REVIEW

ADVANCED
INTELLIGENT
SYSTEMS
Open Access

www.advintellsyst.com

Artificial Intelligence to Power the Future of Materials Science and Engineering

Wuxin Sha, Yaqing Guo, Qing Yuan, Shun Tang, Xinfang Zhang, Songfeng Lu, Xin Guo, Yuan-Cheng Cao,* and Shijie Cheng

Artificial intelligence (AI) has received widespread attention over the last few decades due to its potential to increase automation and accelerate productivity. In recent years, a large number of training data, improved computing power, and advanced deep learning algorithms are conducive to the wide application of AI, including material research. The traditional trial-and-error method is inefficient and time-consuming to study materials. Therefore, AI, especially machine learning, can accelerate the process by learning rules from datasets and building models to predict. This is completely different from computational chemistry where a computer is only a calculator, using hard-coded formulas provided by human experts. Herein, the application of AI in material innovation is reviewed, including material design, performance prediction, and synthesis. The realization details of AI techniques and advantages over conventional methods are emphasized in these applications. Finally, the future development direction of AI is expounded from both algorithm and infrastructure aspects.

1. The Merging of Materials Science and Artificial Intelligence

From the Paleolithic Age to the coming fourth industrial revolution, the millions of years of human history is mainly marked by materials. Material science is mainly to explore the relationship between materials structure, process, properties, and application. The discovery of new materials will play a greater role in promoting the development of human society. After several centuries of development, a large amount of data has been accumulated in the field of materials science.[1] However, the inherent limitations of human cognitive ability make it difficult for human beings to absorb and process the massive literature and data produced every day.[2] Only a small part of data (compared with the whole data volume) can be analyzed in a certain subdivision field. The current material research is mainly a "trial-and-error method" based on a large number of experiments guided by experience, and a small number of computer simulation calculation as a supplement, which consumes a lot of manpower, time, materials, and financial resources.[3] The vast amount of material information data are always silent in the database or used little by little. Therefore, finding a new research method is necessary to accelerate material innovation.[4]

The emergence of artificial intelligence (AI) brings a new dawn to the development of material science.[5] After more than 60 years of development, from the simple perceptron[6] to complex multilayer neural networks,[7] AI has exhibited a primary algorithm framework and a powerful hardware foundation.[8–13] Some advanced AI system even defeated world champions in many domains, such as Chess,[14] Go,[15] quiz game,[16] and other fields.[17–23] The excellent data mining ability of AI has attracted the wide attention of the material science community.[24–27] Jim Gray, the winner of Turing Award, proposed "the fourth paradigm of science" at the NRC-CSTB conference[28] in 2007. It is a data-intensive science that combines big data and AI to compress lots of known information into unknown theories to guide scientific innovation.[29] This method is suitable for dealing with large-scale composite space or nonlinear processes, which reminds some problems in material research. Materials informatics, the combination of materials science and AI techniques, is such an interdiscipline to help scientists to effectively obtain the hidden relationship between different variables, predict the specific properties of materials, guide the chemical synthesis route, optimize the process parameters, and upgrade the existing material characterization methods.

Machine learning (ML) is an important branch of AI which develops rapidly in recent years, and it is also the most promising

W. Sha, Q. Yuan, Dr. X. Zhang, Dr. S. Lu
School of Computer Science and Technology
Huazhong University of Science and Technology
Wuhan 430074, China

W. Sha, Y. Guo, Dr. S. Tang, Prof. Y.-C. Cao, Prof. S. Cheng
State Key Laboratory of Advanced Electromagnetic Engineering and Technology
School of Electrical and Electronic Engineering
Huazhong University of Science and Technology
Wuhan 430074, China
E-mail: yccao@hust.edu.cn

Prof. X. Guo
School of Materials Science and Engineering
Huazhong University of Science and Technology
Wuhan 430074, China

The ORCID identification number(s) for the author(s) of this article can be found under https://doi.org/10.1002/aisy.201900143.

**ADVANCED
SCIENCE NEWS**

www.advancedsciencenews.com

**ADVANCED
INTELLIGENT
SYSTEMS**
Open Access

www.advintellsyst.com

application of AI in the research of material science. The next part introduces the basic knowledge of ML, which lays the foundation to introduce the materials research applications of AI in later text.

2. Basics of ML

ML describes a computer's ability to train on a set of data and then find the regulations or knowledge underlying that data. To be specific, ML is mainly divided into four steps: data collection, data representation, algorithm selection, and model optimization.[30]

2.1. Data Collection

ML is a kind of data-driven algorithms, and data can be obtained by simulations (such as density functional theory [DFT] and molecular dynamics [MDs]), experiments, and online database.[31] Data include physical properties and structural information on some materials. Many data in the field of materials are missing, repeated, and inconsistent because of the limitation of environment and experimental conditions. So, data cleaning, to identify and correct different errors in original data, becomes fairly necessary.[32] For missing values, the average, minimum, or other statistical values of the attribute are used to fill in the vacancy as appropriate.[33–35] For repeated values, the basic idea of eliminating duplicate records is sorting by attribute values and merging records with identical value. The related algorithms include priority queue algorithm, sorted-neighborhood method, and so on. Such methods have been used in perovskite data by merging different entries in the Materials Project database and the Inorganic Crystal Structure Database.[36] For inconsistent values, according to the reasonable value range and mutual relationship of each variable, specific programs can be designed to check whether the data meet the requirements.[37] Data beyond the normal range or conflicting attributes will be deleted appropriately. After cleaning, the data can be used for data representation.

2.2. Data Representation

Data representation is converting the raw data into some forms suitable for an algorithm. The data we collect is usually numeric but may not be appropriate for the algorithm. Just as when we solve mathematical problems, we prefer to list equations or plot-relevant figures to help us understand better. ML algorithms also need an appropriate form of input data to learn better. The more appropriate representation we use, the better the model performs.

One of the methods to represent physical properties and structural information is binary coding. Granda et al. proposed an organic synthesis robot.[38] By binary coding the chemical input, the robot can analyze the reactivity of reagent combinations, and use support vector machine (SVM) model to predict unknown chemical reactions.

2.3. Algorithm Selection

ML is generally classified into supervised learning (such as classification and regression) and unsupervised learning (such as

Wuxin Sha received his bachelor's degree in the School of Materials Science and Engineering from Huazhong University of Science and Technology (HUST) in 2017. He is currently pursuing his Ph.D. degree in the School of Computer Science and Technology, HUST. His research interests focus on AI-assisted materials genome, ML, and solid-state electrolytes lithium batteries.

Yuan-Cheng Cao is currently a professor of the State Key Laboratory of Advanced Electromagnetic Engineering and Technology at Huazhong University of Science and Technology (HUST, Wuhan). He received his Ph.D. degree from HUST in 2006. Then he worked at Nottingham Trent University (UK, 2007–2010), Newcastle University (UK, 2010–2014), and Jianghan University (Wuhan, 2014–2018). His current research interests include solid-state electrolytes in energy-storage batteries, safety and extinguishing control for grid energy storage, eco-friendly recycling, and regeneration of decommissioned batteries.

Shijie Cheng is a professor of Huazhong University of Science and Technology. He received his bachelor's degree from Xi'an Jiaotong University in 1967, his master's degree from HUST in 1981, and his Ph.D. degree from the University of Calgary (Canada) in 1986, respectively, all in electrical engineering. In 2007, he was elected as a member of the Chinese Academy of Sciences. He is currently engaged in research on energy-storage systems for electric power system stability and advanced materials for electrical engineering.

clustering), depending on whether the training data are labeled or not. Due to the recent improvement in materials automation, reinforcement learning and active learning, which need to interact with the environment, are also emerging in the application of materials research. Currently, the most popular algorithms include k-nearest neighbor (KNN), decision tree symbolic regression, and artificial neural networks. A brief introduction about these methods will be provided in the following sections.

KNN is a classification and regression algorithm, which is very simple and effective.[39] Given a training dataset and a new datum, the algorithm finds k entries in a dataset that are nearest with the new datum, and the new datum will be classified in the category which appears most frequently. The algorithm consists of the selection of k, distance measurement, and the rule of classification. The model complexity degree increasing with

ADVANCED
SCIENCE NEWS

www.advancedsciencenews.com

ADVANCED
INTELLIGENT
SYSTEMS
Open Access

www.advintellsyst.com

the k becomes smaller, the approximation error will decrease, and the estimation error will increase. Using different distance to measure similarity with two points may lead to different results. KNN can select Euclidean distance, Manhattan distance, and so on. KNN usually selects majority voting as a rule of classification, because it means empirical error minimization.

Decision tree is one of the simplest and most successful algorithms in ML.[40] A decision tree represents a classifier which takes a series of attribute values as input and outputs a decision. The input and output values can be either discrete or continuous. If the inputs are discrete and output only has two possible values, it is called Boolean classification. A decision tree outputs its decision by performing a set of tests. In decision trees, each node represents a test of the value of one of the input attributes, and the branches from it are possible values of the attribute. Each leaf node is a value which is returned by the function.

Symbolic regression, especially genetic programming-based symbolic regression (GPSR), is a classical AI algorithm.[41] It is different from the traditional numerical regression because the functional relationship between variables is not given. Instead, the functional form is gained by the evolution of chromosomes in each candidate function. The chromosomes consist of a set of internal nodes with mathematical operation symbols and terminal nodes with variables and constants. The depth-first search algorithm can be used to traverse chromosomes to obtain the corresponding function. The error between the experimental data and the fitted data by the function is used as the evaluation function. The candidate functions with the smallest error, and the largest adaptability could create descendants preferentially. Different chromosomes pass through mutation and heredity, and gradually iterate until the best form of function and parameter set for a given problem is found.[42] GPSR is suitable for the field of material research with little prior knowledge and unclear relationship between related variables, such as the magic angle in graphene,[43] the viscosity of normal hydrogen,[44] and the search for descriptors of perovskite stability.[45]

Inspired by the hypothesis that mental activity primarily consists of electrochemical activity in networks of brain cells called neurons, artificial neural networks are created. Neural network consists of nodes connected by directed links. Each link between nodes serves to propagate activation and has a numeric weight associated with it, which determines the strength and sign of the connection. There are two basic ways to connect nodes to form a network. If nodes are connected in one direction, the network is a feed-forward network. If a network feeds its outputs back into its inputs, it is a recurrent network. The most commonly used network consists of more than three layers, including the input layer, the output layer, and hidden layers. The learning process is to find appropriate parameters to minimize the output error rate. After training and testing strategy, the model is well-established.

There are more efficient ML algorithms in addition to that mentioned earlier, such as random forests, kernel methods, convolutional neural networks, and generative adversarial networks (GAN). Whatever algorithm selected, there are some hyperparameters to be estimated by human or other heuristic algorithms. Recently, there are more researches in automatic ML, which aims to make it easier for people to apply ML algorithms.

2.4. Model Optimization

The model which has higher-degree polynomials can fit the training data better, but it will overfit and perform poorly on validation data if the degree is too high. There are two ways to choose the degree of the polynomial: cross-validation and regularization to directly minimize the weighted sum of the empirical loss and the complexity of the model.

To search for a model with as low as possible error rate, loss function is usually used. The loss function is defined to measure the distance of correct values and predicted values. By minimizing the loss function, the best hypothesis can be found. Cross-validation is reliable only when the samples used for training and validation are representative of the whole population.

3. AI Applications for Materials Science and Engineering

In recent years, AI has been applied in more and more fields, and ML research in the field of materials is rapidly developing, especially in that it can synthesize new materials and predict various chemical synthesis.[46,47] In this section, we will explore how ML can help people solve the barriers between designing, synthesizing, and processing materials.[48–54]

3.1. Accelerated Simulation

The research process for computational chemistry and materials science has been updated to the third generation. The first generation refers to the calculation of "structure-performance", which mainly takes advantage of the local optimization algorithm to predict the performance of the materials from the structure. The second is "crystal structure prediction", which mainly adopts global optimization algorithm to predict structure and performance from element composition. The third generation recognized as "statistically driven design," utilizes ML algorithms to predict the composition, structure, and performance of elements from physical and chemical data.[55,56] However, the imperfection of the theory has also brought obstacles to the discovery of high-performance materials and the parameters of the model are not completely consistent with the practical conditions such as mixed phase or grain boundary. For example, the DFT prediction[57] of zirconium-doped lithium tantalum silicate is $10^{-3} \, S \, cm^{-1}$, whereas subsequent experiments have shown that its actual conductivity is about $10^{-5} \, S \, cm^{-1}$.[58] Therefore, finding ways to use ML to make up for the deficiencies of simulation is very important.[59,60]

3.1.1. Atom2vec

Atom2Vec, an unsupervised ML program, reconstructed the periodic table of elements only in a few hours. Atom2Vec first learns to distinguish different atoms by analyzing the list of compounds in the online database. Then, we borrow the simple concept of natural language processing: the characteristics of a word can be derived from other words around it; chemical elements are clustered according to their chemical environment. At the same time, the vectorized atomic descriptor can be used as

the input of many ML models because it carries a large amount of information about the periodic law of elements, which provides an effective new way for the quantitative representation of material data in the future.[61]

3.1.2. Increasing Simulation Scale

Because there are some regular repetitions in the theoretical calculation of atomic force field, once ML finds these repetitive patterns, the corresponding energy or force field can be calculated quickly. The movement of hundreds of atoms in a few picoseconds can be enlarged to that of millions of atoms in a few nanoseconds, which greatly increases the length and time range of the simulation calculation, and achieves better results. Complex material structures (such as amorphous, polycrystalline) and chemical reactions (corrosion, interfacial reactions, etc.) might be simulated.

In large-scale MDs, simulations of surface and interfacial chemical processes, the development of reliable interatomic potentials is a formidable challenge because of the existence of a wide range of atomic environments and very different types of bonds. In recent years, the interatomic potential based on artificial neural networks (NNs) has emerged, which provides an unbiased method for the construction of potential energy surface of systems that are difficult to describe by traditional potential. Artrith et al. used copper and zinc oxide as reference systems to verify the accuracy and validity of the interatomic potential of the artificial neural network and described the CuZnO ternary combination system of oxide-supported copper clusters (**Figure 1**).[62] Generally speaking, the potential energy of the neural network is very precise with the results close to the calculation value of the basic reference electronic structure and several orders of magnitude higher efficiency. Compared with other potential-energy calculation methods, the construction of NN potential energy requires higher computational requirements because of the need for a large number of training points. But the advantages of NN in large-scale applications where traditional electronic structure calculation is hard to solve are evident.

3.1.3. Reducing the Amount of Computation

Due to the massive combination spaces of materials, it is difficult to explore all possible combinations in a reasonable time by traditional simulation calculation. For example, the bimetallic configuration of the smallest known sulfide nanocluster $Au_{15}(SR)_{13}$ exceeds 32 000, and traversing all potential structures is a huge computational challenge. However, if a small part of the data is used to train the ML model, and then the model is used to predict the other combinations, the computational complexity will be greatly reduced and the filtering speed will be increased by several orders of magnitude. Panapitiya et al. proposed a ML model based on stochastic forest method to predict CO adsorption energy of nanoclusters.[63] First, the DFT simulation data training model of Ag-alloyed Au_{25} nanoclusters was used. Using two-step feature selection process and feature engineering method, the authors predicted the adsorption energy with accuracies of 0.78 (R2) and 0.17 (RMSE). After interpreting the key nodes of random forest, the authors found that the distribution of Ag atoms in Au_{25} had the most important effect on CO adsorption sites. The ML model can be easily extended to other nanoclusters based on Au. The model is expected to be used as a screening tool to screen eligible materials for further accurate analysis.

3.2. Predicting the Property of New Materials (Mapping Structure-Property Relationship)

Material researchers generally hope that desired properties of materials can be optimized, such as the conductivity of electrolytes, the Seebeck coefficient of thermoelectric materials, and the power conversion efficiency of organic–inorganic hybrid perovskites.[64–66] A large number of trial-and-error experiments based on theoretical simulation or chemical scientists' intuition typically lead to dissatisfactory results. Fortunately, the applications of ML models can help a lot by predicting the properties and structures of materials with an acceptable accuracy before synthesis. Sendek et al. used the ML model developed in MATLAB to find a small amount of special solid electrolytes in more than 12 000 materials.[67] Using a well-known set of electrolytes and their atomic structures for training, they first combed the scientific literature and found 40 solid crystalline materials. Because of the small size of the dataset, it is necessary to use the "intelligent" feature based on existing physical knowledge for data representation. Therefore, the author downloads the atomic structure of these 40 materials from ICSD as input, and calculates 20 kinds of characteristics according to the atomic position, mass, electronegativity, and atomic radius of the structure, including the volume of each atom, the lithium bond ionicity, the number of lithium adjacent elements, and the minimum anion–anion separation distance, and describes the atomic local arrangement and chemical characteristics of each crystal. Then these 20 features are used as inputs, the experimental values of lithium-ion
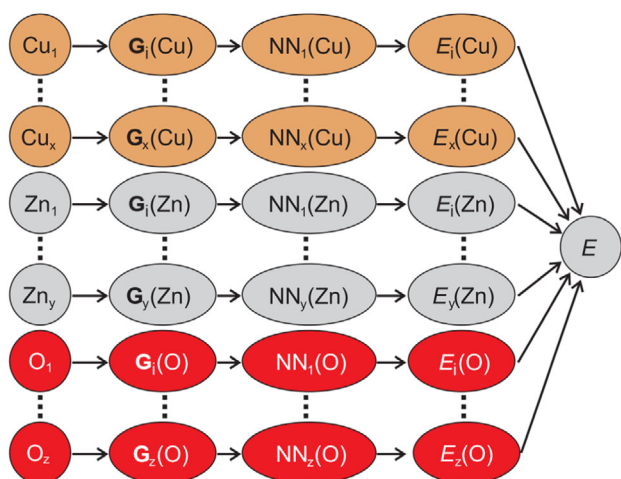

Figure 1. Schematic structure of a high-dimensional neural network potentials for a system of the composition $Cu_xZn_yO_z$. For each atom i in the system there is one line. Each circle on the left side represents the Cartesian coordinate vector of an atom. These are then transformed to symmetry function vectors $\mathbf{G_i}$ describing the local atomic environments. The $\mathbf{G_i}$ are then used as input vectors for atomic NNs yielding the atomic energy contributions E_i to the total energy E. Reproduced with permission.[62] Copyright 2013, Wiley.

conductivity are used as outputs, and 40 known materials constitute the training set of a ML algorithm. After constant parameter adjustment, the model can screen and classify solid electrolytes. Then 317 candidate materials were predicted. The results show that the efficiency of identifying potential new materials using the modified MATLAB model is three times higher than that of random guessing and two times higher than that of Stanford graduate students working in related fields. Compared with DFT results, the F1 score is about 50% (**Figure 2**).

The training data of ML can be not only from experimental tests but also from high-throughput simulations. Li et al. studied the thermodynamic stability of double perovskite halides using high-throughput calculation and ML.[68] First, they established a decomposition energy database based on high-throughput DFT, which was closely related to the thermodynamic stability of 354 perovskite candidates. Based on this database, they trained a ML model. The experimental observation of perovskite formability of 246 $A_2B(I)B(III)X_6$ compounds (F1 score, 95.9%) further verified its prediction performance. This work shows that the ML model prediction is more economical and effective than experimental attempts.

Similar methods have been applied to the design of lead-free organic–inorganic hybrid perovskite,[64] monoatomic catalysts,[69] light-emitting diode (LED),[70] organic light-emitting diode (OLED),[71] and other key materials. The latter two methods have also been verified by experiments. At present, material science is not a complete trial-and-error method. Some theories are still used to reduce the number of experiments, and the demand for reduction will be higher and higher in the future. Or the regression model can be used to select the material with the best interesting performance from a large number of alternative materials, which can effectively reduce the number of error experiments in trial-and-error methods.

3.3. Synthetic Route Planning

Organic synthesis has a standard process that allows scientists to design computer programs to deal with synthetic problems.[72]

As far as computer scientists are concerned, a chemical reaction is a set of data that indicates the relationship or connection of a compound. This presence can be expressed as a data structure, such as a graph or network.[73,74] Then AI could deal with these structural data to guide the synthesis route.[75]

Granda et al. presented an organic synthesis robot that includes online spectral analysis and feedback loop to perform six experiments simultaneously.[38] Its core components include a raw-material tank and a pressure pump assembled with chemicals. These pumps are responsible for feeding reactants into six parallel-operated reaction bottles. In addition, the robot uses the SVM method to automatically classify the reaction mixture into a reactive or nonreactive mixture by real-time evaluation of the reaction using NMR and IR spectroscopy. This method is faster than manual experiments and can predict the reactivity of reagent combinations. Also, after collecting the results of about 10% of the experimental dataset, the robot could predict the reactivity of ≈1000 reaction combinations with a prediction accuracy of over 80% and discovered four new reactions (**Figure 3**).

In addition to data-driven methods, the researchers also used reaction rules to predict retrosynthesis analytic systems and developed logic-based and knowledge-based search strategies to design the reaction route. Therefore, the proposed retrosynthesis method can theoretically obtain a reasonable starting material and a reaction route by analyzing the desired compound. Nowadays, this technology has been applied to synthesize new materials and predict various chemical syntheses.

The difficulty in retrosynthesis is finding ways to express the existing chemical reaction in a data structure amenable to algorithms. Schneider et al. proposed a new chemical reaction fingerprint and classified the organic reaction into 50 models (**Figure 4**).[76] Combining with random forests, Naive Bayes, K-means, and logistic regression methods, they can correctly predict nearly 97% of organic synthesis. In the past 10 years, scientists have used various rule-based algorithms to predict organic reactions. Furthermore, scientists could take advantage of ML to determine which rule the reaction should choose.

Segler et al. first collected about 12.5 million chemical reactions published by 2014.[77] Three different neural networks
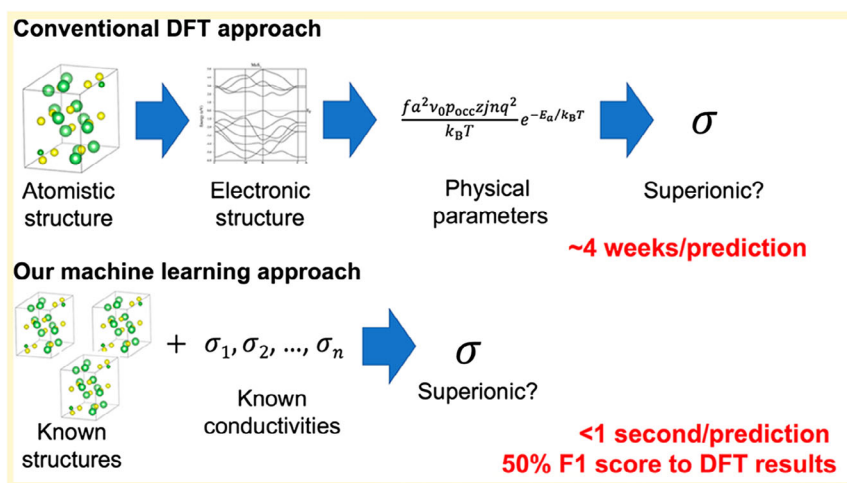

Figure 2. Schematic of comparison between conventional DFT and machine learning approach. Reproduced with permission.[67] Copyright 2018, American Chemical Society.
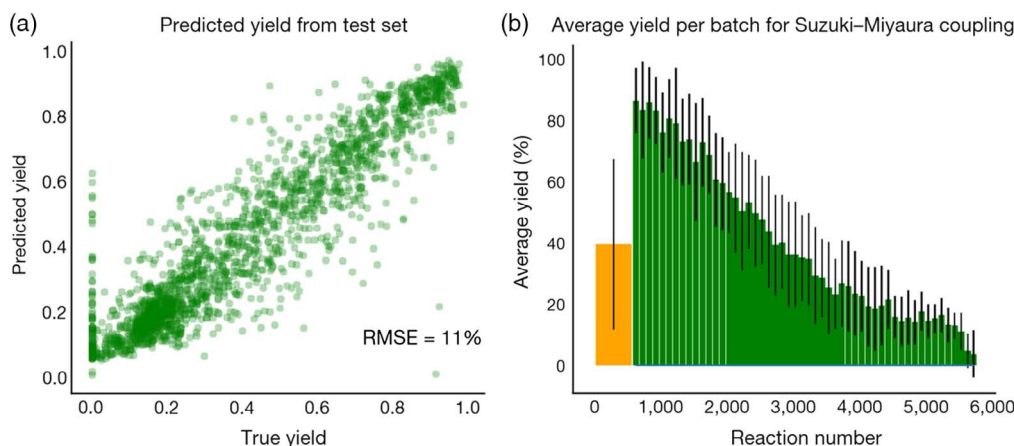
Figure 3. Exploring the Suzuki–Miyaura reaction using ML. a) Validation of the predictive power of the model for a test set of 30% of the reactions (1728 reactions). RMSE, root-mean-square error. b) Simulation of the ML-controlled exploration of this reaction space. The yellow bar shows the initial random choice of 10% of reaction space (576 reactions). The green bars show the next batches of 100 reactions chosen by the ML algorithm. The error bars represent the standard deviation within individual batches for Suzuki–Miyaura coupling. Reproduced with permission.[38] Copyright 2018, Springer Nature.
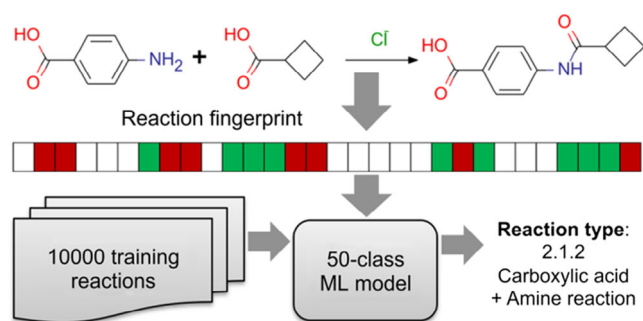

Figure 4. The schematic of ML process for large-scale reaction classification. Reproduced with permission.[76] Copyright 2014, American Chemical Society.

are combined with Monte Carlo tree search (MCTS) to form a new AI algorithm (3N-MCTS) to find the appropriate inverse synthesis route. Three kinds of neural networks are applied to the expansion and display of search nodes (**Figure 5**). Researchers trained these networks using chemical reactions recorded in the Reaxys database before 2015, validated and tested the models using records published after 2015, and finally successfully planned new chemical synthesis routes. In subsequent double-blind experiments, 45 organic synthesizers try to choose synthetic routes for nine complex molecules. 57% of the staff chose the route of 3N-MCTS design and 43% chose the route of literature report. This suggests that even authoritative synthetic chemists find it difficult to distinguish between the software and human chemists. Compared with the traditional synthesis methods, more synthetic routes can be predicted in a shorter time using the new AI technology. This research is a breakthrough in AI applied for chemical synthesis. Mark Waller has also been hailed as the pioneer of "chemical AlphaGo" by the media.
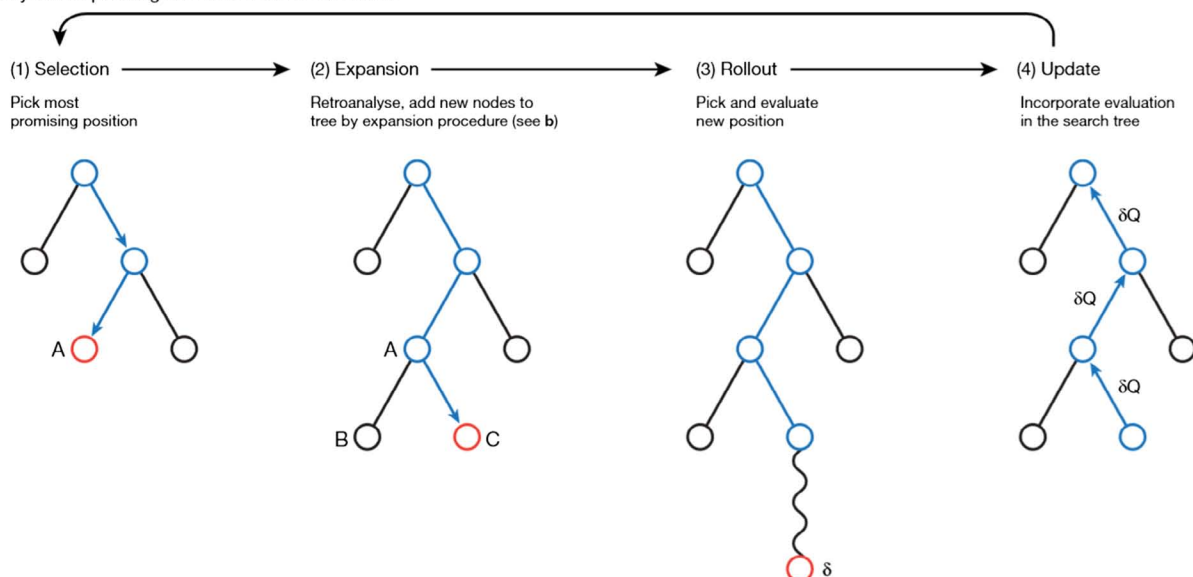
With the aid of simulation calculation and material informatics, the design and performance prediction of new materials can

be completed. However, finding ways to predict the synthesis method of these new materials is the bottleneck in the current material research. Researchers usually need months or even years of repeated trial-and-error experiments to get a mature synthesis method of new compounds, and the corresponding experimental parameters and results varying with the environment will also bring difficulties for wider learning and application. The establishment of material synthesis information database is an important step to overcome this bottleneck.

Kim et al. collaborated to obtain synthetic conditions from published literature using ML and natural language processing techniques.[78] AI platform developed by researchers can automatically analyze literature, and classify them according to the keywords mentioned in the text, such as synthesis temperature, time, equipment name, preparation conditions, and target materials. The results show that the platform has 99% accuracy in identifying passages and 86% accuracy in tagging keywords. Using this platform, the researchers analyzed the synthesis conditions of various metal oxides in 12 900 pieces of literature, and successfully predicted the key parameters needed for hydro-thermal synthesis of titanium dioxide nanotubes based on the obtained data. This technology is an important progress in the Material Genome Project. It is expected to greatly reduce the difficulty in developing new materials and save the time of developing new materials.

Subsequently, Huo et al. constructed a semi-supervised ML method, which was used to obtain and classify inorganic material synthesis information in batches from natural language documents.[79] First, they use the unsupervised algorithm, latent Dirichlet allocation (LDA) model to divide keywords into themes corresponding to specific synthesis steps. They extract information about synthesis methods and steps of materials from more than 2.2 million published documents, such as "grinding", "heating", "dissolution" and "centrifugation". After adding a small number of annotations, the random forest classifier can be associated and divided into different kinds, such as solid-state,
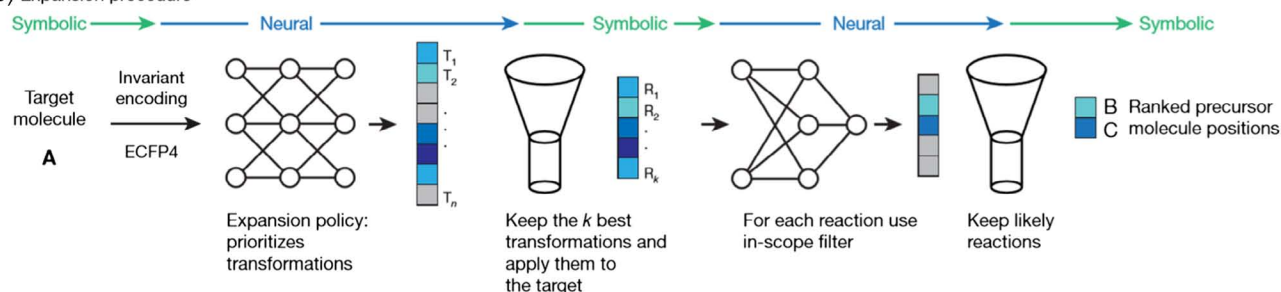
Figure 5. Schematic of MCTS methodology. a) MCTS searches by iterating over four phases. In the selection phase (1), the most urgent node for analysis is chosen on the basis of the current position values. In phase (2), this node may be expanded by processing the molecules of the position A with the expansion procedure (b), which leads to new positions B and C, which are added to the tree. Then, the most promising new position is chosen, and a rollout phase (3) is performed by randomly sampling transformations from the rollout policy until all molecules are solved or a certain depth is exceeded. In the update phase (4), the position values are updated in the current branch to reflect the result of the rollout. b) Expansion procedure. First, the molecule (A) to retroanalyze is converted to a fingerprint and fed into the policy network, which returns a probability distribution over all possible transformations (T_1 to T_n). Then, only the k most probable transformations are applied to molecule A. This yields the reactants necessary to make A, and thus complete reactions R_1 to R_k. For each reaction, the reaction prediction is performed using the in-scope filter, returning a probability score. Improbable reactions are then filtered out, which leads to the list of admissible actions and corresponding precursor positions B and C. Reproduced with permission.[77] Copyright 2018, Springer Nature.

hydrothermal, sol–gel synthesis, and so on. Finally, the flowchart of the possible synthesis process is accurately reconstructed using the Markov chain representation of the order of the experimental steps. The research shows that ML method can not only classify the synthetic process of materials accurately but also reconstruct the synthetic route map of materials, and present the results in a human-readable standardized way, which can be further used to build the synthetic process database.

One of the key challenges in guiding experiments to materials with required properties is finding ways to navigate effectively in a wide composition and structure space. Yuan et al. applied the active learning algorithm, one of the ML methods, to effectively select the sample components to be synthesized and tested in the next step of experiments by exploiting the training data.[52] Only through five iterations, the piezoelectric $(Ba_{0.84}Ca_{0.16})(Ti_{0.90}Zr_{0.07}Sn_{0.03})O_3$ with the largest electrostrain

of 0.23% was synthesized. They also compared four different experimental strategies and found that the strategy of balancing exploration (using uncertainty) and exploitation (only using model prediction) is more efficient in experimental design. This idea can be widely used in the research of new materials.

There is a Chinese proverb, "Failure is the mother of success". Each failure brings researchers one step closer to success. Raccuglia et al. trained ML models using data from unsuccessful hydrothermal reactions in the laboratory, and used the models to predict new reactions.[80] The models were able to successfully predict the synthetic conditions of new organic–inorganic materials with a success rate of 89%. Literature published by researchers in the field of chemistry usually only include examples of successful reactions, but in fact, a large number of unreported failed experiments also contain information about synthetic conditions. The information contained in

these failed experiments is also of great value in predicting the boundary conditions of successful and failed reactions. A large number of laboratory failure reaction data were collected. An SVM model was trained to predict the reaction results of the test set. The accuracy of the model was 78% and the prediction of the reaction of vanadium-selenite system was achieved. The accuracy was 79%. By transforming the SVM model into a decision tree model for human understanding, we can further understand the mechanism of the reaction and guide the new synthetic reaction.

3.4. Experimental Parameter Optimization

In traditional material developments, a large number of parameters need to be analyzed and adjusted manually in synthesis, processing, and device assembly processes. The efficiency is very low and may not be able to find the optimal parameters. ML has powerful nonlinear regression ability to find the best location in the huge parameter space.[81]

This idea has been applied in the welding process. Friction stir welding (FSW) is a relatively new solid-state welding process, which has been widely used in aerospace, shipbuilding, automobile, and other industries. Du et al. collected 108 independent experimental data from authoritative literature to train ML models, including neural networks and decision trees, and explored the effects of original welding parameters such as temperature, maximum shear stress on tool pins, torque and strain rate, and potential causative variables on void formation.[82] The results show that the two algorithms can predict the formation of defects well, and the highest prediction accuracy is 96.6%. With this model, the optimization of parameters in the welding process can be completed, and the formation of unfavorable factors such as void formation in FSW from ML can be avoided.

Similar examples have been applied in 3D printing. Aerosol jet printing (AJP) is a noncontact 3D printing technology, which is often used to fabricate microelectronic devices on flexible substrates. It has the deposition ability of special patterns, but the complex relationship between the main process parameters is complex, and it will have a significant impact on the printing quality. Zhang et al. proposed a new hybrid ML method to determine the best operating process window of AJP process in different design spaces.[83] This method consists of classical ML methods, including experimental sampling, data clustering, classification, and knowledge transfer. The method is based on the Latin hypercube sampling experiment design, and the 2D design space is fully explored at a certain printing speed. Then, the influence of sheath gas flow rate (SHGFR) and carrier gas flow rate (CGFR) on the quality of printing line was analyzed by K-means clustering method, and the optimal operation process window was determined by support vector machine (**Figure 6**). To effectively identify more operation process windows at different printing speeds, the transfer learning method is used to make use of the correlation between different operation process windows. Therefore, under the new printing speed, the number of row samples used to identify the new operation process window is greatly reduced. Finally, to balance the complex relationship between SHGFR, CGFR and printing speed, an incremental classification method is used to determine a 3D operation process window. Unlike the experiment-based quality optimization method in 3D printing technology, this method is developed based on knowledge discovery and data mining theory. Therefore, the knowledge of different design spaces can be fully excavated and transmitted to optimize printing line quality.

In the future, when the material synthesis process is fully automated, it will be integrated with industrial manufacturing 4.0, such as programmable high-throughput synthesis platform for polymers.[84] In the early stage of this high-throughput synthesis, ML is needed to explore the parameter space to determine how the ratio of raw materials and the rate of catalyst supply can be used to synthesize ideal organic compounds with appropriate molecular weight, narrow distribution, and few side reactions.
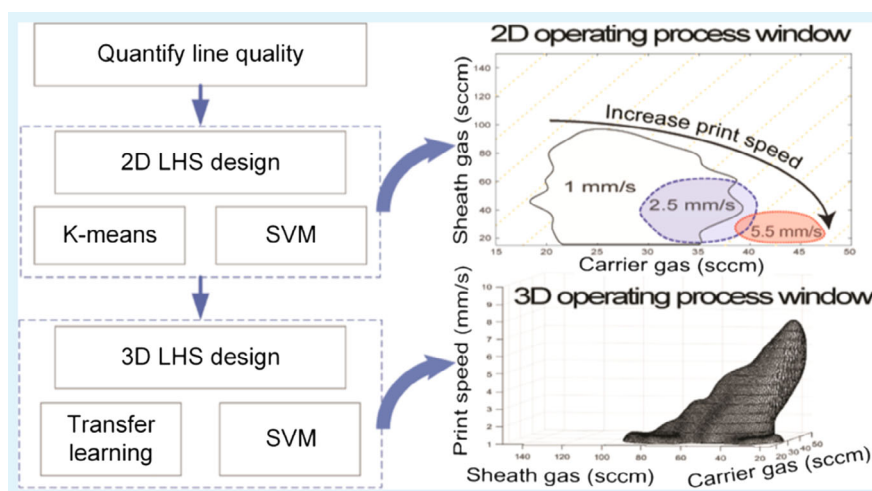

Figure 6. Schematic of process of printing parameters optimization via hybrid ML method. Reproduced with permission.[83] Copyright 2019, American Chemical Society.

ADVANCED
SCIENCE NEWS
www.advancedsciencenews.com

ADVANCED
INTELLIGENT
SYSTEMS
Open Access
www.advintellsyst.com

3.5. Upgrading of Characterization Methods

The great advances in materials science since the last century have been largely due to advances in representational methods, which have enabled scientists to observe atomic-level structures and track atomic-level movements, thus discovering more laws of materials science. With the development of Material Genome Project, high-throughput materials preparation and analysis with AI will become inevitable.[85–88]

The successful application of convolutional neural networks in deep learning has made great achievements in image recognition.[89] This pattern-recognition ability can be easily transferred to the image characterization of micromaterials. Electron microscopy and defect analysis are the cornerstones of material science because they provide detailed insights into the microstructures and properties of various materials and material systems. If a powerful and flexible platform is established for automatic defect recognition and classification in electron microscopy, the analysis can be completed more quickly after image recording and even during image acquisition. However, a large number of images are needed to extract statistically significant information, and recognition is still done manually, which is not only time-consuming but also inconsistent. Recently, Li et al. obtained information about the size and type of defects by combining ML, computer vision, and image analysis techniques (**Figure 7**).[90] At present, the performance of the program is consistent with the manual analysis of quality. Further improvement in the program can make real-time analysis of large datasets.

X-ray diffraction (XRD) data can also be analyzed by ML.[91] In the face of large-scale measurement data with high-throughput characterization, it will undoubtedly consume a lot of time and energy if we analyze them one by one and find sample data of interest from them. ML can help researchers improve the efficiency of analysis and discover hidden rules in data.

By depositing ternary Fe—Ga—Pd compound films on a single silicon wafer, Long et al. obtained 535 samples of the size of 1.75×1.75 mm^2 with continuously changing ternary Fe—Ga—Pd composition.[92] The diffraction data of 273 samples were obtained by XRD characterization. Then, with the help of ML, 273 XRD sample data are clustered by hierarchical clustering algorithm in unsupervised learning, and single-phase samples are merged into the same cluster as far as possible. Only representative sample data in each cluster are analyzed, which greatly improves the efficiency of analysis. The aforementioned results show that dimensionality reduction and clustering algorithm in ML can help to efficiently analyze high-throughput XRD data, identify the phase distribution and the intersection of different phases, and help researchers quickly find regions of interest.

The capacity of lithium-ion batteries decreases with the increase inf the times of cycles. The cycle life of batteries has always been one of the most concerned performances of battery researchers. Severson et al. have developed a new large data-driven model.[93] Without analyzing the mechanism of battery decay, the ability to use neural networks to explore the law of high-dimensional data can predict the whole life of commercial lithium iron phosphate/graphite batteries only by using the charge and discharge data of the first few cycles. In the regression setup, the author uses the first 100 cycles, and the prediction error is only 9.1%. In the classification setup, the author uses the data of the first five cycles, and the prediction error is only 4.9%, which achieves the accurate prediction. This brings new opportunities for battery production, cascade utilization and optimization. For example, battery manufacturers can accelerate battery development cycles, quickly validate new manufacturing processes, and classify new batteries according to their life expectancy. Similarly, consumers can estimate the life expectancy of batteries in their electronic products. Generally speaking, the work emphasizes the combination of data generation and data-driven modeling, which has broad prospects in understanding and developing complex systems such as lithium-ion batteries.
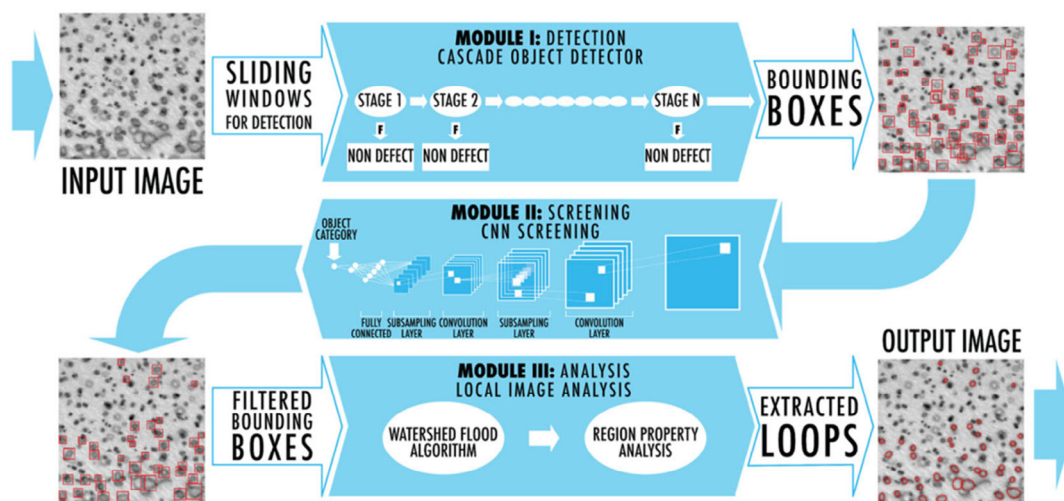

Figure 7. Schematic flowchart of the proposed automated detection approach. Input micrographic images go through the pipeline of module I—Cascade Object Detector, module II—CNN Screening, and module III—Local Image Analysis. After module I, the loop locations and bounding boxes are identified and then further refined to remove false positives using module II. Then module III determines the loop shape and size. Reproduced with permission.[90] Copyright 2018, Springer Nature.

ML can also help researchers get rid of the confusion in impedance data analysis. Electrochemical impedance spectroscopy (EIS) is a very powerful method in the research and diagnosis of electrochemical batteries and future electrochemical energy storage systems. However, it is quite difficult to analyze a large number of EIS data. Typical optimization algorithms are not complete. In practice, it means that researchers must accurately construct the equivalent circuit (EC) model, select the appropriate initial values of the parameters of each component of the model, and constantly verify the output in the process to ensure the correct convergence of the fitting. Buteau and Dahn proposed an inverse model of ML, which transformed 100 000 independent fitting optimization problems into a single optimization problem.[94] The error rate of solving a single optimization problem was less than 1% by applying various viewpoints in ML literature. If an open-source system is assembled for EIS test, it can be easily adapted to various impedance spectrograms, and the parameters of the physical model can be reliably fitted to the measured data. This method has high reliability, good consistency, and no need of manual supervision. The code used in this work can be obtained at -https://github.com/samuelbuteau/eisfitting.

At present, material science research has been self-derided as "stir-fried dishes". It adds salt and water, and discovers new materials through trial and error. By ML and high-throughput computing, material scientists can speed up the efficiency of trial and error and save labor.

In the future, the development of material AI may require some free open-source software platform, which combines the functions of AI data analysis with the appropriate operating interface. AI could track each scientific research topic and provide possible alternative analysis solutions for the problems in representation. Researchers can also upload their own experimental process and corresponding results, so as to facilitate everyone to solve and think about the experimental difficulties.

In conclusion, AI will not completely replace synthetic chemists. Synthetic chemists will discover new reactions in practical scientific research and expand the theoretical basis of chemistry, but AI will certainly become a powerful assistant to chemists to help them find synthetic routes faster and better. Supported by existing experimental data and theoretical basis, combined with ML technology, AI-aided material design, synthesis, characterization, and application research will greatly promote the research efficiency of scientists in the field of materials and help the rapid development of material science.

4. Prospects and Future

AI is making more and more contributions in materials research.[95–100] This article reviews the representative research progress of materials AI including the realization details and advantages over conventional methods. In general, the future development of material informatics requires high-throughput experiments, high-throughput simulation calculations, and high-throughput characterization. The following will be the outlook from both software and hardware aspects.

4.1. Algorithm Upgradation

ML is data analysis (statistical method) and the required data pursues quantity, comprehensiveness, and objectivity. Previous studies of material informatics were limited by the computed properties without enough accuracy. The datasets composed of more accurate experimental results will make a big difference. However, the current experimental samples is uncomprehensive because of the excessive centralization of hot research spots. Fortunately, some models are suitable for dealing with small datasets such as autoencoders, generative adversarial networks, active learning and transfer learning.

In addition, ML models need to be translated into actual knowledge or physical pictures to avoid the "Black Box" characteristic. Calculating the average of neurons that respond to the descriptors could provide certain interpretation. Or more explanatory models, such as decision trees which can reflect the impact of relevant factors by the weights of nodes and branches of the tree, could be applied to boost the development of materials informatics.

4.2. Infrastructure Construction

Effective training of ML models usually requires abundant data. Such data could come from online databases, published papers, or high-throughput experimental equipment.

Online databases are a trend for the application of deep learning, such as ImageNet. The development of material informatics also need similar platforms. For example, Hatakeyama-Sato et al. built up a database to accumulate the information of electrolytes, including ionic conductivity, transference number, and chemical stability.[101] Published articles also contain vast materials data. Researchers can search for desired information easily by natural-language-processing technology once these papers are arranged by standardized article formats.

More sensors and software can be integrated into the high-throughput synthesis or characterization equipment. The results collected by these equipment are directly fed back to AI models for the optimization of experimental parameters. Then, the samples with ideal properties can be obtained by adjusting the parameters. Materials informatics will finally map the relationship between "composition-structure-property-processing-application" through these efforts.

AI will not completely replace humans at the work of material research but will serve as a powerful tool to accelerate the progress of materials discovery. We material researchers all need to learn to master this tool to decrease the trial error times, solve more difficult material problems in more fields, and find more rules that govern the nature we live.

Conflict of Interest

The authors declare no conflict of interest.

Keywords

[1] K. Rajan, *Mater. Today* **2012**, *15*, 470.
[2] A. F. Zahrt, J. J. Henle, B. T. Rose, Y. Wang, W. T. Darrow, S. E. Denmark, *Science* **2019**, *363*, eaau5631.
[3] Y. Liu, T. Zhao, W. Ju, S. Shi, *J. Mater.* **2017**, *3*, 159.
[4] W. Lu, R. Xiao, J. Yang, H. Li, W. Zhang, *J. Mater.* **2017**, *3*, 191.
[5] R. R. Kline, *IEEE Ann. Hist. Comput.* **2011**, *33*, 5.
[6] W. S. McCulloch, W. Pitts, *Bull. Math. Biol.* **1943**, *5*, 115.
[7] D. T. Tran, S. Kiranyaz, M. Gabbouj, A. Iosifidis, *IEEE Trans. Neural Networks Learn. Syst.* **2019**, *31*, 710.
[8] C. J. C. Burges, *Data Min. Knowl. Discovery* **1998**, *2*, 121.
[9] S. K. Pal, S. Mitra, *IEEE Trans. Neural Networks* **1992**, *3*, 683.
[10] M. Uccellari, F. Facchini, M. Sola, E. Sirignano, G. M. Vitetta, A. Barbieri, S. Tondelli, *IET Microwaves Antennas Propag.* **2017**, *12*, 302.
[11] D. E. Rumelhart, G. E. Hinton, R. J. Williams, *Nature* **1986**, *323*, 533.
[12] G. Hinton, L. Deng, D. Yu, G. Dahl, A.-R. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. Sainath, B. Kingsbury, *IEEE Signal Process. Mag.* **2012**, *29*, 82.
[13] Y. LeCun, Y. Bengio, G. Hinton, *Nature* **2015**, *521*, 436.
[14] P. Bory, *Convergence Int. J. Res. New Media Technol.* **2019**, *25*, 627.
[15] D. Silver, J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A. Guez, T. Hubert, L. Baker, M. Lai, A. Bolton, Y. Chen, T. Lillicrap, F. Hui, L. Sifre, G. van den Driessche, T. Graepel, D. Hassabis, *Nature* **2017**, *550*, 354.
[16] A. K. Baughman, W. Chuang, K. R. Dixon, Z. Benz, J. Basilico, *IEEE Trans. Comput. Intell. AI* **2014**, *6*, 55.
[17] N. T. BrownSandholm, T. Sandholm, *Science* **2019**, *365*, 885.
[18] J. Pei, L. Deng, S. Song, M. Zhao, Y. Zhang, S. Wu, G. Wang, Z. Zou, Z. Wu, W. He, F. Chen, N. Deng, S. Wu, Y. Wang, Y. Wu, Z. Yang, C. Ma, G. Li, W. Han, H. Li, H. Wu, R. Zhao, Y. Xie, L. Shi, *Nature* **2019**, *572*, 106.
[19] Y. Yao, X. Li, X. Liu, P. Liu, Z. Liang, J. Zhang, K. Mai, *Int. J. Geogr. Inf. Sci.* **2016**, *31*, 825.
[20] A. W. Senior, R. Evans, J. Jumper, J. Kirkpatrick, L. Sifre, T. Green, C. Qin, A. Zidek, A. W. R. Nelson, A. Bridgland, H. Penedones, S. Petersen, K. Simonyan, S. Crossan, P. Kohli, D. T. Jones, D. Silver, K. Kavukcuoglu, D. Hassabis, *Nature* **2020**, *577*, 706.
[21] M. Popova, O. Isayev, A. Tropsha, *Sci. Adv.* **2018**, *4*, eaap7885.
[22] D. Zhang, R. Cao, S. Wu, *Inform. Fusion* **2019**, *52*, 268.
[23] Y. Liu, F. Han, F. Li, Y. Zhao, M. Chen, Z. Xu, X. Zheng, H. Hu, J. Yao, T. Guo, W. Lin, Y. Zheng, B. You, P. Liu, Y. Li, L. Qian, *Nat. Commun.* **2019**, *10*, 2409.
[24] T. Zhou, Z. Song, K. Sundmacher, *Engineering* **2019**, *5*, 595.
[25] K. K. Yang, Z. Wu, F. H. Arnold, *Nat. Methods* **2019**, *16*, 687.
[26] J. Wei, X. Chu, X. Y. Sun, K. Xu, H. X. Deng, J. Chen, Z. Wei, M. Lei, *InfoMat* **2019**, *1*, 338.
[27] R. Jose, S. Ramakrishna, *Appl. Mater. Today* **2018**, *10*, 127.
[28] A. Agrawal, A. Choudhary, *APL Mater.* **2016**, *4*, 053208.
[29] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. van den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. Lillicrap, M. Leach, K. Kavukcuoglu, T. Graepel, D. Hassabis, *Nature* **2016**, *529*, 484.
[30] K. Rajan, *Mater. Today* **2005**, *8*, 38.
[31] A. Zakutayev, N. Wunder, M. Schwarting, J. D. Perkins, R. White, K. Munch, W. Tumas, C. Phillips, *Sci. Data* **2018**, *5*, 180053.
[32] X. Xu, Y. Lei, Z. Li, *IEEE Trans. Ind. Electron.* **2020**, *67*, 2326.
[33] X. Shen, Z. J. Zhu, *Bioinformatics* **2019**, *35*, 2870.
[34] G. Delaporte, M. Cladière, V. Camel, *Chemom. Intell. Lab. Syst.* **2019**, *188*, 54.
[35] J. Yang, K. K. Tan, M. Santamouris, S. E. Lee, *Buildings* **2019**, *9*, 204.
[36] Q. Xu, Z. Li, M. Liu, W. J. Yin, *J. Phys. Chem. Lett.* **2018**, *9*, 6948.
[37] P. Li, C. Dai, W. Wang, *Symmetry* **2019**, *11*, 575.
[38] J. M. Granda, L. Donina, V. Dragone, D. L. Long, L. Cronin, *Nature* **2018**, *559*, 377.
[39] S. Bermejo, J. Cabestany, *Pattern Recognit.* **1999**, *32*, 2077.
[40] Y. Xia, C. Liu, Y. Li, N. Liu, *Expert Syst. Appl.* **2017**, *78*, 225.
[41] Y. Wang, N. Wagner, J. M. Rondinelli, *MRS Commun.* **2019**, *9*, 793.
[42] E. J. Vladislavleva, G. F. Smits, D. den Hertog, *IEEE Trans. Evol. Comput.* **2009**, *13*, 333.
[43] Y. Cao, V. Fatemi, S. Fang, K. Watanabe, T. Taniguchi, E. Kaxiras, P. Jarillo-Herrero, *Nature* **2018**, *556*, 43.
[44] C. D. Muzny, M. L. Huber, A. F. Kazakov, *J. Chem. Eng. Data* **2013**, *58*, 969.
[45] B. Weng, R. Zhu, Q. Yan, Q. Sun, C. G. Grice, Y. Yan, W. J. Yin, **2019**, https://arxiv.org/abs/1908.06778.
[46] J. J. Möller, W. Körner, G. Krugel, D. F. Urban, C. Elsässer, *Acta Mater.* **2018**, *153*, 53.
[47] P. V. Balachandran, B. Kowalski, A. Sehirlioglu, T. Lookman, *Nat. Commun.* **2018**, *9*, 1668.
[48] N. Artrith, A. M. Kolpak, *Nano Lett.* **2014**, *14*, 2670.
[49] Y. Tan, H. Matsui, N. Ishiguro, T. Uruga, D.-N. Nguyen, O. Sekizawa, T. Sakata, N. Maejima, K. Higashi, H. C. Dam, M. Tada, *J. Phys. Chem. C* **2019**, *123*, 18844.
[50] J. Timoshenko, C. J. Wrasman, M. Luneau, T. Shirman, M. Cargnello, S. R. Bare, J. Aizenberg, C. M. Friend, A. I. Frenkel, *Nano Lett.* **2019**, *19*, 520.
[51] C. Kim, A. Chandrasekaran, A. Jha, R. Ramprasad, *MRS Commun.* **2019**, *9*, 866.
[52] R. Yuan, Z. Liu, P. V. Balachandran, D. Xue, Y. Zhou, X. Ding, J. Sun, D. Xue, T. Lookman, *Adv. Mater.* **2018**, *30*, 1702884.
[53] V. Stanev, C. Oses, A. G. Kusne, E. Rodriguez, J. Paglione, S. Curtarolo, I. Takeuchi, *npj Comput. Mater.* **2018**, *4*, 29.
[54] O. Isayev, C. Oses, C. Toher, E. Gossett, S. Curtarolo, A. Tropsha, *Nat. Commun.* **2017**, *8*, 15679.
[55] M. Schmidt, H. Lipson, *Science* **2009**, *324*, 81.
[56] H. Salmenjoki, M. J. Alava, L. Laurson, *Nat. Commun.* **2018**, *9*, 5307.
[57] X. He, Y. Zhu, Y. Mo, *Nat. Commun.* **2017**, *8*, 15893.
[58] Q. Wang, J. F. Wu, Z. Lu, F. Ciucci, W. K. Pang, X. Guo, *Adv. Funct. Mater.* **2019**, *29*, 1904232.
[59] F. Brockherde, L. Vogt, L. Li, M. E. Tuckerman, K. Burke, K. R. Muller, *Nat. Commun.* **2017**, *8*, 872.
[60] V. L. Deringer, M. A. Caro, G. Csanyi, *Adv. Mater.* **2019**, *31*, 1902765.
[61] Q. Zhou, P. Tang, S. Liu, J. Pan, Q. Yan, S. C. Zhang, *Proc. Natl. Acad. Sci.* **2018**, *115*, E6411.
[62] N. Artrith, B. Hiller, J. Behler, *Phys. Status Solidi B* **2013**, *250*, 1191.
[63] G. Panapitiya, G. Avendano-Franco, P. Ren, X. Wen, Y. Li, J. P. Lewis, *J. Am. Chem. Soc.* **2018**, *140*, 17508.
[64] S. Lu, Q. Zhou, Y. Ouyang, Y. Guo, Q. Li, J. Wang, *Nat. Commun.* **2018**, *9*, 3405.

**ADVANCED
SCIENCE NEWS**

www.advancedsciencenews.com

**ADVANCED
INTELLIGENT
SYSTEMS**
Open Access

www.advintellsyst.com

[65] H. Sahu, W. Rao, A. Troisi, H. Ma, *Adv. Energy Mater.* **2018**, *8*, 1801032.

[66] K. Fujimura, A. Seko, Y. Koyama, A. Kuwabara, I. Kishida, K. Shitara, C. A. J. Fisher, H. Moriwake, I. Tanaka, *Adv. Energy Mater.* **2013**, *3*, 980.

[67] A. D. Sendek, E. D. Cubuk, E. R. Antoniuk, G. Cheon, Y. Cui, E. J. Reed, *Chem. Mater.* **2018**, *31*, 342.

[68] Z. Li, Q. Xu, Q. Sun, Z. Hou, W.-J. Yin, *Adv. Funct. Mater.* **2019**, *29*, 1807280.

[69] M. Sun, T. Wu, Y. Xue, A. W. Dougherty, B. Huang, Y. Li, C.-H. Yan, *Nano Energy* **2019**, *62*, 754.

[70] Y. Zhuo, A. Mansouri Tehrani, A. O. Oliynyk, A. C. Duke, J. Brgoch, *Nat. Commun.* **2018**, *9*, 4377.

[71] R. Gomez-Bombarelli, J. Aguilera-Iparraguirre, T. D. Hirzel, D. Duvenaud, D. Maclaurin, M. A. Blood-Forsythe, H. S. Chae, M. Einzinger, D. G. Ha, T. Wu, G. Markopoulos, S. Jeon, H. Kang, H. Miyazaki, M. Numata, S. Kim, W. Huang, S. I. Hong, M. Baldo, R. P. Adams, A. Aspuru-Guzik, *Nat. Mater.* **2016**, *15*, 1120.

[72] B. Sanchez-Lengeling, A. Aspuru-Guzik, *Science* **2018**, *361*, 360.

[73] T. Xie, A. France-Lanord, Y. Wang, Y. Shao-Horn, J. C. Grossman, *Nat. Commun.* **2019**, *10*, 2667.

[74] B. A. Grzybowski, K. J. Bishop, B. Kowalczyk, C. E. Wilmer, *Nat. Chem.* **2009**, *1*, 31.

[75] A. F. de Almeida, R. Moreira, T. Rodrigues, *Nat. Rev. Chem.* **2019**, *3*, 589.

[76] N. Schneider, D. M. Lowe, R. A. Sayle, G. A. Landrum, *J. Chem. Inf. Model.* **2015**, *55*, 39.

[77] M. H. S. Segler, M. Preuss, M. P. Waller, *Nature* **2018**, *555*, 604.

[78] E. Kim, K. Huang, A. Saunders, A. McCallum, G. Ceder, E. Olivetti, *Chem. Mater.* **2017**, *29*, 9436.

[79] H. Huo, Z. Rong, O. Kononova, W. Sun, T. Botari, T. He, V. Tshitoyan, G. Ceder, *npj Comput. Mater.* **2019**, *5*, 62.

[80] P. Raccuglia, K. C. Elbert, P. D. Adler, C. Falk, M. B. Wenny, A. Mollo, M. Zeller, S. A. Friedler, J. Schrier, A. J. Norquist, *Nature* **2016**, *533*, 73.

[81] P. M. Attia, A. Grover, N. Jin, K. A. Severson, T. M. Markov, Y. H. Liao, M. H. Chen, B. Cheong, N. Perkins, Z. Yang, P. K. Herring, M. Aykol, S. J. Harris, R. D. Braatz, S. Ermon, W. C. Chueh, *Nature* **2020**, *578*, 397.

[82] Y. Du, T. Mukherjee, T. DebRoy, *npj Comput. Mater.* **2019**, *5*, 68.

[83] H. Zhang, S. K. Moon, T. H. Ngo, *ACS Appl. Mater. Interfaces* **2019**, *11*, 17994.

[84] B. Lin, J. L. Hedrick, N. H. Park, R. M. Waymouth, *J. Am. Chem. Soc.* **2019**, *141*, 8921.

[85] Y. T. Wang, B. Li, X. J. Xu, H. B. Ren, J. Y. Yin, H. Zhu, Y. H. Zhang, *Food Chem.* **2020**, *303*, 125404.

[86] S. Kiyohara, T. Miyata, K. Tsuda, T. Mizoguchi, *Sci. Rep.* **2018**, *8*, 13548.

[87] A. Maksov, O. Dyck, K. Wang, K. Xiao, D. B. Geohegan, B. G. Sumpter, R. K. Vasudevan, S. Jesse, S. V. Kalinin, M. Ziatdinov, *npj Comput. Mater.* **2019**, *5*, 12.

[88] M. Ziatdinov, A. Maksov, S. V. Kalinin, *npj Comput. Mater.* **2017**, *3*, 1.

[89] A. Krizhevsky, I. Sutskever, G. E. Hinton, *Commun. ACM* **2017**, *60*, 84.

[90] W. Li, K. G. Field, D. Morgan, *npj Comput. Mater.* **2018**, *4*, 36.

[91] A. Sanchez-Gonzalez, P. Micaelli, C. Olivier, T. R. Barillot, M. Ilchen, A. A. Lutman, A. Marinelli, T. Maxwell, A. Achner, M. Agaker, N. Berrah, C. Bostedt, J. D. Bozek, J. Buck, P. H. Bucksbaum, S. C. Montero, B. Cooper, J. P. Cryan, M. Dong, R. Feifel, L. J. Frasinski, H. Fukuzawa, A. Galler, G. Hartmann, N. Hartmann, W. Helml, A. S. Johnson, A. Knie, A. O. Lindahl, J. Liu, et al., *Nat. Commun.* **2017**, *8*, 15461.

[92] C. J. Long, J. Hattrick-Simpers, M. Murakami, R. C. Srivastava, I. Takeuchi, V. L. Karen, X. Li, *Rev. Sci. Instrum.* **2007**, *78*, 072217.

[93] K. A. Severson, P. M. Attia, N. Jin, N. Perkins, B. Jiang, Z. Yang, M. H. Chen, M. Aykol, P. K. Herring, D. Fraggedakis, M. Z. Bazant, S. J. Harris, W. C. Chueh, R. D. Braatz, *Nat. Energy* **2019**, *4*, 383.

[94] S. Buteau, J. R. Dahn, *J. Electrochem. Soc.* **2019**, *166*, A1611.

[95] Y. Mao, X. Wang, S. Xia, K. Zhang, C. Wei, S. Bak, Z. Shadike, X. Liu, Y. Yang, R. Xu, P. Pianetta, S. Ermon, E. Stavitski, K. Zhao, Z. Xu, F. Lin, X. Q. Yang, E. Hu, Y. Liu, *Adv. Funct. Mater.* **2019**, *29*, 1900247.

[96] Z. Li, Z. Zhang, J. Shi, D. Wu, *Rob. Comput.-Integr. Manuf.* **2019**, *57*, 488.

[97] W. Li, J. Zhu, Y. Xia, M. B. Gorji, T. Wierzbicki, *Joule* **2019**, *3*, 2279.

[98] M. X. Li, S. F. Zhao, Z. Lu, A. Hirata, P. Wen, H. Y. Bai, M. Chen, J. Schroers, Y. Liu, W. H. Wang, *Nature* **2019**, *569*, 99.

[99] R. P. Joshi, J. Eickholt, L. Li, M. Fornari, V. Barone, J. E. Peralta, *ACS Appl. Mater. Interfaces* **2019**, *11*, 18494.

[100] S. Honrao, B. E. Anthonio, R. Ramanathan, J. J. Gabriel, R. G. Hennig, *Comput. Mater. Sci.* **2019**, *158*, 414.

[101] K. Hatakeyama-Sato, T. Tezuka, M. Umeki, K. Oyaizu, *J. Am. Chem. Soc.* **2020**, *142*, 3301.