are added to the simulation, the more computing power is required. In conclusion, DFT and in situ characterization techniques are still unable to reveal the catalytic process at present. By using AI technology, DFT simulation and In-situ data are used as data sources to establish ML models, which can finally achieve a reasonable prediction of the kinetic data in the catalytic process or the structural properties of the catalyst.

Designing and synthesizing new catalytic materials using AI technology presents both opportunities and challenges. At present, more and more researchers are no longer satisfied with the traditional way of research and development. Therefore, to better use computer technology to assist the research and development of new materials, it is necessary for computer experts and chemical experts to work together to reveal the catalytic mechanism of materials and apply intelligent algorithms more widely.

Declaration of competing interest

The authors declare that they have no conflicts of interest in this work.

Acknowledgments

This work was supported by the National Key Research and Development Program of China (2022YFB3807500); the National Natural Science Foundation of China (22220102003); the Beijing Natural Science Foundation (JL23003); Talent cultivation of State Key Laboratory of Organic-Inorganic Composites; “Double-First-Class” construction projects (XK180301, XK1804–02), National Key Research and Development Program of China Stem Cell and Translational Research.

References

[1] S.S. Siwal, W. Yang, Q. Zhang, Recent progress of precious-metal-free electrocatalysts for efficient water oxidation in acidic media, J. Energy Chem. 51 (2020) 113–133.
[2] Y.J. Lee, H.E. Kim, H. Oh, et al., Lens-shaped carbon particles with perpendicularly-oriented channels for high-performance proton exchange membrane fuel cells, ACS Nano 16 (2) (2022) 2988–2996.
[3] L. Fan, H. Deng, Y. Zhang, et al., Towards ultralow platinum loading proton exchange membrane fuel cells, Energy Environ. Sci. 16 (4) (2023) 1466–1479.
[4] J.W.M. Crawley, I.E. Gow, N. Lawes, et al., Heterogeneous trimetallic nanoparticles as catalysts, Chem. Rev. 122 (6) (2022) 6795–6849.
[5] B.M. Hunter, H.B. Gray, A.M. Müller, Earth-abundant heterogeneous water oxidation catalysts, Chem. Rev. 116 (22) (2016) 14120–14136.
[6] Y. Wen, P. Chen, L. Wang, et al., Stabilizing highly active Ru sites by suppressing lattice oxygen participation in acidic water oxidation, J. Am. Chem. Soc. 143 (17) (2021) 6482–6490.
[7] P. Sudarsanam, R. Zhong, S. Van den Bosch, et al., Functionalised heterogeneous catalysts for sustainable biomass valorisation, Chem. Soc. Rev. 47 (22) (2018) 8349–8402.
[8] A. West, Heterogeneous catalysis: Tuning up a hybrid catalyst, Nature Rev. Chem. 2 (4) (2018).
[9] H. Xiong, A.K. Datye, Y. Wang, Thermally stable single-atom heterogeneous catalysts, Adv. Mater. 33 (50) (2021) 2004319.
[10] X. Miao, L. Zhang, L. Wu, et al., Quadruple perovskite ruthenate as a highly efficient catalyst for acidic water oxidation, Nat. Commun. 10 (1) (2019) 3809.
[11] Y. Liang, W. Zhou, Y. Shi, et al., Unveiling in situ evolved in/in(2)o(3-)(x) heterostructure as the active phase of in(2)o(3) toward efficient electroreduction of CO(2) to formate, Sci. Bull. 65 (18) (2020) 1547–1554.
[12] Y.-S. Liu, P.-A. Glans, C.-H. Chuang, et al., Perspectives of in situ/operando resonant inelastic x-ray scattering in catalytic energy materials science, J. Electron. Spectros. Relat. Phenomena 200 (2015) 282–292.
[13] S. Fukuzumi, Production of liquid solar fuels and their use in fuel cells, Joule 1 (4) (2017) 689–738.
[14] Y. Gu, Y. Liu, X. Cao, Evolving strategies for tumor immunotherapy: Enhancing the enhancer and suppressing the suppressor, Natl. Sci. Rev. 4 (2) (2017) 161–163.
[15] N. Tian, C. Hu, J. Wang, et al., Layered bismuth-based photocatalysts, Coord. Chem. Rev. 463 (2022) 214515.
[16] E. Fabbri, A. Habereder, K. Waltar, et al., Developments and perspectives of oxide-based catalysts for the oxygen evolution reaction, Catal. Sci. Technol. 4 (11) (2014) 3800–3821.
[17] X. Guo, S. Lin, J. Gu, et al., Simultaneously achieving high activity and selectivity toward two-electron O_2 electroreduction: The power of single-atom catalysts, ACS Catal. 9 (12) (2019) 11042–11054.
[18] S. Chen, H. Huang, P. Jiang, et al., Mn-doped RuO_2 nanocrystals as highly active electrocatalysts for enhanced oxygen evolution in acidic media, ACS Catal. 10 (2) (2019) 1152–1160.
[19] A. Jain, Y. Shin, K.A. Persson, Computational predictions of energy materials using density functional theory, Nat. Rev. Mater. 1 (1) (2016) 15004.
[20] H. Niu, X. Wang, C. Shao, et al., Revealing the oxygen reduction reaction activity origin of single atoms supported on g-C_3N_4 monolayers: A first-principles study, J. Mater. Chem. A 8 (14) (2020) 6555–6563.
[21] H. Xu, D. Cheng, D. Cao, et al., A universal principle for a rational design of single-atom electrocatalysts, Nat. Catal. 1 (5) (2018) 339–348.
[22] A. Mohajeri, N.L. Dashti, Cooperativity in bimetallic sacs: An efficient strategy for designing bifunctional catalysts for overall water splitting, J. Phys. Chem. C 123 (51) (2019) 30972–30980.
[23] M. Krenn, J. Landgraf, T. Foesel, et al., Artificial intelligence and machine learning for quantum technologies, Phys. Rev. A 107 (1) (2023) 010101.
[24] H. Masood, C.Y. Toe, W.Y. Teoh, et al., Machine learning for accelerated discovery of solar photocatalysts, ACS Catal. 9 (12) (2019) 11774–11787.
[25] Q. Zhu, F. Zhang, Y. Huang, et al., An all-round ai-chemist with a scientific mind, Natl. Sci. Rev. 9 (10) (2022) nwac190.
[26] E.O. Pyzer-Knapp, J.W. Pitera, P.W.J. Staar, et al., Accelerating materials discovery using artificial intelligence, high performance computing and robotics, NPJ Comput. Mater. 8 (1) (2022) 84.
[27] C. Fare, P. Fenner, M. Benatan, et al., A multi-fidelity machine learning approach to high throughput materials screening, NPJ Comput. Mater. 8 (1) (2022) 257.
[28] H. Mai, T.C. Le, D. Chen, et al., Machine learning for electrocatalyst and photocatalyst design and discovery, Chem. Rev. 122 (16) (2022) 13478–13515.
[29] P. Xu, X. Ji, M. Li, et al., Small data machine learning in materials science, NPJ Comput. Mater. 9 (1) (2023) 42.
[30] B.J. Shields, J. Stevens, J. Li, et al., Bayesian reaction optimization as a tool for chemical synthesis, Nature 590 (7844) (2021) 89–96.
[31] K.M. Jablonka, D. Ongari, S.M. Moosavi, et al., Big-data science in porous materials: Materials genomics and machine learning, Chem. Rev. 120 (16) (2020) 8066–8129.
[32] Z. Fu, W. Liu, C. Huang, et al., A review of performance prediction based on machine learning in materials science, Nanomaterials 12 (17) (2022) 2957.
[33] B. Han, Y. Lin, Y. Yang, et al., Deep-learning-enabled fast optical identification and characterization of 2d materials, Adv. Mater. 32 (29) (2020) e2000953.
[34] Y. Jiang, Y. Yu, M. Kong, et al., Artificial intelligence for retrosynthesis prediction, Engineering 25 (2022) 32–50.
[35] Y. Shen, J.E. Borowski, M.A. Hardy, et al., Automation and computer-assisted planning for chemical synthesis, Nat. Rev. Methods Primers 1 (1) (2021) 23.
[36] B. Burger, P.M. Maffettone, V.V. Gusev, et al., A mobile robotic chemist, Nature 583 (7815) (2020) 237–241.
[37] B. S-L, A. A-G, Inverse molecular design using machine learning: Generative models for matter engineering, Science 361 (6400) (2018) 360–365.
[38] A. Wolos, D. Koszelewski, R. Roszak, et al., Computer-designed repurposing of chemical wastes into drugs, Nature 604 (7907) (2022) 668–676.
[39] Y. Zhang, Y. Lin, T. Duan, et al., Interfacial engineering of heterogeneous catalysts for electrocatalysis, Mater. Today 48 (2021) 115–134.
[40] X. Zhong, B. Gallagher, S. Liu, et al., Explainable machine learning in materials science, NPJ Comput. Mater. 8 (1) (2022) 204.
[41] C. Gao, X. Min, M. Fang, et al., Innovative materials science via machine learning, Adv. Funct. Mater. 32 (1) (2021) 2108044.
[42] K.T. Butler, D.W. Davies, H. Cartwright, et al., Machine learning for molecular and materials science, Nature 559 (7715) (2018) 547–555.
[43] Z. Lian, M. Yang, F. Jan, et al., Machine learning derived blueprint for rational design of the effective single-atom cathode catalyst of the lithium-sulfur battery, J. Phys. Chem. Lett. 12 (29) (2021) 7053–7059.
[44] S. Koutsoukos, F. Philippi, F. Malaret, et al., A review on machine learning algorithms for the ionic liquid chemical space, Chem. Sci. 12 (20) (2021) 6820–6843.
[45] Y. Liu, C. Niu, Z. Wang, et al., Machine learning in materials genome initiative: A review, J. Mater. Sci. Technol. 57 (2020) 113–122.
[46] V. Dhayalan, S.C. Gadekar, Z. Alassad, et al., Unravelling mechanistic features of organocatalysis with in situ modifications at the secondary sphere, Nat. Chem. 11 (6) (2019) 543–551.
[47] L.M. Ghiringhelli, J. Vybiral, S.V. Levchenko, et al., Big data of materials science: Critical role of the descriptor, Phys. Rev. Lett. 114 (10) (2015) 105503.
[48] L. Simon-Vidal, O. Garcia-Calvo, U. Oteo, et al., Perturbation-theory and machine learning (ptml) model for high-throughput screening of parham reactions: Experimental and theoretical studies, J. Chem. Inf. Model. 58 (7) (2018) 1384–1396.
[49] S. Singh, M. Pareek, A. Changotra, et al., A unified machine-learning protocol for asymmetric catalysis as a proof of concept demonstration using asymmetric hydrogenation, Proc. Natl. Acad. Sci. U.S.A. 117 (3) (2020) 1339–1345.
[50] B. Medasani, A. Gamst, H. Ding, et al., Predicting defect behavior in b2 intermetallics by merging ab initio modeling and machine learning, NPJ Comput. Mater. 2 (1) (2016) 1.
[51] Y. Sun, H. Liao, J. Wang, et al., Covalency competition dominates the water oxidation structure–activity relationship on spinel oxides, Nat. Catal. 3 (7) (2020) 554–563.

[52] Z. Wang, Y. Gu, L. Zheng, et al., Machine learning guided dopant selection for metal oxide-based photoelectrochemical water splitting: The case study of Fe_2O_3 and CuO, Adv. Mater. 34 (10) (2022) e2106776.
[53] M. Zhong, K. Tran, Y. Min, et al., Accelerated discovery of CO_2 electrocatalysts using active machine learning, Nature 581 (7807) (2020) 178–183.
[54] Z. Guo, P. Liu, J. Liu, et al., Neural network inspired design of highly active and durable N-doped carbon interconnected molybdenum phosphide for hydrogen evolution reaction, ACS Appl. Energy Mater. 1 (2018) 5437–5445.
[55] D. Cheng, Z.J. Zhao, G. Zhang, et al., The nature of active sites for carbon dioxide electroreduction over oxide-derived copper catalysts, Nat. Commun. 12 (1) (2021) 395.
[56] A.J. Medford, M.R. Kunz, S.M. Ewing, et al., Extracting knowledge from data through catalysis informatics, ACS Catal. 8 (8) (2018) 7403–7429.
[57] J.S. Smith, O. Isayev, A.E. Roitberg, Ani-1: An extensible neural network potential with dft accuracy at force field computational cost, Chem. Sci. 8 (4) (2017) 3192–3203.
[58] B.B. Hoar, S. Lu, C. Liu, Machine-learning-enabled exploration of morphology influence on wire-array electrodes for electrochemical nitrogen fixation, J. Phys. Chem. Lett. 11 (12) (2020) 4625–4630.
[59] M.H.S. Segler, M. Preuss, M.P. Waller, Planning chemical syntheses with deep neural networks and symbolic ai, Nature 555 (7698) (2018) 604–610.
[60] P. Pankajakshan, S. Sanyal, O.E. de Noord, et al., Machine learning and statistical analysis for materials science: Stability and transferability of fingerprint descriptors and chemical insights, Chem. Mater. 29 (10) (2017) 4190–4201.
[61] X. Duan, F. Yang, E. Antono, et al., Unsupervised data mining in nanoscale x-ray spectro-microscopic study of ndfeb magnet, Sci. Rep. 6 (2016) 34406.
[62] X. Liu, Y. Zhang, W. Wang, et al., Transition metal and N doping on alp monolayers for bifunctional oxygen electrocatalysts: Density functional theory study assisted by machine learning description, ACS Appl. Mater. Interfaces 14 (1) (2022) 1249–1259.
[63] C. Sutton, M. Boley, L.M. Ghiringhelli, et al., Identifying domains of applicability of machine learning models for materials science, Nat. Commun. 11 (1) (2020) 4428.
[64] K.M. Jablonka, D. Ongari, S.M. Moosavi, et al., Using collective knowledge to assign oxidation states of metal cations in metal-organic frameworks, Nat. Chem. 13 (8) (2021) 771–777.
[65] E.J. Corey, W.T. Wipke, Computer-assisted design of complex organic syntheses, Science 166 (1969) 178.
[66] H. Lu, D.J. Diaz, N.J. Czarnecki, et al., Machine learning-aided engineering of hydrolases for pet depolymerization, Nature 604 (7907) (2022) 662–667.
[67] C.T. Madsen, J.C. Refsgaard, F.G. Teufel, et al., Combining mass spectrometry and machine learning to discover bioactive peptides, Nat. Commun. 13 (1) (2022) 6235.
[68] B. Mikulak-Klucznik, P. Golebiowska, A.A. Bayly, et al., Computational planning of the synthesis of complex natural products, Nature 588 (7836) (2020) 83–88.
[69] Y.F. Shi, P.L. Kang, C. Shang, et al., Methanol synthesis from CO_2/CO mixture on Cu-Zn catalysts from microkinetics-guided machine learning pathway search, J. Am. Chem. Soc. 144 (29) (2022) 13401–13414.
[70] F. Liu, C. Duan, H.J. Kulik, Rapid detection of strong correlation with machine learning for transition-metal complex high-throughput screening, J. Phys. Chem. Lett. 11 (19) (2020) 8067–8076.
[71] I.G. Clayson, D. Hewitt, M. Hutereau, et al., High throughput methods in the synthesis, characterization, and optimization of porous materials, Adv. Mater. 32 (44) (2020) e2002780.
[72] L. An, C. Wei, M. Lu, et al., Recent development of oxygen evolution electrocatalysts in acidic environment, Adv. Mater. 33 (20) (2021) 2006328.
[73] Z.W. Seh, J. Kibsgaard, C.F. Dickens, et al., Combining theory and experiment in electrocatalysis: Insights into materials design, Science 355 (6321) (2017) eaad4998.
[74] X. Han, W. Zhang, X. Ma, et al., Identifying the activation of bimetallic sites in $NiCo_2S_4$@g-C_3N_4-CNT hybrid electrocatalysts for synergistic oxygen reduction and evolution, Adv. Mater. 31 (18) (2019) 1808281.
[75] S. Lu, Q. Zhou, Y. Ouyang, et al., Accelerated discovery of stable lead-free hybrid organic-inorganic perovskites via machine learning, Nat. Commun. 9 (1) (2018) 3405.
[76] I.C. Man, H.Y. Su, F. Calle-Vallejo, et al., Universality in oxygen evolution electrocatalysis on oxide surfaces, ChemCatChem 3 (7) (2011) 1159–1165.
[77] C. Liu, J. Qian, Y. Ye, et al., Oxygen evolution reaction over catalytic single-site Co in a well-defined brookite TiO_2 nanorod surface, Nat. Catal. 4 (1) (2020) 36–45.
[78] S. Zhai, H. Xie, P. Cui, et al., A combined ionic lewis acid descriptor and machine-learning approach to prediction of efficient oxygen reduction electrodes for ceramic fuel cells, Nat. Energy 7 (9) (2022) 866–875.
[79] B. Tang, Y. Lu, J. Zhou, et al., Machine learning-guided synthesis of advanced inorganic materials, Mater. Today 41 (2020) 72–80.
[80] Y. Guo, X. He, Y. Su, et al., Machine-learning-guided discovery and optimization of additives in preparing Cu catalysts for CO_2 reduction, J. Am. Chem. Soc. 143 (15) (2021) 5755–5762.
[81] H. Chun, E. Lee, K. Nam, et al., First-principle-data-integrated machine-learning approach for high-throughput searching of ternary electrocatalyst toward oxygen reduction reaction, Chem. Catal 1 (4) (2021) 855–869.
[82] C.C. Price, A. Singh, N.C. Frey, et al., Efficient catalyst screening using graph neural networks to predict strain effects on adsorption energy, Sci. Adv. 8 (2022) eabq5944.
[83] J.A. Esterhuizen, B.R. Goldsmith, S. Linic, Uncovering electronic and geometric descriptors of chemical activity for metal alloys and oxides using unsupervised machine learning, Chem. Catal 1 (4) (2021) 923–940.
[84] W. Xu, H. Tang, H. Gu, et al., Research progress of asymmetrically coordinated single-atom catalysts for electrocatalytic reactions, J. Mater. Chem. A 10 (28) (2022) 14732–14746.
[85] M. Babucci, A. Guntida, B.C. Gates, Atomically dispersed metals on well-defined supports including zeolites and metal-organic frameworks: Structure, bonding, reactivity, and catalysis, Chem. Rev. 120 (21) (2020) 11956–11985.
[86] S. Ji, Y. Chen, X. Wang, et al., Chemical synthesis of single atomic site catalysts, Chem. Rev. 120 (21) (2020) 11900–11955.
[87] X. Zhao, X. Liu, B. Huang, et al., Hydroxyl group modification improves the electrocatalytic ORR and OER activity of graphene supported single and bi-metal atomic catalysts (Ni, Co, and Fe, J. Mater. Chem. A 7 (42) (2019) 24583–24593.
[88] L. Wu, T. Guo, T. Li, Machine learning-accelerated prediction of overpotential of oxygen evolution reaction of single-atom catalysts, iScience 24 (5) (2021) 102398.
[89] Y. Ying, K. Fan, X. Luo, et al., Unravelling the origin of bifunctional OER/ORR activity for single-atom catalysts supported on C_2N by dft and machine learning, J. Mater. Chem. A 9 (31) (2021) 16860–16867.
[90] M. Sun, T. Wu, Y. Xue, et al., Mapping of atomic catalyst on graphdiyne, Nano Energy 62 (2019) 754–763.
[91] F. Sandfort, F. Strieth-Kalthoff, M. Kühnemund, et al., A structure-based platform for predicting chemical reactivity, Chem 6 (6) (2020) 1379–1390.
[92] Y. Jiang, D. Salley, A. Sharma, et al., An artificial intelligence enabled chemical synthesis robot for exploration and optimization of nanomaterials, Sci. Adv. 8 (2022) eabo2626.
[93] L. Xing, Z. Chen, W. Chen, et al., Robotic platform for accelerating the high-throughput study of silver nanocrystals in sensitive/selective hg2+ detection, Chem. Eng. J. 466 (2023) 143225.
[94] H. Zhao, W. Chen, H. Huang, et al., A robotic platform for the synthesis of colloidal nanocrystals, Nat. Synth. 2 (6) (2023) 505–514.
[95] C. Zheng, K. Mathew, C. Chen, et al., Automated generation and ensemble-learned matching of x-ray absorption spectra, NPJ Comput. Mater. 4 (1) (2018) 12.
[96] K. Mathew, C. Zheng, D. Winston, et al., High-throughput computational x-ray absorption spectroscopy, Sci. Data 5 (2018) 180151.

Author profile

Linkai Han is now a PhD student in chemical engineering at the Beijing University of Chemical Technology (BUCT) under the supervision of Prof. Zhonghua Xiang. His current research direction is the theoretical design and synthesis of water electrolytic anode reaction catalytic materials

Zhonghua Xiang (BRID: 03851.00.19661) is a professor and director of the Molecular Energy Materials R&D Center at BUCT. He received his PhD in 2013 at BUCT and was a postdoctoral researcher at Case Western Reserve University (2013–2014). His research interests are focused on the design and synthesis of molecular energy materials, mainly including covalent–organic frameworks for fuel cells and flow battery.