ADVANCED SCIENCE Open Access
www.advancedscience.com

Machine Learning in the Development of Adsorbents for Clean Energy Application and Greenhouse Gas Capture

Haoxin Mai, Tu C. Le, Dehong Chen, David A. Winkler,* and Rachel A. Caruso**

Addressing climate change challenges by reducing greenhouse gas levels requires innovative adsorbent materials for clean energy applications. Recent progress in machine learning has stimulated technological breakthroughs in the discovery, design, and deployment of materials with potential for high-performance and low-cost clean energy applications. This review summarizes basic machine learning methods—data collection, featurization, model generation, and model evaluation—and reviews their use in the development of robust adsorbent materials. Key case studies are provided where these methods are used to accelerate adsorbent materials design and discovery, optimize synthesis conditions, and understand complex feature–property relationships. The review provides a concise resource for researchers wishing to use machine learning methods to rapidly develop effective adsorbent materials with a positive impact on the environment.

1. Introduction

The 2016 Paris agreement aimed to limit global warming to less than 2 °C by 2100, an essential target for addressing climate change.[1,2] To achieve this, the research community must develop innovative processes for generating clean energy and reducing greenhouse gas emissions.[3–5] More environmentally friendly gaseous fuels with high energy generation efficiency, such as H_2, are attractive alternatives to nonrenewable fossil fuels.[6,7] Although their gravimetric energy density is excellent, their volumetric energy density under normal temperatures and pressures is low, leading to storage and transport challenges.[8,9] Conventional technologies for storage and transport, like compression and liquefaction, require high pressures or low temperatures.[10,11] The environmental impact of greenhouse gases can be ameliorated by their capture and storage, or conversion to useful chemicals.[12,13] Currently, chemical absorption using amine aqueous solutions is the most common method of absorbing CO_2. However, safety and cost issues related to corrosion, energy consumption, and amine loss remain.[14,15] Therefore, cheaper and safer technologies for storage and transport of gaseous fuels and the large-scale separation and adsorption of CO_2 are a high priority.

Among the solid adsorbents, porous materials such as metal–organic frameworks (MOFs),[9,16,17] covalent–organic frameworks (COFs),[18,19] porous carbons,[20] and zeolites[21] have excellent H_2, CH_4, and CO_2 adsorption abilities, low cost, scalable production, and tunable structural features. However, rapid development of these materials is hampered by two obstacles.[16,22] First, the structural and compositional space of these materials is vast, making full experimental exploration impossible. Second, the relationships between materials features and their desired properties (e.g., uptake capacity and selectivity) are complex and often nonlinear. This can make physics-based computational modeling and experimental characterization to identify the most relevant features complicated, time-consuming and expensive, hindering material innovation. More effective strategies must be developed to shorten discovery timelines for efficient adsorbents.

Data acquired from experiments and high-throughput computation can be used to train machine learning (ML) models of the properties of porous materials that can be used to expedite the design, discovery, and optimization of adsorbents.[23–25] Data-driven ML methods such as neural networks (NN), support vector machines (SVM), and Bayesian methods can uncover

H. Mai, D. Chen, R. A. Caruso
Applied Chemistry and Environmental Science
School of Science
STEM College
RMIT University
Melbourne, Victoria 3001, Australia
E-mail: dehong.chen@rmit.edu.au; rachel.caruso@rmit.edu.au
T. C. Le
School of Engineering
STEM College
RMIT University
GPO Box 2476, Melbourne, Victoria 3001, Australia
D. A. Winkler
Monash Institute of Pharmaceutical Sciences
Monash University
Parkville, VIC 3052, Australia
E-mail: D.winkler@latrobe.edu.au
D. A. Winkler
School of Biochemistry and Chemistry
La Trobe University
Kingsbury Drive, Bundoora 3042, Australia
D. A. Winkler
School of Pharmacy
University of Nottingham
Nottingham NG7 2RD, UK

The ORCID identification number(s) for the author(s) of this article can be found under https://doi.org/10.1002/advs.202203899

DOI: 10.1002/advs.202203899

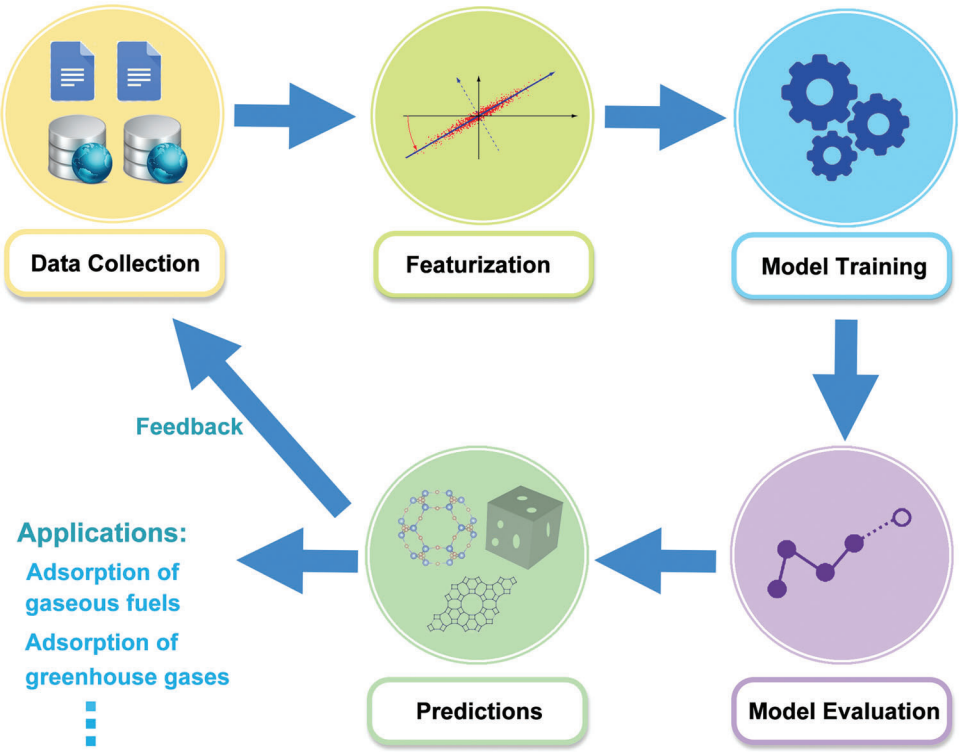


Figure 1. Workflow for applying ML techniques in the development of porous materials for gas adsorption. ML model construction includes data collection, featurization, model training, and model evaluation. The ML models are then used to make predictions. Predicted materials that are subsequently synthesized and assessed will be added to the original dataset for model improvement, and the materials with better properties will be progressed to practical applications.

complex relationships between porous materials structural, physicochemical, and process properties and useful properties such as uptake and selectivity.[26] These models can make quantitative predictions of these properties for new porous materials yet to be synthesized.[27] Given the exponential growth of research data, ML approaches are driving new materials discovery and elucidating complex feature–property relationships in large porous materials chemistry spaces.[28–31]

This review provides a broad introduction to ML methods used for the design (using chemical intuition/skill or computational guidance to generate similar structures), discovery (screening unknown material spaces by high-throughput experiments or computations), and optimization (modifying the structures of lead materials or the synthesis conditions to improve their properties) of functional materials for clean energy applications, such as adsorption and separation of H_2 and CH_4, and reduction of greenhouse gases. Recent reviews that combine data-science and porous materials are available, some covering specific areas (e.g., high-throughput methods, nanotechnology, and materials evolution), in which ML is a part of the strategy.[32,33] Other reviews are substantially ML focused, and potentially less relevant to materials scientists lacking strong data science backgrounds.[24] As many readers may not be familiar with ML techniques, our review summarizes the application of ML to different types of adsorbent materials, introduces the most recent advances in adsorbent materials developed by ML-assisted strategies, and provides readers with a practical guide to selecting ML methods for research on a wide range of adsorbent materials. More extensive, recent reviews on the applications of ML to materials design in general are available for interested readers.[24,34,35] The focus here is on using ML methods to develop high performing, diverse adsorbent materials. Section 2 introduces key steps in applying ML methods to adsorbent materials. Recent examples of adsorbent design and optimization using ML are summarized in Section 3. The final section presents conclusions and provides a perspective on likely progress in this field in the short to medium term. We hope this review will stimulate and encourage the use of ML techniques to accelerate development of effective adsorbents, leading to improved renewable energy technologies.

2. Developing Machine Learning Models

ML uses training data and an appropriate algorithm to model diverse relationships in physical or biological systems.[28,36] It is an empirical alternative to complicated static and dynamic first-principles electronic structure and molecular dynamics calculations and provides insight into nonlinear, multidimensional, feature–property relations. The process of constructing a ML model consists of four steps as follows (**Figure 1**): 1) data collection, 2) feature generation and selection, 3) algorithm selection and mapping, and 4) model validation and prediction.

Data collection is extremely important for material informatics. Data must be reliable, of sufficient volume, low in noise and bias, and the modeled property should have a reasonable

dynamic range of values (i.e., a model cannot be generated if all data points have similar properties).[37]

Data for ML model training include features (from molecular structures, physicochemical properties, experimental conditions of synthesis, etc.) and the target properties. Target properties for materials in the dataset are usually derived from laboratory and computational experiments.[38] Models are most successful when they are trained on data with a wide range of properties—materials with poor properties as well as the successful materials.[39,40] A paucity of large, high-quality datasets has necessitated the use of data collected from the literature, ideally compiled into data repositories.[41–44] These materials databases are important sources of curated data. Examples of databases for adsorbent materials include molecule databases (ChEMBL,[45] GDB-13,[46] GDB-17,[47] ZINC,[48] etc.), inorganic compound databases (Atomic-FLOWLIB,[49] Inorganic Crystal Structure Database,[50] Materials Project,[51] Open Quantum Materials Database, NOMAD,[52] etc.), and databases for specific materials, such as the Cambridge Structural Database MOF subset,[53] the Computation-Ready, Experimental MOF Database,[54] the International Zeolite Association (IZA) database, the Predicted Crystallography Open Database (mainly for silicates, phosphates, sulfates, zeolites, and fluorides),[55] the NIST isotherm database (ISODB),[56] and the Metal Organic Framework Database (MOFDB).[57] Many of these databases contain materials structures, and some record important properties of materials. For example, the Materials Project has DFT-calculated electronic properties available for a large number of materials. NIST ISODB and MOFDB are databases of adsorbents and isotherms.

Once sufficient data are collected, information about the materials must be converted into mathematical representations (descriptors, or features) suitable for training ML models. The quality and relevance of these representations play a major role in the quality of the subsequent model, accuracy of predictions of properties for new materials, and the ability to interpret the models in terms of chemistry (useful for deciding which material to synthesize next). The conversion of relevant attributes of materials into features is called featurization. The performance of ML models is optimum when the features used in training are most relevant to the property being modeled. Featurization is of critical importance to the quality and interpretability of the models generated. Ideally, materials and data scientists should work together to identify the features with the most promise.

A large number of features can be calculated or measured for adsorbent materials, and the importance of these features for a material is strongly context dependent.[28,31,36] For example, experimental features (temperature, pH value, pressure, reaction time, and the amount of the reactants) would be used to construct models for synthesis optimization of adsorbent materials.[58] Topographical features (pore size, volume, surface area, topological shape) and compositional features are frequently used when searching for adsorbent materials with the best texture and composition for gas uptake.[59] Atomic features (e.g., atomic radii, mass, number of valence electrons) and electronic features (e.g., electronegativity, ionization energy, polarizability) are frequently used for selection of the coordinating metals of MOFs.[35,60] For MOF linkers, electronic features, such as bandgap, dipole moment, highest occupied molecular orbital (HOMO), lowest unoccupied molecular orbital (LUMO), and other structural features such as the Coulomb matrix,[61] atom-centered symmetry functions (ACSF),[62] simplified molecular-input line-entry system (SMILES),[63] Voronoi tessellations,[64] and Smooth Overlap of Atomic Positions (SOAP),[65] are used to describe their molecular structures.[35] Distances between neighboring atoms are often used to describe or encode the local structure of adsorbent materials.[66] Some physical coefficients, such as Henry's law constant, which is closely related to the adsorption isotherm, can be used as a feature or target property.[67–69] In addition, energy-based features, including Voronoi energy[70] and energy histogram,[71] are used to improve the model performance in the cases when the training set is small and the data diversity is high. Fanourgakis et al. developed a set of new features related to the MOF energy surfaces by inserting probe atoms of different sizes into the MOF, and found that model predictions were improved by using these features.[72] We suggest that despite computational complexity, structural features and energy-based features play more important roles in gas uptake prediction than electronic and atomic features,[73] while topographical features and composition-based features are frequently used due to their ease of calculation.[74,75] Therefore, there is a trade-off between the computational complexity and model accuracy and interpretability. Combining multiple features provides a more accurate description of materials.

Although many features may correlate with target properties, the number of features must be limited to avoid overfitting and degradation of model predictivity by the presence of features of low relevance (noise).[31] Large numbers of features also increase the complexity of the model, increasing the computation expense, compromising the prediction ability (optimally sparse models have the best predictive abilities), and making model interpretation more difficult.[36] As a rule of thumb, the number of fitted variables in a model should be less than half of the number of the data points, preferably much less.[36] Down-selection and dimensionality reduction are two strategies often used to reduce the number of features. In down-selection, statistical methods, such as the least absolute shrinkage and selection operator (LASSO) or random forest (RF), are used to assess the importance of the features, and the least important features can be discarded from the feature set.[24] However, the performance of the down-selection algorithms depends on the choice of hyperparameters. Dimensionality reduction is the alternative method to shrink the feature set. The principle of this strategy is to project the data points from a high dimensional feature space to a low dimensional feature space. As it is well known that sparser models have better generalization ability and interpretability,[36] some features in these models may be highly correlated or do not strongly relate to the target property. During dimensionality reduction, these features are combined, and new features are generated. Therefore, relevant feature information is retained and only the redundant information is lost when an appropriate dimensionality reduction algorithm is used. The ability of models to predict the properties of new, unseen materials will be enhanced after such dimensionality reduction, thus the performance of the dimensionality reduction algorithm can be estimated by model evaluation. The most popular linear dimensionality reduction algorithms are principal component analysis (PCA) and linear discriminant analysis (LDA).[76–78] The PCA method generates orthogonal features that decrease or remove correlations.[79] This

is an unsupervised projection algorithm, computing a group of orthogonal vectors (principal component) as new features, where the data point variances are maximum. In contrast to PCA, LDA looks for the orthogonal vectors on which the variances of the data points among different classes are maximum. Both PCA and LDA can simplify ML models and remove the issue of dimensionality, but they suffer from the assumption that the relationship between features and modeled property is linear, which is often not the case. Nonlinear projection algorithms, such as Isomap, Local Linear Embedding, Laplacian Eigenmaps, t-distributed stochastic neighbor embedding (t-SNE), and uniform manifold approximation and projection (UMAP) can perform nonlinear dimensionality reduction.[80,81] Among these algorithms, t-SNE and UMAP have been extensively applied to data visualization and evaluating the domain of applicability of ML models on different datasets. Note that less relevant information is retained after some dimensional reduction methods, sometimes resulting in reduced model performance.

When the materials in a dataset are described by a relevant subset of features, an ML algorithm is chosen to train models using these data. The choice of algorithm depends on the nature of the dataset and problem to be solved. For example, supervised learning is used for gas uptake capacity prediction and adsorbent screening. In supervised learning, all the training data must be labelled by the property of interest.[36] Unsupervised learning is preferred when the aim is to identify the patterns and trends in unlabelled data.[93] The largest difference in the performance of ML algorithms is between linear and nonlinear methods. **Table 1** summarizes algorithms commonly used in materials science.

Among the nonlinear algorithms, NNs have been successfully applied to materials science, particularly for the development of adsorbent materials.[24,94–102] NNs are composed of an input layer, an output layer and interconnected hidden layers.[103,104] In each hidden layer, there are a series of units (neurons) containing nonlinear transfer functions that pass inputs forward and errors backward to allow the weights and biases of each unit to be adjusted. The number of hidden layer neurons depends on the nonlinearity of the problem to be solved. Simple neural networks usually contain a single hidden layer with relatively few neurons. A deep neural network (DNN) contains multiple hidden layers, each layer containing many neurons (**Figure 2**a). DNNs have numerous applications in adsorbent materials science, including design of adsorbent materials with high gas adsorption rate and illustration of the underlying structure–adsorption relationships.[90,91,105,106] A subset of DNNs, convolutional neural networks (CNNs), are very useful for image recognition and analysis.[104] Unlike other DNNs, the hidden layers of CNN consist of a number of convolutional and pooling layers (Figure 2b). The convolutional layer maps the input tensor to a feature map using multiple kernel filters then transmits the output to the pooling layer that performs downsampling and further convolution. The final feature map with more abstracted features is transmitted to the fully connected layer for regression or classification. For adsorbent materials investigations CNNs have been trained to extract the chemical and physical characteristics from their topology images or diverse spectra.[107–111]

Finally, the accuracy of the model predictions must be evaluated. For regression models, the error metrics are the coefficient of determination (r^2), root mean square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE). The RMSE and MAE are preferred over r^2 values as they are not dependent on the number of data points and number of parameters in the model.[112,113] MAE values are less biased by one or two large outliers in the predictions than RMSE values. Likewise, MAPE is independent of scale and easy to interpret, but it will become infinite when there are actual values close to zero. For the classification models, accuracy, F1 score, geometric mean of recall and precision (G-mean), and the area under the receiver operating characteristic curve (AUC) are the metrics widely used for scoring model performance. F1 score, G-mean, and AUC are suitable for unbalanced classification models where one class is more highly represented than the other. By using these metrics, we can evaluate how accurately the models can predict the properties of the training data used to generate the model, and the test data that is not used in training. A good regression model should have the r^2 close to 1, while RMSE and MAE should be close to 0 on both training set and test set. For a good classification, the accuracy, F1 score, G-mean and AUC should all be close to 1 on both training set and test set. A model is underfitting when the performance on the training set is poor, while overfitting is identified when good performance is obtained on the training set but poor performance on the test set. Both underfitting and overfitting can be avoided by increasing the size of the dataset, using fewer or greater relevance features, and modeling using linear and nonlinear ML algorithms.

Model interpretation is of critical importance in material research. Linear regression and decision tree-based models are intrinsically interpretable and provide global interpretations. In some cases, however, interpretation can fail to capture true feature–property relationships. For example, a linear model cannot explain nonlinear relationships no matter how much regularization is carried out. To address these issues and allow interpretation of nonlinear models, new methods have been developed. A commonly used method is permutation feature importance, which estimates the importance of a feature by calculating the increases of the model error after permutating this feature.[114] This method is fast, easy to understand, and gives global interpretation of features that span a wide range for nonlinear relationships, but it may be bias when there are correlated features, it cannot illustrate the effects of features on predicted values, and the results may lack reproducibility as it adds randomness in the calculation. Shapley additive explanation (SHAP) is another popular method in ML studies.[115] SHAP can measure the contribution of each feature to the prediction of an individual sample. It attempts to generate global interpretations, usually spanning a range of values and signs due to the fact that feature importance is a local property for nonlinear models.[116–118] SHAP can also give unreliable results when features are correlated, and thus the results should be scrutinized by domain specialists. For NN models, salience methods (e.g., class activation maps),[119] attention masks,[120] and partial derivatives (sensitivity analysis)[121] are used to interpret these "black box" models. Interested readers can find more details about model interpretations in the recent review by Oviedo et al.[122]

Good practice in ML research requires good quality data.[26,36] All data used, especially that obtained from different sources, should be reproducible and comparable. To avoid overfitting, featurization must be applied to optimize the number of

Table 1. The characteristics and disadvantages of different ML algorithms and their applications to adsorbent materials.

Algorithm	Purposes	Characteristics	Disadvantages	Examples
Linear regression	Regression	• Simple and fast • Good performance on small datasets • Good interpretability	• Poor performance when feature–property relations are nonlinear	[71, 75, 78, 82]
Logistic regression	Classification	• Simple and fast • Good performance on small datasets • Easy to be updated with new data	• Poor performance when feature–property relations are nonlinear • Poor performance on high dimensional feature spaces	[83]
Kernel ridge regression (KRR)	Regression	• Provide nonlinear solution	• Slow on large datasets	[43, 84, 85]
k-nearest neighbors (kNN)	Classification, regression	• Simple principles • Insensitive to outliers • Nonlinear analysis • Easy to be updated with new data	• Numbers of neighbors (k) are defined by user • Slow on large datasets • Poor performance on biased samples	[78, 75, 86]
Naive Bayes	Classification	• Fast • Insensitive to missing data and irrelevant features • Multi-class predictions • Easy to be updated with new data • Good interpretability	• Each feature should have independent and equal contribution to the outcome	[86]
Support vector machine (SVM)	Classification, regression	• Available on high dimensional feature spaces • Provide nonlinear solution • Provide small-sample solution • Insensitive to outliers • Global solution (no local minima issue) • Good generalization ability	• Slow on large datasets • No general rule for choosing kernel function • Only available for binary classification • Sensitive to missing data • Poor interpretability • Tends to overfit models	[78, 75, 83, 86, 87]
Random forest (RF)	Classification, regression	• Ensemble of decision trees • Available on large datasets with high feature space dimensionality • Can handle missing data • Insensitive to outliers • Good generalization ability • Can evaluate feature importance (thus can be used in feature selection) • Resistance to overfitting	• High complexity • Slow (when the number of decision trees is large) • Each decision tree should be independent • Biased for small datasets	[78, 75, 83, 86, 70, 88]
Extremely randomized trees (EXT)	Classification, regression	• Similar to RF except using the whole original sample instead of bootstrap • Choose cut points randomly instead of optimum split • Less variance than RF • Better generalization than RF	• Generates more decision trees in the model than RF • Larger bias than RF	[75, 88]
Gradient boosting trees (GBT)	Classification, regression	• Ensemble of decision trees • Can handle missing data • Can evaluate feature importance (thus can be used in feature selection)	• Sensitive to outliers • Tend to overfit when the number of decision trees is large • Slow for large datasets	[43, 75, 88]
XGBoost	Classification, regression	• More accurate than GBT • The in-built regularization can prevent overfitting • Can handle missing data	• Slow for large datasets • High space complexity	[88]
Neural network (NN)	Classification, regression	• High accuracy • Intensive learning ability • Can store information in the network • Robust to missing data and noise	• No objective method for choosing architecture unless Bayesian regularized • Poor interpretability • Local minima issues	[43, 78, 83, 89]

(*Continued*)

Table 1. (Continued).

Algorithm	Purposes	Characteristics	Disadvantages	Examples
Deep neural networks	Classification, regression	• High accuracy • Intensive learning ability • Can store information in the network • Robust to missing data and noise • Can generate useful latent descriptors from simple molecular representations	• Slow • Many weights requiring large training datasets • Poor interpretability • High complexity • Requires effective regularization to avoid overfitting	[90–92]
k-means	Clustering	• Simple and fast • Unsupervised	• Numbers of clusters (k) are defined by user • Poor performance when the shapes of clusters are irregular • Sensitive to noise	[78]
DBSCAN	Clustering	• Numbers of clusters are obtained by the algorithm • Can handle clusters with irregular shapes • Can identify clusters and noise • Unsupervised	• Slow, especially on large datasets • Poor performance when the clusters have very different densities	[23]

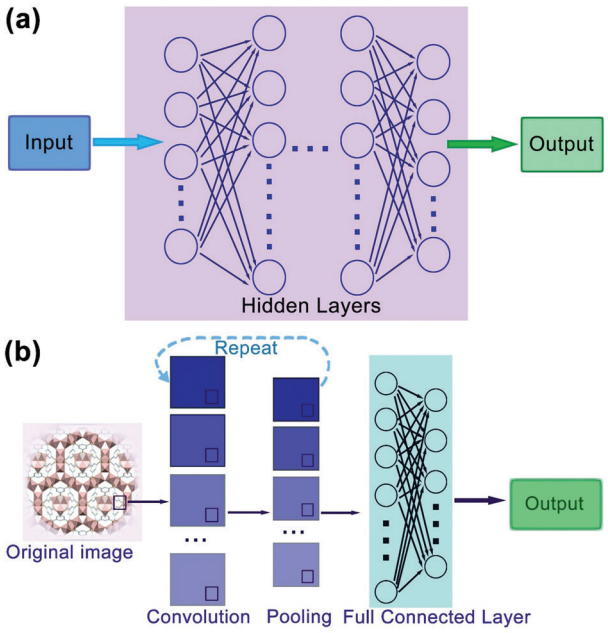


Figure 2. Representations of a) deep neural networks (DNN) and b) convolutional neural networks (CNN). The DNN comprises one input layer, one output layer, and a few hidden layers, each of the hidden layer contains multiple neurons, where the inputs were passed forward and errors backward to adjust the weights and bias of each node. The CNN is a class of DNN in which the hidden layers consist of several convolutional and pooling layers. The full connected layer is a traditional multilayer NN used to predict the value or the class of the images.

features.[24,26,31,36] Some features, such as structural and energy features, are more important for predicting gas adsorption for adsorbent materials than others, but the complexity of their measurement and calculation should also be considered. The research objective and nature of the structure–property relationship determine the selection of the ML model. Clearly, models with good predictivity and interpretability are preferred, while training time, size of training set, and domains of applicability (the range of properties in which the model makes reliable predictions) should also be considered.[24,36,123] Using informative features improves model predictivity and interpretability. Selecting such features and model interpretation requires collaboration between materials scientists and data scientists. Ideally, ML models should be consistent with the known physicochemical theories or provide new insight for materials science.

3. Examples of Machine Learning Approaches for Adsorbent Materials

ML models can be used in materials science in three main ways: discovery, design, and optimization. In the development of adsorbent materials, high-throughput first principles calculations can be useful to simulate some properties but are limited by the time and cost of the calculations. The gas adsorption capacity of an adsorbent material, one of the most important metrics for these materials in commercial use, is influenced by local properties such as structure, topology, and sorption sites for gas molecules; properties difficult to simulate by physics-based methods on a large scale. However, ML is not only effective in predicting local properties but also in properties that are influenced by intrinsic and external factors (e.g., gas adsorption and selectivity).[24,28,31,124] ML models can be trained on the data generated by high-throughput calculations to rapidly predict the target properties of related materials.[30,125,126] New materials are best discovered using ML models with broad domains of applicability to screen large databases of real or hypothetical materials to identify candidates with potentially useful properties.[37,127] The domain of a model is the range of feature space and property space represented by the training data. The further away from the domain that predictions are made, the less accurate they will be. Moreover, target properties of materials can be improved by modifying the structure, composition, or other features that are suggested to have critical effects by ML models.[39,128,129] In this

section, ML applications to adsorbent materials are discussed, including metal–organic frameworks, porous carbons, zeolites, covalent–organic frameworks, porous polymers networks, and intermetallics.

3.1. Metal–Organic Frameworks

MOFs are a class of hybrid inorganic-organic nanoporous materials formed by the self-assembly of metal clusters and polydentate organic linkers as structural units that create open crystalline frameworks.[17] Since their discovery, MOFs have garnered significant interest for a wide range of applications such as gas separation and storage, catalysis, sensing, nonlinear optics and as light absorbers due to their highly tunable nature.[130–134] In principle, databases of millions of new MOF structures can be readily generated by systematic functionalization of the well-known organic linkers, from which new MOFs with excellent target properties may be identified.[135–137] However, finding an optimal MOF structure for a given application is challenging. For example, for Zn_2(1,4-benzenedicarboxylate)$_2$(pyrazine) (ZBP), when the four symmetric substitution points in this compound are functionalized by a small library of 35 functional groups, there are a total of 35^4 possible combinations of new MOF structures.[138] It is impractical to locate the optimal MOF structures in such large databases using experiments or physics-based computational methods. Therefore, ML methods can alleviate the computational burden by preselecting candidates with predicted high performance.[40,139–144]

The porous structure, large surface area, and tunability of MOFs provide exceptional performance in gas separation and storage, especially storage of hydrogen, carbon dioxide, and methane.[133,137,145] Many new MOF structures with high gas uptake values and good selectivity have been discovered with the assistance of ML techniques. Fernandez et al. described a robust SVM classifier that rapidly identified promising MOFs for CO_2 capture.[146] A database of 324 500 hypothetical MOF structures was generated by combining 66 structural building units and 19 functional groups, of which 10% were randomly selected to form a training set. Atomic property-weighted radial distribution function (AP-RDF) descriptors were used to represent the atomic properties and electronic structural information of the MOF structures. A grand canonical Monte Carlo (GCMC) simulation was carried out to label the data points with the CO_2 uptake values at 0.15 and 1 bar CO_2 at 298 K.[147] When screening a material space of 292 050 MOFs, the 0.15 bar classifier successfully identified 945 of the 1000 MOFs with the highest CO_2 adsorption capacity in the dataset. As the properties of only 10% of the MOFs in the database needed to be calculated, and a high accuracy prediction was achieved, the ML approach proved useful for accelerated screening of large search spaces. Burner et al. developed a NN model to predict the CO_2 uptake capacity and CO_2/N_2 selectivity of MOFs under low pressure.[148] The model was trained on a dataset of 340 000 MOFs with over 1000 topologies. They found that the model had the best pe rformance when six geometric descriptors together with AP-RDF and chemical motifs were used as features. The model identified 994 MOFs with the highest CO_2 adsorption capacity from a test set of ≈70 000 MOFs.

ML has been used to elucidate feature–property relations. Fernandez et al. investigated the relationships between the geometric features of MOFs and their CO_2 and N_2 adsorption using an RF model.[78] 81 679 MOFs with unique frameworks were collected from the Northwestern University database, from which 16 000 data points were selected as a training set. Five geometric descriptors (dominant pore size, maximum pore size, void fraction, volumetric surface area, and gravimetric surface area) used to train the RF classifier yielded an accuracy >94% for both gases. It identified over 70% and 60% of MOFs known to have high performance for CO_2 and N_2 capture, respectively, in a vast search space of ≈65 000 MOFs. They also developed a binary decision tree model to suggest the optimal combination of the five descriptors that enhance the CO_2 uptake under low pressure, previously only achieved by a sophisticated radial distribution function model (**Figure 3**). Anderson et al. studied the effects of geometric and chemical features on the prediction accuracy of CO_2 capture using several ML models.[149] The ML methods provided an unbiased approach to evaluating the importance of the descriptors on materials performance, providing useful insight into structure–property relationships. As is often the case, improvement in CO_2 capture prediction depended strongly on the chemical descriptors, while the absolute values of CO_2 capture prediction were mostly related to the geometric descriptors. In addition, ML techniques could be used to bypass the complicated first-principle calculations and predict the partial charges of MOFs, with which the CO_2 adsorption properties could be accurately calculated.[116]

Another important application for MOFs is gaseous fuel storage. The effects of different ML algorithms and descriptors on fuel gas adsorption prediction accuracies have been investigated. Pardakhti et al. constructed four ML models of CH_4 uptake using chemical and crystal structure descriptors.[150] The RF model had the best performance, and incorporation of chemical descriptors greatly enhanced prediction accuracy while maintaining computational efficiency. Kim et al. studied the CH_4 uptake isotherm at a range of temperatures from the isotherm at 298 K using three ML models. Texture features such as surface area and total pore volume, obtained from gas adsorption–desorption experiments, were the major determinants of CH_4 uptake capacity of MOFs at different temperatures.[151] Instead of using experimentally derived features, Gurnani et al. created fingerprints to represent 137 953 hypothetical MOFs.[73] They used a series of atomic features describing the coordinating metals (e.g., electronegativity, ionization energy, atomic radii, etc.), and exploited the SMILES strings for the linkers. The speed and generalization ability of the model for CH_4 uptake capacity of MOFs were analyzed in this report. Wang et al. used the molecular graph (the way the atoms are connected) of MOFs as descriptors and developed a CNN model to predict their CH_4 adsorption properties.[152] This model could reliably recapitulate the properties of the test set so was used to screen a database of 330 000 hypothetical MOFs to discover four MOFs with excellent predicted CH_4 adsorption ability. Moreover, this model showed good transferability and could be used to predict the CH_4 adsorption for COF and zeolitic imidazolate framework (ZIF) materials.

Clearly, feature choice significantly affects the model accuracy. Anderson et al. built a NN model to predict the amount of H_2 adsorbed at different temperatures and pressures.[153] To simplify

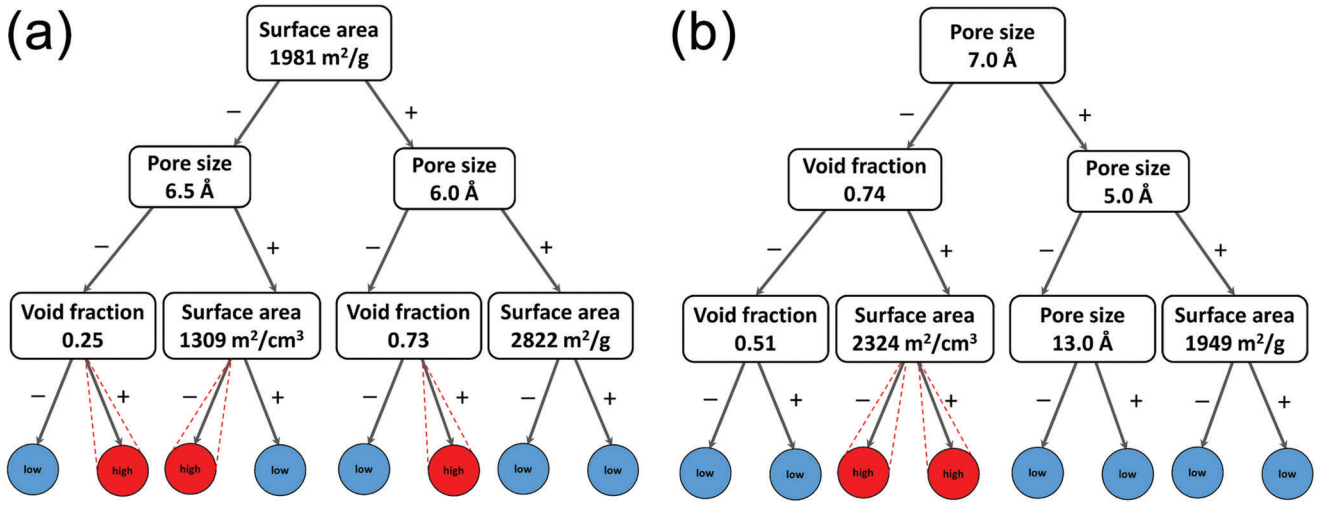


Figure 3. Binary decision tree models of MOFs with a) high CO_2 capacity (higher than 1 mmol g^{-1}) and b) high N_2 capacity (higher than 0.5 mmol g^{-1}). The output nodes referring to low and high gas uptake capacity are highlighted in blue and red color, respectively. Reproduced with permission.[78] Copyright 2016, American Chemical Society.

the model, they used seven features that could be readily obtained by calculations: void fraction, framework density, largest cavity diameter, pore limiting diameter, volumetric surface area, alchemical catecholate site number density, and the epsilon for the interaction of hydrogen with the alchemical sites. This model suggested that a large reduction of pressure (from 100 to 35 bar) only slightly influenced the H_2 adsorption capacity. Such a reduction of H_2 pressure could improve safety and compression costs in commercial use. Borboudakis et al. studied CO_2 and H_2 adsorption with ensemble learning from three ML models. They represented the structures of the MOFs by encoding the presence or absence of the building blocks (such as organic linker, metal cluster, and functional groups) as a binary parameter.[154] Although the accuracy of this method was acceptable, it was not able to predict gas adsorption of MOFs whose linkers or metals were outside of the domain of the training set. To address this, the building block features were substituted by the atom type number density (calculated by the numbers of a particular atom in the MOF unit cell over the unit cell volume) in each structure, and bonds, angles, torsions, and pair interactions were used to represent the elements and connectivity types.[155] An RF model was established after training 100 times with randomly selected training sets ranging in size from 50 to 10 000. The effects of three feature families (structural features alone, structural features with MOF building blocks, and structural features with atom type number density) were studied. For CO_2 and CH_4 adsorption under the pressures examined, significant improvements were obtained when atom type number density was used (**Figure 4**). These features allowed the model to be extended to different porous materials such as COFs. Ma et al. trained a DNN model on the H_2 adsorption data with 13 506 MOFs at 100 bar and 243 K.[92] The MOFs were represented by five physical features (void fraction, volumetric surface area, gravimetric surface area, pore limiting diameter, and largest cavity diameter). The good generalizability of this model made it useful for predicting H_2 adsorption at 130 K as well as being applicable to predicting CH_4 adsorption under similar conditions. Unsurprisingly, the model showed poor performance when applied to Xe adsorption, indicating the large differences in feature–adsorption relations between fuel gases and inert gases.

The use of solid adsorbents for noble gas adsorption and separation has progressed with the help of ML techniques. Liang et al. built a XGBoost model with seven physical features to predict the Xe/Kr adsorption and selectivity of MOFs.[156] Xe is an important propellant used for ion thrusters in spacecraft, therefore, the separation of Xe from Kr is critical for aerospace energy applications. They found that the density, porosity, pore volume, and pore limiting diameter of MOFs are crucial features affecting the Xe/Kr adsorption. Surprisingly, this model could be extended to screen the MOFs for the separation of a CH_4/CO_2 mixture.

ML methods can also be used to design MOFs, providing guidance for synthesizing MOFs and other porous materials with bespoke properties. Zhang et al. reported a combined computational approach using a Monte Carlo tree search (MCTS) (an algorithm analogous to reinforcement learning)[157] and recurrent neural networks (RNN, a type of NN developed to tackle sequential data)[158] to design MOFs for CO_2 adsorption (**Figure 5**).[159] This approach begins with a given metal vertex, a MOF topology, and the target property (CO_2 adsorption). An RNN model was trained on 168 130 SMILES strings representing linkers (edges) collected from the ZINC database. The MCTS built a tree in which each node denoted one symbol from the SMILES string by repeating four steps: selection, expansion, simulation, and backpropagation. In the first step, a path from the root (metal node) to a node at *i* level of the current tree was built by choosing the child nodes with maximum upper confidence bound that considered the sum of the target property after *i* simulations. After reaching the leaf of the current tree, child nodes (any valid symbols in SMILES string) were expanded under the nodes at level *i*. Then, the RNN model created the remainder of the strings to simulate a complete linker based on the partial string already built. Using this simulated linker, metal node, and the topology, a MOF was constructed whose CO_2 adsorption capability was simulated using GCMC. This predicted CO_2 adsorption value was then back-

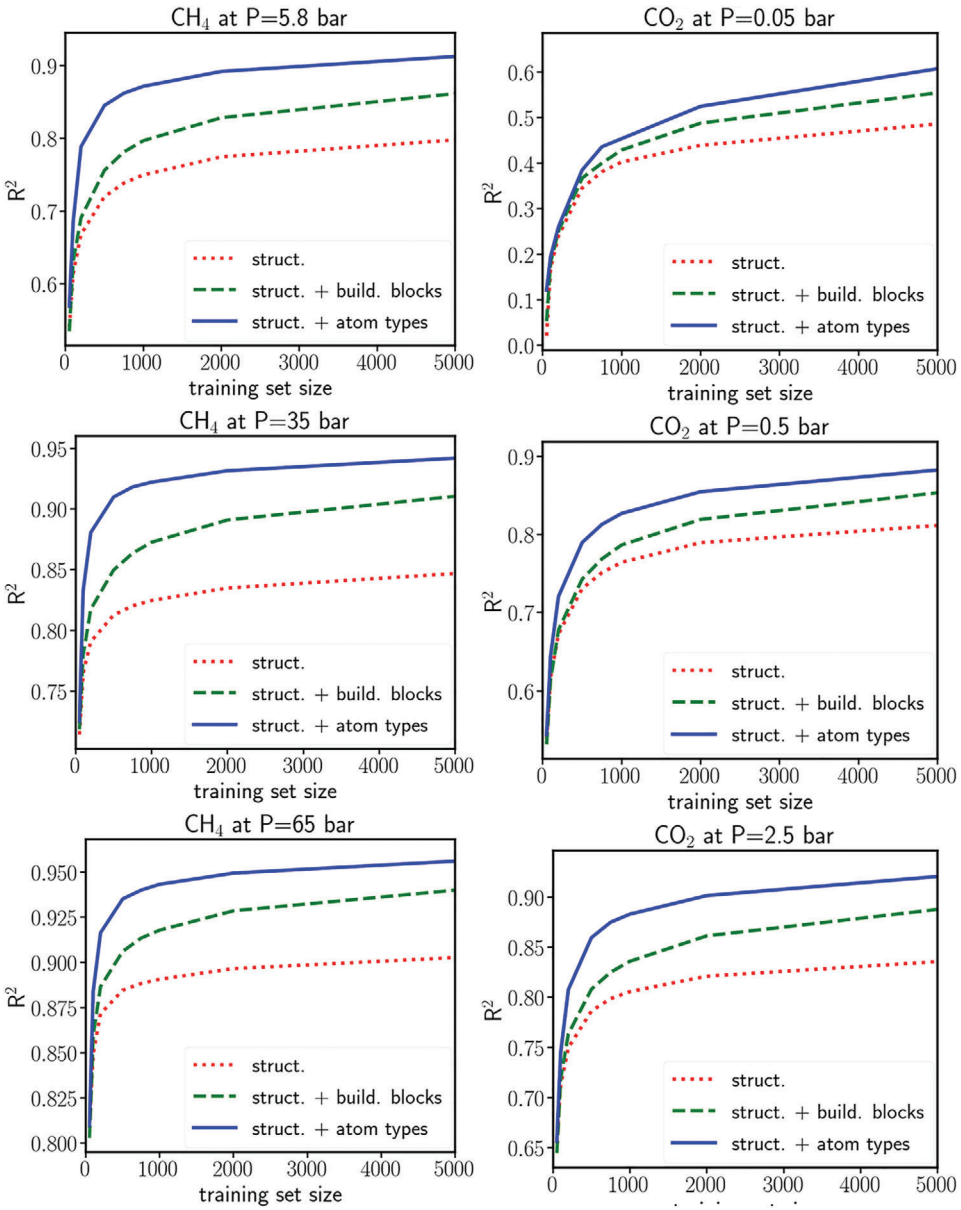


Figure 4. Variation of the R^2 versus the training set size for (left column) CH_4 and (right column) CO_2 under different pressures. Reproduced with permission.[155] Copyright 2020, American Chemical Society.

propagated to the tree and used to update the upper confidence bound on each node. These four steps were repeated iteratively until the string hit the terminal symbol or the maximum length. Several MOFs with high CO_2 adsorption were thus designed using 10 combinations of metal nodes and topologies extracted from experimental MOFs reported in the literature. Moreover, by applying the topological data analysis, new MOFs with diverse topologies could also be designed. Despite the success of this approach, there can still be difficulties in synthesis or self-assembly of the MOFs, hence their ability to form stable materials with the expected structures. To address this issue, Collins et al. used a genetic algorithm to optimize and discover MOFs.[160] To optimize ZBPs [Zn_2(1,4-benzenedicarboxylate)$_2$(pyrazine)], they used 28 common functional groups to generate 96 156 hypothetical, stable structures. The materials genome used by the genetic algorithm was the sequence of equivalent sites and their associated functional groups, while the CO_2 uptake was the fitness function. After genetic algorithm optimization, a 4.8-fold increase in the CO_2 uptake was achieved by a new structure. The method was extended to optimize 141 experimentally characterized MOFs, giving rise to 1035 functionalized structures that were predicted to have exceptional CO_2 uptake (>3 mmol g^{-1} at 0.15 atm and 298 K). Using a different approach, Moosavi et al. attempted to optimize the synthesis conditions and build knowledge on accessible chemistries based on successful and failed synthesis experiments.[161] A robot was used to synthesize Cu-BTC[162] (BTC is benzene-1,3,5-tricarboxylic acid) by manipulating nine synthesis parameters (**Figure 6**). Since experimental

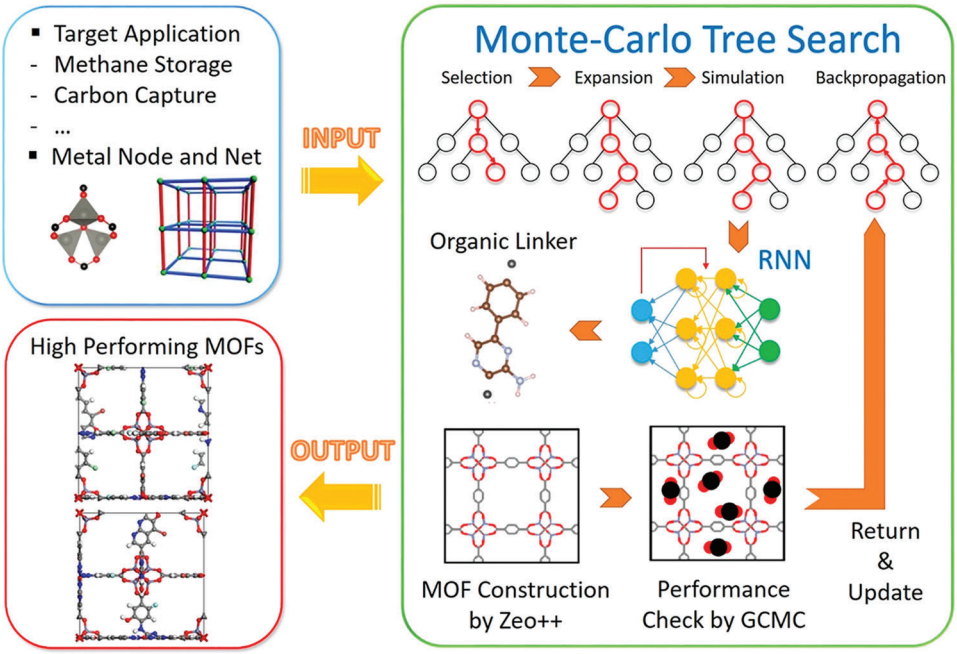


Figure 5. Schematic of an algorithm to design an application-specific MOF. With given inputs about target application and type of metal node and net, organic linkers were generated by combining the MCTS and RNN. A new MOF constructed by Zeo++ underwent a performance check for the target application and then internal parameters in MCTS were updated. Reproduced with permission.[159] Copyright 2020, American Chemical Society.

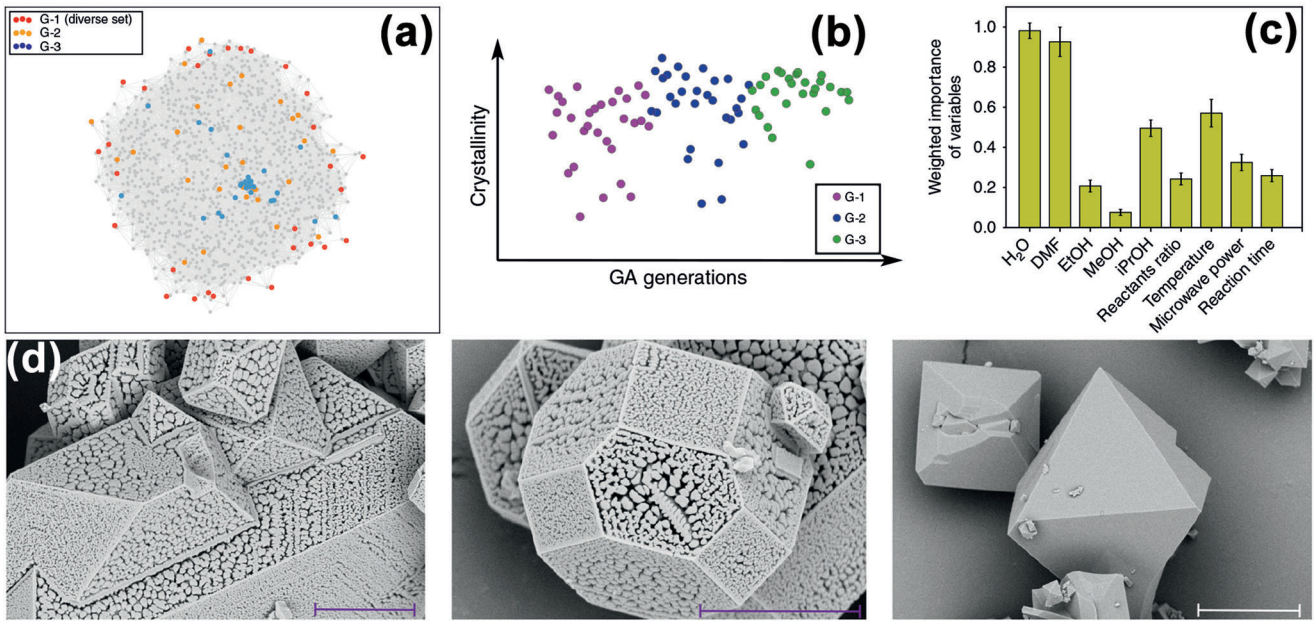


Figure 6. a) Projection from a 9D parameter space to a 2D plane. Grey dots denote the known synthesis conditions. b) Progress in crystallinity in different generations. c) Relative importance of the nine parameters on Cu-BTC synthesis. d) Scanning electron microscopy images of several Cu-BTC samples. Scale bars were 5, 4, and 10 μm for the samples shown from left to right, respectively. Reproduced with permission.[161] Copyright 2019, Springer Nature.

exploration of these parameters is infeasible, a genetic algorithm was used to accelerate the optimization, with crystallinity, phase purity, and surface area as fitness functions. An RF model was trained to rank parameter importance. It was found that the amount of water and DMF played the largest roles in the synthesis, while the temperature had three times more impact than changing the reactant ratios. Accordingly, a new synthesis could be designed for optimal targets.

There is also increasing interest in the electronic properties and stability of MOFs.[130–132] Although most MOFs are insulators with bandgaps over 2 eV, conductive MOFs have been discovered by experiments and theoretical simulations.[163,164]

Because of their high porosity and large surface area, conductive MOFs are largely used for electrochemical energy conversion and energy storage.[165–167] ML methods have been used to accelerate the screening of large search spaces to discover new conductive MOFs. Despite the large amount of information on MOFs in databases, bandgap information is seldom provided. This is an important characteristic when searching for conductive MOFs. Transfer learning, where the model stores knowledge gained from one property then applies it to solve problems in other properties, was employed to tackle this issue.[83] Four ML models (logistic regression, SVM, NN, and RF) were trained using 52 300 inorganic compounds from the Open Quantum Material Database and 45 optimized descriptors. Subsequently, t-SNE was used to reduce the 45D space to a 2D space, with data points exhibiting some overlap. The authors proposed that this bandgap model could be used to predict the bandgaps of MOFs. To increase the accuracy of prediction, a consensus of predicted bandgaps from four models was used to find those likely to be conductors. From a pool of 2932 MOFs, nine were predicted to be conductive, with six subsequently confirmed as conductive by ab initio calculations. In addition to transfer learning, CNN models were found to accurately predict the bandgap of MOFs. Kernel ridge regression (KRR) models with SOAP fingerprints or composition-based features predicted bandgaps less accurately than the CNN model.[168] The water stability of MOFs is an important property for commercial applications in gas storage, which has also been investigated by ML techniques. ML classifiers have been constructed using features encoding metal electronic properties, linker SMILES strings, molar ratios of the linkers, numbers of O, OH, H_2O species with respect to metals, culminating in discovery of several MOFs with aqueous stability.[169] ML methods were used to avoid the detrimental effects of water on MOF gas adsorption, identifying two water-stable MOFs by a computational screen of 300 000 MOFs.[170]

Artifical intelligence approaches have also been applied to design MOF-based devices. For example, the fabrication of gas (CH_4) sensors composed of MOF arrays was optimized by a genetic algorithm.[171] By optimizing the parameters of the arrays, both the selectivity and the sensitivity of the sensor were significantly enhanced. This study had practical applications in detecting and preventing natural gas leaks in the methane fuel industry.

These examples have exemplified the fact that ML methods can reduce the computational cost and accelerate screening of extremely large MOF spaces. New MOFs with diverse topologies, coordinating metals, and molecular structures for a range of applications have been designed using ML methods.[84] These methods can be extended to different frameworks and will widen the applications of MOFs.

3.2. Porous Carbon

Porous carbon is a promising material for gas capture due to its low cost, fast adsorption–desorption kinetics, large surface area and pore volume, and tunable pore structure.[172,173] ML approaches can elucidate the relationship between physical properties and gas adsorption ability of carbon.[32] Zhang et al. trained a NN model on a set of ≈1000 CO_2 adsorption data points from literature and experiments (**Figure 7**).[174] The three descriptors that had the most significant effect on the adsorption model were surface area, mesopore volume, and micropore volume. This model accurately predicted CO_2 adsorption properties of porous carbon materials. Zhu et al. illustrated in detail the effects of different types of pores on CO_2 adsorption.[74] They trained a RF model on 6244 CO_2 adsorption data points generated for 155 porous carbons and found that increasing the volume of micropores and mesopores had a negative effect on CO_2 adsorption under low pressure. However, the model indicated that increasing the volume of ultra-micropores improved CO_2 adsorption when the pressure increased. To study the selectivity of CO_2 adsorption, Wang et al. trained a DNN model on experimental data for CO_2 and N_2 uptake on porous carbons and concluded that high CO_2 selectivity could be achieved when the porous carbons possessed moderate micropore (0.4–0.6 $cm^3\ g^{-1}$) and mesopore volumes (0.4–1 $cm^3\ g^{-1}$).[175] To further elucidate the effects of porosity on CO_2 selectivity, a CNN model was built to predict the separation performance of porous carbon, using a CO_2/N_2 mixture as a test case.[176] The model suggested that the best porous carbon with high CO_2 adsorption selectivity should have pores with a bimodal size distribution, in which the pore size was in the range of 3–7 nm or less than 2 nm.

The effects of the features of porous carbon on fuel gas adsorption has been studied by ML methods by Zhang et al. who trained a feedforward NN model on the literature data. They used this model to understand the relationships between physical properties of the adsorbent and CH_4 adsorption.[177] Kusdhany et al. also trained an RF model on a dataset of 1745 data points from 68 porous carbons.[118] The model showed that pressure and surface area played critical roles in H_2 uptake capacity prediction, Unlike previous studies, they found that oxygen content was also an important factor in predicting H_2 uptake, while pore volume had little effect.

These examples indicate how ML can guide the design of highly efficient gas adsorbents and provide a better understanding of the gas adsorption kinetics by highlighting the importance of physical parameters that may have been previously unrecognized. It is expected that ML models will become even more useful tools for designing and optimizing porous materials and uncovering new physical insights about gas adsorption mechanisms when larger and more reliable datasets become available.

3.3. Zeolites

Zeolites are microporous crystalline aluminosilicate materials[178] with well-defined cavities and pores that make them very useful for catalysis, adsorption, ion exchange, renewable energy conversion, and water purification.[21,178,179] Much effort has been devoted to designing and tailoring zeolites for specific applications.[21,180,181] Gas adsorption in zeolites has been more extensively studied by ML techniques than most other porous materials. Pai et al. used ML to optimize the operating conditions for CO_2 adsorption and selectivity for zeolites, suggesting that this technique could be used to develop zeolites for post-combustion CO_2 capture.[182] Göltl et al. investigated the CO and NO adsorption of zeolites SSZ-13 and pentasil zeolite mordenite by a linear regression model.[183] The variables in this model were optimized by a genetic algorithm. By analyzing

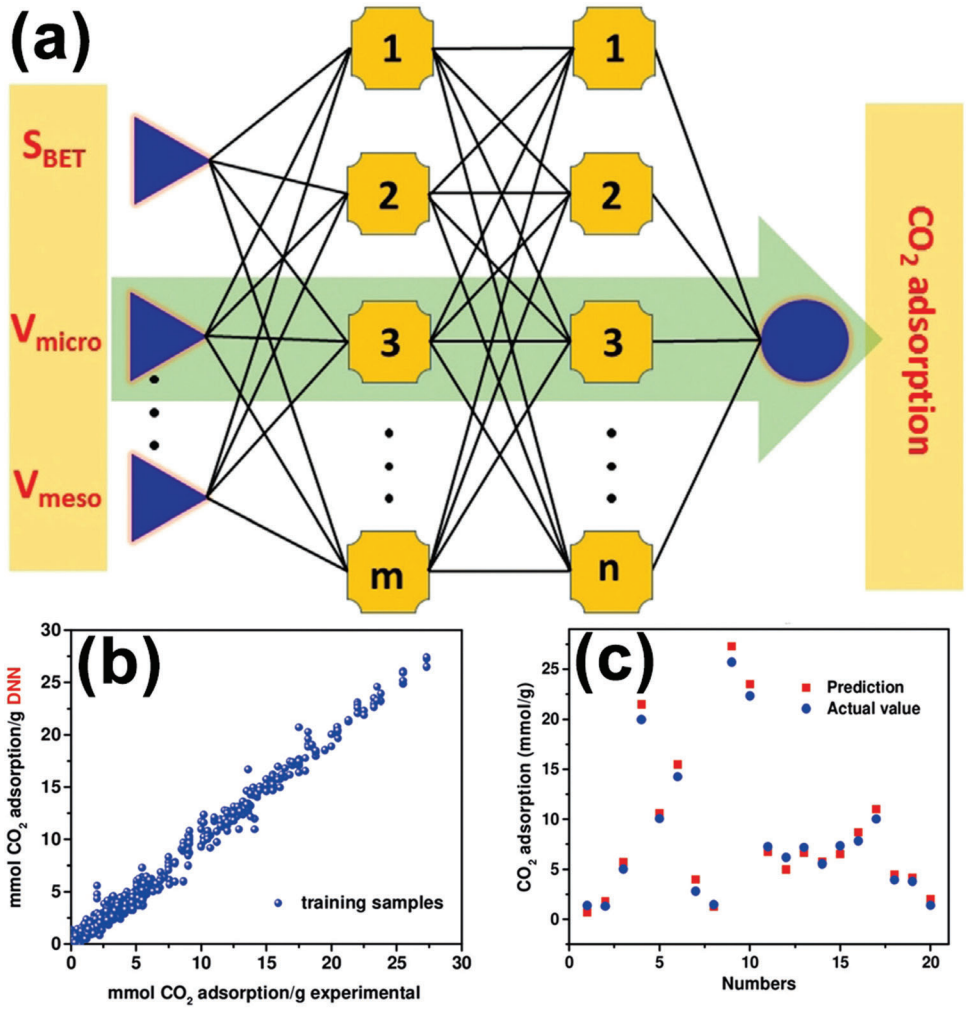


Figure 7. a) Schematic architecture of a DNN model. Inputs were surface area (S_{BET}), mesopore volume (V_{meso}), and micropore volume (V_{micro}), and output was CO_2 capture capacity. Each line between two nodes represented a weight. By tuning these weights, the input–output relation can be simulated. b) The experimental predicted CO_2 adsorption versus model predicted CO_2 adsorption. c) The correlation between 20 experimental CO_2 adsorption data points and the corresponding model predicted values. Reproduced with permission.[174] Copyright 2019, Wiley-VCH.

the correlations between the descriptors, they found that the position of the s orbital, the number of valence electrons at the active site, and the HOMO–LUMO gap of the adsorbent had the largest impact on gas adsorption. The reconstruction of the active sites also had a noticeable effect on adsorption. As CO and NO are useful molecular probes in studies of adsorption and conversion of industrial and car exhaust gases (CO, CO_2, NO_x) by zeolites, this investigation provides a rational basis for the design of next generation zeolites with improved capacity and activity for the adsorption and conversion of greenhouse gases and toxic exhaust gases.

State-of-the-art computational techniques have also been used in the zeolite design. Kim et al. implemented a generative adversarial NN to produce novel zeolites for CH_4 capture (**Figure 8**a).[184] The model was trained on 31 713 zeolites, pairing the positions of oxygen and silicon atoms and the CH_4 potential energy grids. A total of 121 new porous structures were identified to have the desired heat of adsorption for CH_4, and this model could be extended to predict the heat of adsorption of other gases on other porous materials including MOFs and COFs. Cho et al. constructed a 3D CNN model on 6500 hypothetical zeolites that exhibited high prediction accuracy for CH_4 adsorption.[185] To enhance the model generalization ability, Sun et al. developed a meta-learning model to predict H_2 adsorption for a series of adsorbent materials under a wide range of pressure and temperature.[186] Meta learning is a technique that uses ML algorithms to determine the best combination of individual models for a new target (but related to the targets on which the individual models have been trained) with a small amount of training data.[187] In the study of gas sorption isotherms, a model showing good performance on one subset (e.g., gas adsorption of a series of materials, or gas adsorption under a range of pressure and temperature) might predict the adsorption poorly on another subset. Therefore, it is necessary to employ meta learning to find correlations between the model performances and the subsets, with which the meta learning model is built to adapt to the gas adsorption dataset generated from multiple materials or under a large range of pressures and temperatures (Figure 8b). Using the

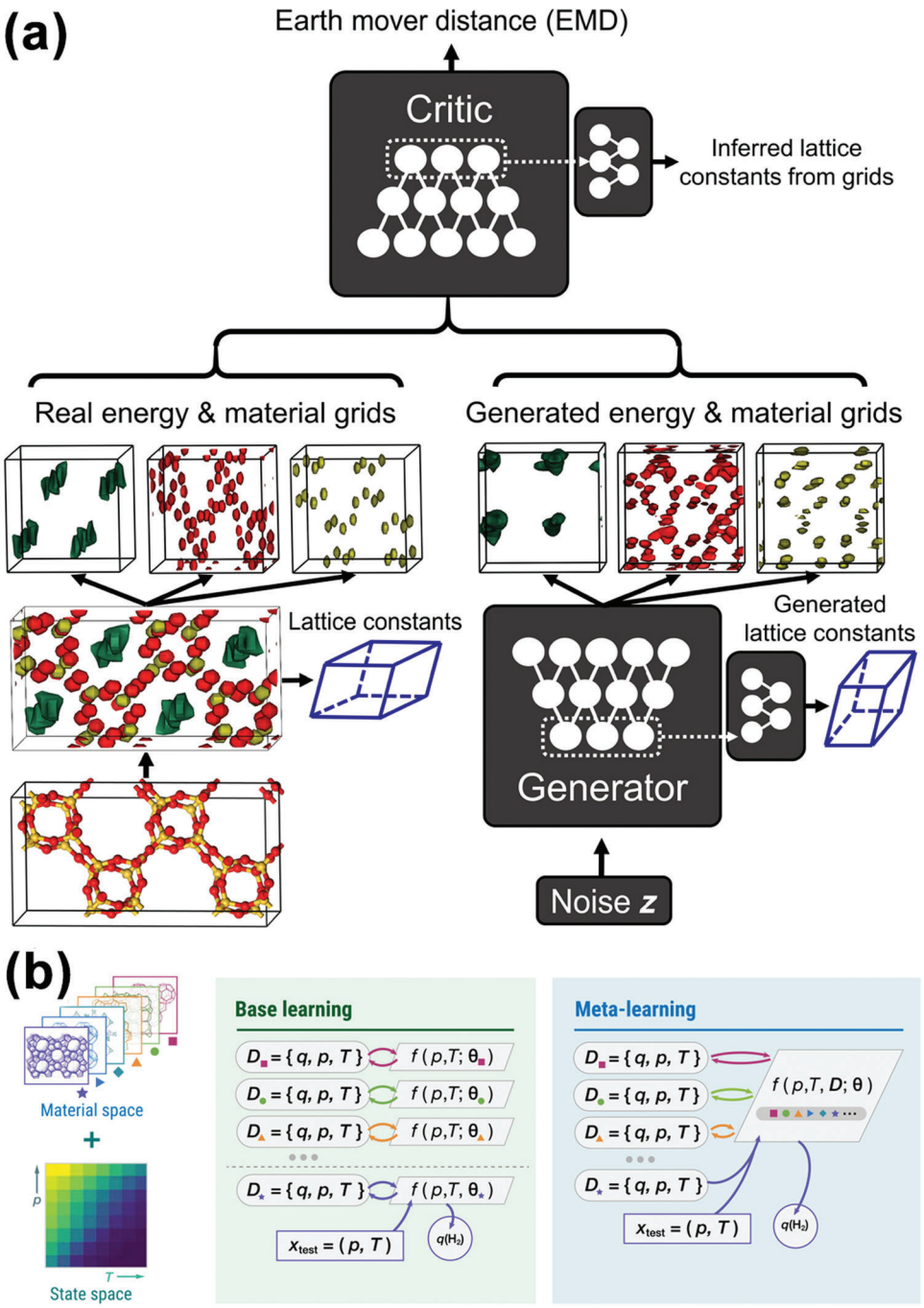


Figure 8. a) Schematic of the generative adversarial NN for zeolite design. Green dots referred to methane potential energy, and material grids indicated silicon (red) and oxygen (yellow) atoms. The energy and material grids of generated zeolites were evolved from Gaussian noise distribution, and the earth mover distance (EMD) between the real and generated energy & material grids was the metric to determine the convergence of training. Periodic padding, feature matching, and lattice constant generating network were added into the "Critic network" to infer rational lattice constants from grids. Reproduced with permission.[184] Copyright 2020, American Association for the Advancement of Science. b) Schematic of the meta-learning technique. Instead of building individual models on different subsets (base learning), meta learning consolidated the prediction of all materials into a single model. Reproduced with permission.[186] Copyright 2021, American Association for the Advancement of Science.

meta learning technique, the meta learning model was available for predicting the optimal H_2 storage temperature under a given pressure for a large variety of adsorbents including all-silica zeolites, hyper-crosslinked polymers, and MOFs.[186]

As gas adsorption and separation often require high pressure, the mechanical properties of zeolites are an important consideration for their commercial use. Evans et al. studied the elastic properties of zeolites using a GBR model and identified important features of SiO_2 polymorphs that modulate their elastic response.[188] Kim et al. used an active learning technique to find the zeolite structures with the highest shear moduli.[66] Starting from the International Zeolite Association (IZA) database where only a few zeolites were labeled by their shear moduli, they trained a ML regression model to predict the shear moduli of the rest of the zeolites in the IZA database. Then, they chose the zeolites that were most likely to have good mechanical properties from the test materials via the Bayesian optimization method, labeled their shear moduli by DFT calculation, and returned them to the training set to re-train the regression model. This process was repeated until the model predictions and the DFT calculations were concordant. 23 novel zeolite structures having excellent shear moduli were discovered using this active learning technique. As with the other porous materials classes, the ZIF examples also illustrate the great potential of ML approaches for design and optimization of zeolite adsorbents.

Apart from property improvement, design of synthesis routes and optimization of synthesis conditions are another important application of ML.[189] Most zeolites are generated by hydrothermal synthesis that is controlled by multiple correlated parameters and complex crystallization kinetics.[179] This makes it difficult to rationally optimize the synthesis conditions, hitherto relying on trial-and-error or theoretical simulation methods to uncover feature-property relationships. Consequently, ML approaches have been developed to guide the synthesis of zeolites with bespoke properties.[190] Daeyaert et al. trained a NN model on a set of 4781 organic structure directing agents, using molecular features as input and the stabilization energy for polymorph A zeolite beta, an important zeolite for enantiospecific catalysis and gas separation, as the predicted property.[191] The accuracy of the ML model predictions of stabilization energy was comparable to that from more computationally demanding molecular dynamics simulations. This model was used to search a much larger materials space, and several molecules were identified as structure directing agents in terms of their stabilization energy. These new molecules were potentially useful for the synthesis of polymorph A zeolite beta. Ma et al. reported a ML-based atomic simulation method to guide design of new $Si_xAl_yP_zO_2H_{y-z}$ zeolites[192] that are useful for gaseous fuel adsorption and separation.[193,194] They discovered that structure directing agents were important for the formation of micropores for aluminophosphates, silicoaluminophosphates, and pure silica zeolites, while strong alkali was much more important than structure directing agents for the formation of aluminosilicates. Similar ML modeling techniques are being increasingly used in zeolite design and screening.[37,190,195]

Jensen et al. extracted information on the synthesis of CHA and SFW zeolites from literature using a combination of natural language processing, HTML and XML parsing, and regular expressions.[196] A RF model trained on these data indicated the importance of specific synthesis conditions, the Si/Ge molar ratio, the H_2O/T molar ratio (T is the TO_4 tetrahedron in zeolites), and the volume of the organic structure directing agent on zeolite framework design. CHA and SFW zeolites are promising materials for the mitigation of pollutant gases and adsorption of H_2 and CH_4. This study provides a pathway to materials with improved clean energy storage and useful environmental remediation properties.[197,198] Corma et al. used ML methods to predict synthesis conditions for successful zeolite syntheses,[199] a well-known greenhouse gas adsorbent.[200,201] They elucidated the relationships between different synthesis parameters with the performance of Ti-silicates using a NN model, and used this to optimize the synthesis of the next generation of materials using a genetic algorithm. Specifically, they found that the catalytic performance of the zeolite was enhanced by decreasing the amount of organic modifier while maintaining the OH/Si ratio of ≈0.2. To further illustrate synthesis condition–structure relationships and to provide direction for synthesis of unknown zeolites, Muraoka et al. used an extreme gradient boosting RF model. They extended its prediction space through a similarity network of crystal structures based on structural features and synthesis parameters (**Figure 9**).[202] This model was initially trained on a set of experimental data with synthesis parameters. The good accuracy in predicting the synthesis of zeolites with various structures allowed it to be extended to the prediction of synthesis of zeolites with structures outside the training set. They also generated structure fingerprints for each zeolite and merged them into the feature set. The zeolites were grouped by a k-means clustering algorithm where the similarity of two zeolites involved both structural and synthesis similarity. With the assistance of this similarity network, the model established optimal conditions for the synthesis of some novel zeolites and thus extended the diversity of the available dataset.

3.4. Other Adsorbent Materials

COFs (a class of porous polymer framework) have been widely used for H_2, CH_4, and CO_2 storage in clean energy applications.[18] Unlike MOFs and ZIFs, COFs are composed entirely of light elements (e.g., H, C, B, O).[203] These elements are linked by covalent bonds to form porous structures. High-throughput COF construction has been facilitated by an evolutionary algorithm, and large COF databases have been constructed.[204] However, it takes significant time and resources to explore the large chemical space of COFs with desired properties by traditional calculations (e.g., GCMC). ML approaches have been adopted to accelerate these property predictions. Desgranges et al. created ensemble models by averaging the outputs of the NN models with diverse architectures,[205] which were applicable to a broad range of applications such as prediction of CO_2 adsorption in IRMOF-1 ($Zn_4O(BDC)_3$, where BDC^{2-} = 1,4-benzodicarboxylate), H_2 adsorption by COF-102, and the separation of methane and ethane by COF-102 and COF-108. Optimization of ML model performance was achieved by appropriate choices of algorithms and model descriptors. Yang et al. used a tree-based pipeline optimization tool (TPOT) in an automated ML platform to analyze the CH_4 uptake by 403 959 COFs.[206] TPOT optimized the model parameters using genetic algorithms, and outperformed other traditional ML models (RF, SVM, etc.).

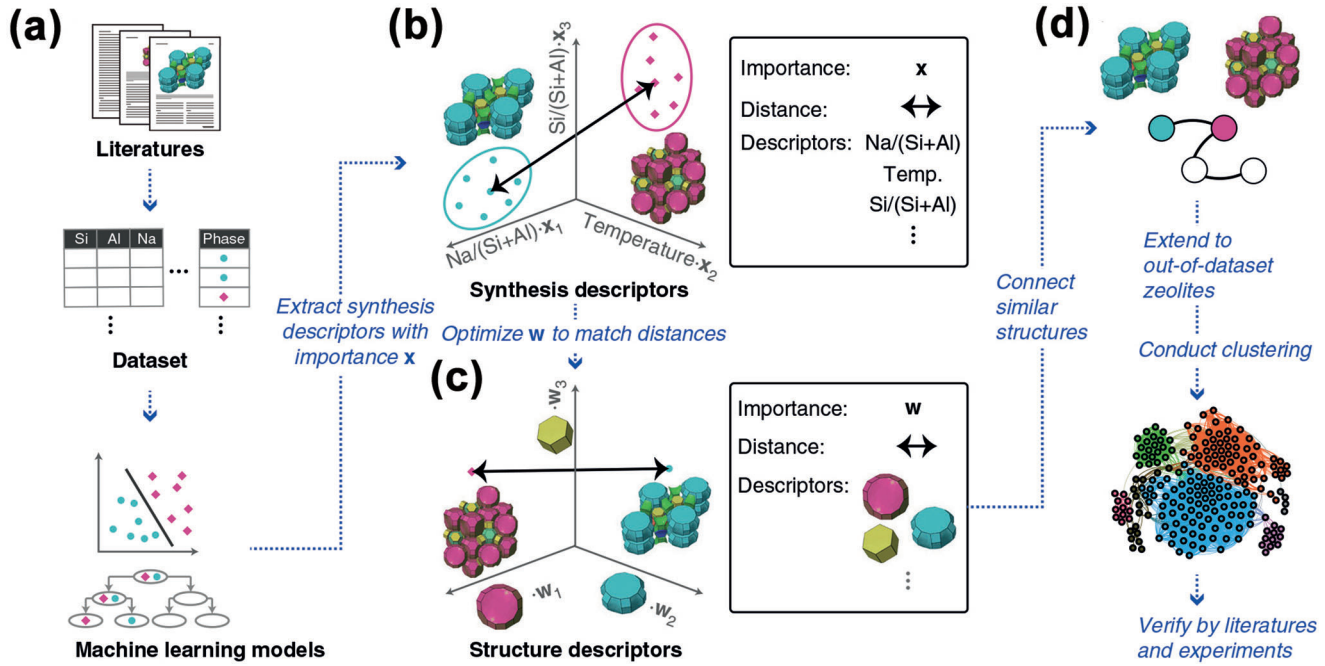


Figure 9. Workflow to link synthesis parameters to structure features in zeolites. a) ML models were constructed from literature data. b) Synthesis parameters mapped the synthesizable domains of zeolites onto a multidimensional phase diagram. x_i encoded the importance of each synthesis parameter assessed by the ML models. The synthesis similarity was represented by the distance between the centers of the synthesis conditions for each phase. c) Structure features defined the structural similarity in a multidimensional space representing the presence or absence of building units. d) A network was constructed by connecting structurally similar zeolites based on the structure features. The resulting clustering was verified with data in the literature and experiments. Reproduced with permission.[202] Copyright 2019, Springer Nature.

Fanourgakis et al. studied the performance of models of the CH_4 uptake trained on 69 840 COFs and 4763 MOFs.[207] Their results showed that the use of relevant materials features resulted in excellent predictivity for materials properties when models were trained on a small subset of the training data rather than the entire training set (**Figure 10**). This approach could significantly reduce the computational cost of the construction of the training set using expensive physics-based methods.

Porous polymer networks (PPNs) are another new type of adsorbent material. They possess a reticular structure with robust covalent bonds of organic linkers. They exhibit superior surface areas and much better stability than MOFs, making them popular choices for gas storage and separation.[208–210] Pardakhti et al. reported quantitative relationships between the chemical features and CH_4 uptake of PPNs using RF models trained on 17 846 PPNs.[211] Chemical features such as number and type of atoms, electronegativity and degree of unsaturation, played important roles in the CH_4 uptake under low pressure, while physical features such as surface area and void fraction dominated the adsorption under high pressure. This study highlighted the contributions of surface atoms to gas adsorption that are helpful for adsorbent materials screening and design.

Intermetallics are important materials for gas storage, particularly hydrogen.[212] Jäger et al. used a KRR model with local structural descriptors (e.g., SOAP, ACSF) to accurately predict the hydrogen adsorption energy on an Au-Cu alloy surface. However, construction of local structural descriptors was complex.[213] Witman et al. developed a GBT model using the descriptors derived from intermetallic composition only, rather than any structural or hydride information, to predict the log equilibrium pressure of H_2, $\ln P_{eq}$.[214] 145 compositional descriptors were used to train the model, and descriptor relevance analysis identified the specific volume per atom for a given composition as most important. Since this has limited physical interpretability, a new descriptor encoding the volume per atom in a crystal was generated and a similar relationship with $\ln P_{eq}$ was confirmed. This ML model enabled researchers to predict the hydrogen storage capacity of intermetallic materials from compositional information. ML models have also been employed to predict the CO adsorption energy of a thiolated Au-Ag nanoalloy. It was found that the CO adsorption largely depended on structural features of the Au-Ag alloy, and the ML model allowed very fast screening of candidates for further analysis.[94]

4. Summary and Outlook

ML techniques are becoming invaluable for adsorbent materials discovery and design. Coupled with resource-intensive DFT and GCMC calculations and experiments, ML has robustly and effectively predicted gas uptake, discovered unknown feature–adsorption relations, and optimized synthesis conditions. Many ML models have achieved high accuracy, comparable to first principles calculations, and can elucidate complex feature–property relations efficiently given sufficient reliable data. The use of ML methods allows rapid exploration of large material spaces, provides a rational basis for material design with bespoke properties, and provides new physical insights from large and complex datasets. **Table 2** summarizes the adsorbents discovered

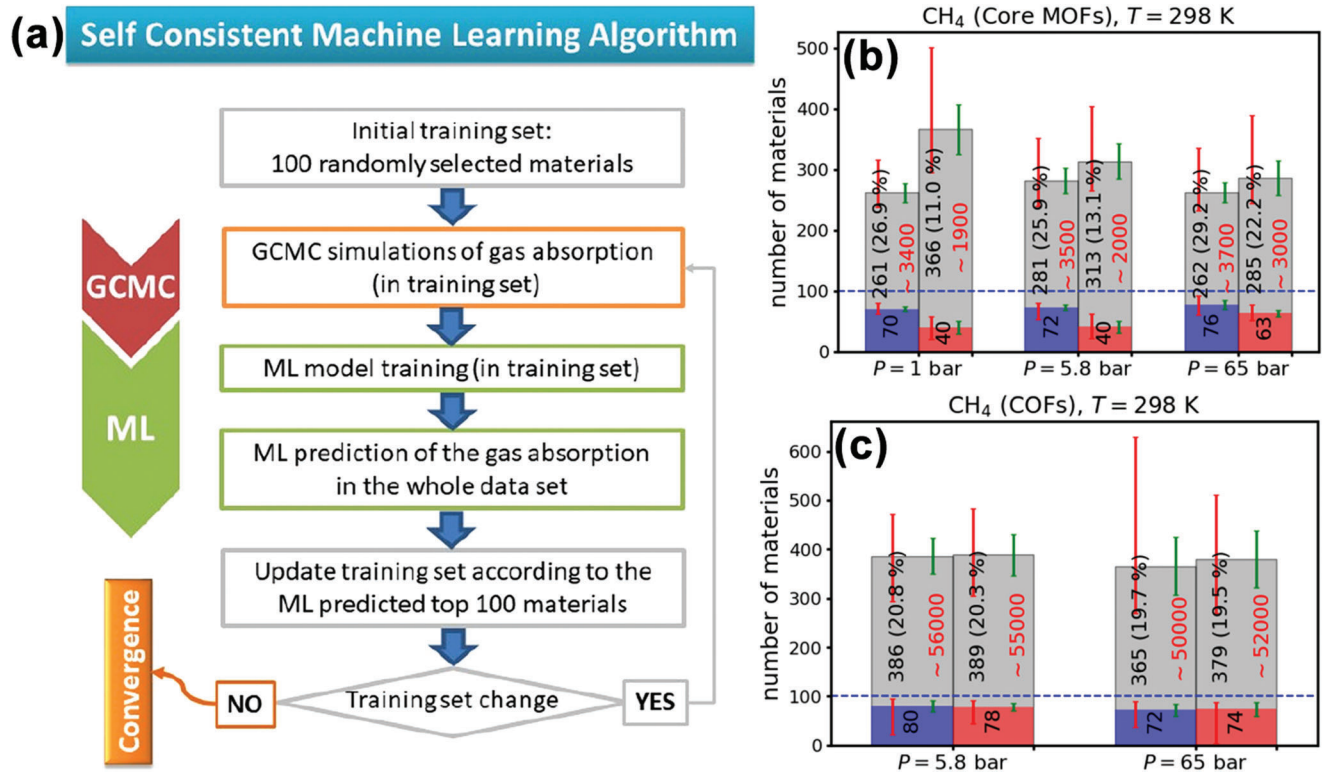


Figure 10. a) Flowchart of the self-consistent ML approach. From the top of the flowchart, it can be seen that this approach started from randomly picking 100 materials from a database as the initial training set. After labeling by GCMC, a ML model was trained on this training set. The ML model was then used to predict the gas adsorption of all materials in the database, and the top 100 materials would be selected as the new training set until the top 100 materials were the same as the those in the training set. Average results for the b) MOFs and c) COFs obtained from 100 individual runs. Each pair of bars corresponded to calculations with the "high" (blue bar) and the "low" (red bar) accuracy features, respectively. The average number of MOFs and COFs with top performance found during the 100 individual runs was denoted inside these bars. The gray bars show the average number of the total structures included in the final training set. Inside these bars, the number of structures is denoted in black, together with the percentage of the top performing structures found in the final training set. The approximate number of structures required to be randomly selected from the original training set was denoted in red. The red error bar showed the minimum and the maximum value found during the 100 individual runs. The corresponding standard deviation was shown with a green error bar. Reproduced with permission.[207] Copyright 2020, American Chemical Society.

Table 2. The adsorbent materials with leading gas uptake capacity discovered by ML techniques.

Adsorbents	Year of discovery	ML features	ML algorithms	Gas	Results	Note	Ref.
MOF (qtz-sym-4-mc-Si-L2)	2019	Topographical features	NN	H_2	Uptake capacity: 62 g L^{-1} under 100 bar/77 K to 5 bar/160 K	Simulated value. The highest deliverable capacity of H_2 that can be attained without extreme pressure conditions	[153]
MOF (MFU-4l (Zn))	2019	Energy histogram, structural features	LASSO	H_2	Uptake capacity: 47 g L^{-1} under 100 bar/77 K to 5 bar/160 K	Experimental value	[71]
MOF (DUT-23 (Cu))	2022	Topographical features	GBR	CH_4	Uptake capacity: 373 cm^3 (STP) cm^{-3} under 250 bar/120 K to 65 bar/298 K	Experimental value	[215]
MOF (MIL-47)	2016	Structural features	Genetic algorithm	CO_2	Uptake capacity: ≈4 mmol g^{-1} at 0.15 atm/298 K 0	Experimental value	[160]
MOF (Al-PMOF)	2019	Structural features	ML assisted data mining	CO_2	Uptake capacity: 6 mmol g^{-1} at 2000 mbar/313 K	Experimental value	[170]

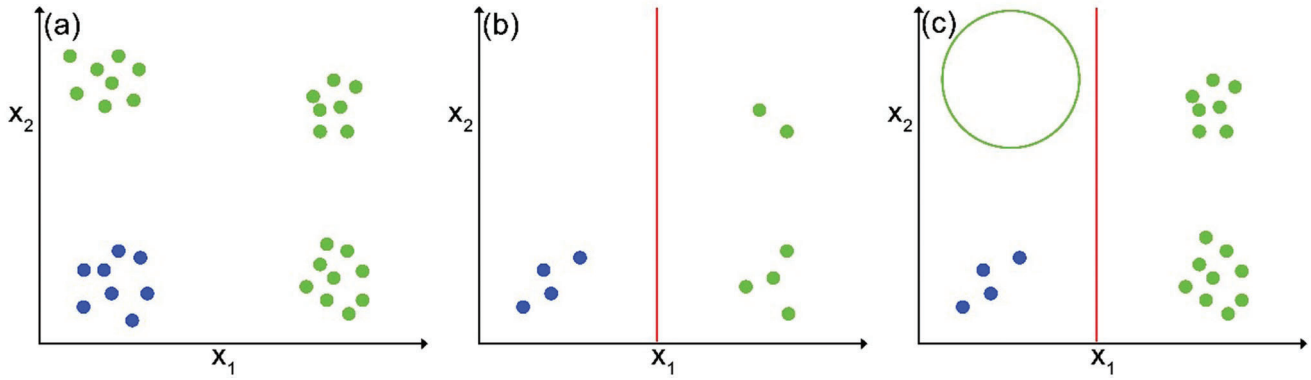


Figure 11. a) The true distribution of the green dots and blue dots in a space, where the green dots are our target. b) On the basis of the initial training set, only a single feature (x_1) may appear important, resulting in a hyperplane perpendicular to the x_1-axis (red line). c) Once the hyperplane is learnt, the continuous discovery of green dots on the right side of the hyperplane may strengthen the confidence of the model to look for items on the right side of the hyperplane, and thus a correct hyperplane is never determined.

or designed by ML techniques to have substantially improved gas uptake capacity. The application of ML to adsorbent science and engineering is an important step to fast-track the discovery and optimization of adsorbent materials to address climate change challenges.

The studies summarized in this review show how ML accelerates material development; however, many of the outputs were hypothetical materials. These adsorbent materials may have complicated synthesis procedures or synthesis may not be possible at all. To increase the likelihood of successful synthesis of the adsorbent materials proposed by ML property models, training data and screening scope could also be restricted to materials that have previously been synthesized.[89] However, new ML models have now also made significant inroads into predicting the synthesizability of porous materials.[202] The results reviewed here also emphasize the importance of close collaboration between computer scientists and experimental experts, both to provide the essential data for training models, but also to allow predictions of ML models to be tested experimentally. Clearly, models are much more useful and generate greater confidence when their predictions are subject to experimental validation. Thus, computational and experimental researchers should work together from the very beginning of projects to establish the ML strategies that achieve materials with optimal properties for a given application.

A major challenge limiting the application of ML to the development of adsorbent materials is the size, range, and quality of the dataset. Despite the rise in the number of porous materials databases, collecting calculated and experimental data labeled by target properties (e.g., gas uptake, selectivity, mechanic properties) is expensive and time-consuming. Therefore, it is important to develop techniques for generating reliable ML models from small samples, especially using high-throughput and robotic methods. In this review, some cutting-edge solutions have been described, such as active learning, transfer learning and meta learning, that have been applied to address this issue for adsorbent material studies. However, care must be taken to avoid the attentional learning trap and biases depicted in **Figure 11**.[216] Avoiding this issue requires the involvement of a human operator who, for example, can reduce the rewards in reinforcement learning to force the model to explore new routes, or add different rewards. The limited ability to predict outside the domain of the training data is another limitation, however this will be ameliorated by the increasing availability of data from high-throughput experiments and computation. Moreover, by identifying an optimally sparse subset of relevant features, overfitting can be avoided when robust models are built using relatively small datasets, and model interpretation simplified. Recently, new techniques based on evolutionary algorithms, such as symbolic learning[217] and the sure independence screening and sparsifying operator (SISSO),[218] have been developed to generate informative features from large feature pools. The generation of meaningful features is essential to generate robust and predictive models that can usefully guide material development and optimization.

In addition to small sample techniques and feature generation, training data can be expanded by data sources other than databases. Experimental and computational information in the literature can be batch extracted by text mining techniques. Supervised natural language techniques and unsupervised word embedding techniques have been employed to capture the knowledge in the materials science literature.[43,89,219,220] However, the sparsity and inhomogeneity of the experimental information from diverse literature sources limits their use for ML model construction. The reproducibility of experimental and computational information is another obstacle to compiling data from heterogeneous sources. Experimental data ideally should be collected by conducting experiments under the same experimental setup and conditions. However, it has been reported replicated syntheses of MOF adsorbent materials is low.[221] Some key features of adsorbent materials, such as surface area, are difficult to reproduce because of differences in calculation approaches and ambiguities in molecular structure.[222] Eliminating these issues requires authors to provide additional metadata in their publications of synthesis and characterization methods, thereby ensuring the quality and reproducibility of the reported material. Further development and wider use of materials ontologies should also assist in improving reproducibility of syntheses and experimental characterization of adsorbent (and other) materials. Likewise, the reproducibility of the computational information should also be guaranteed by providing open access to the data, input and output

files, and the software or codes used for computation.[223] In addition, high throughput experiments promise to generate large quantities of data for specific materials systems,[32,224,225] but we stress that experimental data on poorly performing materials are also valuable for training the most robust ML models.[40] An interesting and very recent development is autonomous laboratories that merge ML techniques with robotics,[30,128,226,227] where synthesis and characterization are carried out without human intervention. This is a potentially valuable future solution to collecting high-quality experimental data on the large scale and autonomously discovering potential adsorbent materials (e.g., porous perovskites,[228] porous spinel[229,230]) with multiple favorable properties (e.g., gases or ions adsorption ability, porosity, selectivity, synthesizability, stability, cost) simultaneously. Further integration of ML, materials science, and engineering will accelerate adsorbent material discovery and find solutions for energy diversification and for combatting climate change.

Acknowledgements

The authors acknowledge the support from the Australian Research Council (ARC) in the form of Discovery Project DP180103815 and DP220100945, and the support through RMIT Sustainable Development Research Grants.

Conflict of Interest

The authors declare no conflict of interest.

Keywords

covalent–organic frameworks, hydrogen, intermetallics, metal–organic frameworks, porous carbons, porous polymers networks, zeolites

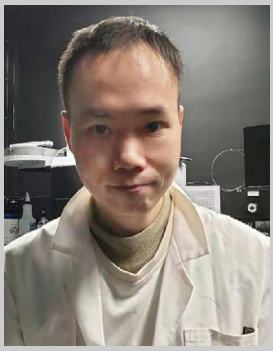
Received: July 7, 2022
Revised: September 27, 2022
Published online: October 26, 2022

[1] UNFCCC, Paris, France **2015**.
[2] J. Rogelj, M. den Elzen, N. Höhne, T. Fransen, H. Fekete, H. Winkler, R. Schaeffer, F. Sha, K. Riahi, M. Meinshausen, *Nature* **2016**, *534*, 631.
[3] M. Meinshausen, N. Meinshausen, W. Hare, S. C. B. Raper, K. Frieler, R. Knutti, D. J. Frame, M. R. Allen, *Nature* **2009**, *458*, 1158.
[4] B. Obama, *Science* **2017**, *355*, 126.
[5] S. Carley, D. M. Konisky, *Nat. Energy* **2020**, *5*, 569.
[6] S. van Renssen, *Nat. Clim. Change* **2020**, *10*, 799.
[7] R. L. Martin, C. M. Simon, B. Smit, M. Haranczyk, *J. Am. Chem. Soc.* **2014**, *136*, 5006.
[8] H. Furukawa, K. E. Cordova, M. O'Keeffe, O. M. Yaghi, *Science* **2013**, *341*, 974.
[9] S. Yuan, L. Feng, K. Wang, J. Pang, M. Bosch, C. Lollar, Y. Sun, J. Qin, X. Yang, P. Zhang, Q. Wang, L. Zou, Y. Zhang, L. Zhang, Y. Fang, J. Li, H. C. Zhou, *Adv. Mater.* **2018**, *30*, 1704303.
[10] H. Barthelemy, M. Weber, F. Barbier, *Int. J. Hydrogen Energy* **2017**, *42*, 7254.
[11] D. J. Durbin, C. Malardier-Jugroot, *Int. J. Hydrogen Energy* **2013**, *38*, 14595.
[12] C. Kunze, H. Spliethoff, *Appl. Energy* **2012**, *94*, 109.
[13] R. Ben-Mansour, M. A. Habib, O. E. Bamidele, M. Basha, N. A. A. Qasem, A. Peedikakkal, T. Laoui, M. Ali, *Appl. Energy* **2016**, *161*, 225.
[14] M. G. Plaza, S. García, F. Rubiera, J. J. Pis, C. Pevida, *Chem. Eng. J.* **2010**, *163*, 41.
[15] X. Yang, R. J. Rees, W. Conway, G. Puxty, Q. Yang, D. A. Winkler, *Chem. Rev.* **2017**, *117*, 9524.
[16] M. Zhao, Y. Huang, Y. Peng, Z. Huang, Q. Ma, H. Zhang, *Chem. Soc. Rev.* **2018**, *47*, 6267.
[17] J. F. Olorunyomi, S. T. Geh, R. A. Caruso, C. M. Doherty, *Mater. Horiz.* **2021**, *8*, 2387.
[18] S.-Y. Ding, W. Wang, *Chem. Soc. Rev.* **2013**, *42*, 548.
[19] K. Geng, T. He, R. Liu, S. Dalapati, K. T. Tan, Z. Li, S. Tao, Y. Gong, Q. Jiang, D. Jiang, *Chem. Rev.* **2020**, *120*, 8814.
[20] M. R. Benzigar, S. N. Talapaneni, S. Joseph, K. Ramadass, G. Singh, J. Scaranto, U. Ravon, K. Al-Bahily, A. Vinu, *Chem. Soc. Rev.* **2018**, *47*, 2680.
[21] Y. Li, L. Li, J. Yu, *Chem* **2017**, *3*, 928.
[22] X. Yu, Z. Tang, D. Sun, L. Ouyang, M. Zhu, *Prog. Mater Sci.* **2017**, *88*, 1.
[23] Y. Yan, T. N. Borhani, S. G. Subraveti, K. N. Pai, V. Prasad, A. Rajendran, P. Nkulikiyinka, J. O. Asibor, Z. Zhang, D. Shao, L. Wang, W. Zhang, Y. Yan, W. Ampomah, J. You, M. Wang, E. J. Anthony, V. Manovic, P. T. Clough, *Energy Environ. Sci.* **2021**, *14*, 6122.
[24] K. M. Jablonka, D. Ongari, S. M. Moosavi, B. Smit, *Chem. Rev.* **2020**, *120*, 8066.
[25] S. M. Moosavi, K. M. Jablonka, B. Smit, *J. Am. Chem. Soc.* **2020**, *142*, 20273.
[26] H. Mai, T. C. Le, D. Chen, D. A. Winkler, R. A. Caruso, *Chem. Rev.* **2022**, *122*, 13478.
[27] M. I. Jordan, T. M. Mitchell, *Science* **2015**, *349*, 255.
[28] K. T. Butler, D. W. Davies, H. Cartwright, O. Isayev, A. Walsh, *Nature* **2018**, *559*, 547.
[29] E. N. Muratov, R. Amaro, C. H. Andrade, N. Brown, S. Ekins, D. Fourches, O. Isayev, D. Kozakov, J. L. Medina-Franco, K. M. Merz, T. I. Oprea, V. Poroikov, G. Schneider, M. H. Todd, A. Varnek, D. A. Winkler, A. V. Zakharov, A. Cherkasov, A. Tropsha, *Chem. Soc. Rev.* **2021**, *50*, 9121.
[30] D. P. Tabor, L. M. Roch, S. K. Saikin, C. Kreisbeck, D. Sheberla, J. H. Montoya, S. Dwaraknath, M. Aykol, C. Ortiz, H. Tribukait, C. Amador-Bedolla, C. J. Brabec, B. Maruyama, K. A. Persson, A. Aspuru-Guzik, *Nat. Rev. Mater.* **2018**, *3*, 5.
[31] C. Chen, Y. X. Zuo, W. K. Ye, X. G. Li, Z. Deng, S. P. Ong, *Adv. Energy Mater.* **2020**, *10*, 1903242.
[32] I. G. Clayson, D. Hewitt, M. Hutereau, T. Pope, B. Slater, *Adv. Mater.* **2020**, *32*, 2002780.
[33] W. Chaikittisilp, Y. Yamauchi, K. Ariga, *Adv. Mater.* **2022**, *34*, 2107212.
[34] M. Opanasenko, M. Shamzhy, Y. Wang, W. Yan, P. Nachtigall, J. Čejka, *Angew. Chem., Int. Ed.* **2020**, *59*, 19380.
[35] C. Altintas, O. F. Altundal, S. Keskin, R. Yildirim, *J. Chem. Inf. Model.* **2021**, *61*, 2131.
[36] T. Le, V. C. Epa, F. R. Burden, D. A. Winkler, *Chem. Rev.* **2012**, *112*, 2889.
[37] Y. Zhang, C. Ling, *npj Comput. Mater.* **2018**, *4*, 25.
[38] G. Lo Dico, Á. P. Nuñez, V. Carcelén, M. Haranczyk, *Chem. Sci.* **2021**, *12*, 9309.
[39] C. W. Coley, W. H. Green, K. F. Jensen, *Acc. Chem. Res.* **2018**, *51*, 1281.
[40] P. Raccuglia, K. C. Elbert, P. D. F. Adler, C. Falk, M. B. Wenny, A. Mollo, M. Zeller, S. A. Friedler, J. Schrier, A. J. Norquist, *Nature* **2016**, *533*, 73.
[41] S. R. Kalidindi, M. De Graef, in *Annual Review of Materials Research*, Vol. *45* (Ed: D. R. Clarke), **2015**, p. 171.

[42] M. Krallinger, O. Rabal, A. Lourenco, J. Oyarzabal, A. Valencia, *Chem. Rev.* **2017**, *117*, 7673.
[43] A. Nandy, C. Duan, H. J. Kulik, *J. Am. Chem. Soc.* **2021**, *143*, 17535.
[44] J. Zhou, D. Liu, M. Ye, Z. Liu, *Ind. Eng. Chem. Res.* **2021**, *60*, 13727.
[45] A. Gaulton, L. J. Bellis, A. P. Bento, J. Chambers, M. Davies, A. Hersey, Y. Light, S. Mcglinchey, D. Michalovich, B. Al-Lazikani, J. P. Overington, *Nucleic Acids Res.* **2012**, *40*, D1100.
[46] L. C. Blum, J.-L. Reymond, *J. Am. Chem. Soc.* **2009**, *131*, 8732.
[47] L. Ruddigkeit, R. van Deursen, L. C. Blum, J.-L. Reymond, *J. Chem. Inf. Model.* **2012**, *52*, 2864.
[48] J. J. Irwin, B. K. Shoichet, *J. Chem. Inf. Model.* **2005**, *45*, 177.
[49] S. Curtarolo, W. Setyawan, S. Wang, J. Xue, K. Yang, R. H. Taylor, L. J. Nelson, G. L. W. Hart, S. Sanvito, M. Buongiorno-Nardelli, N. Mingo, O. Levy, *Comput. Mater. Sci.* **2012**, *58*, 227.
[50] F. H. Allen, G. P. Shields, in *Implications of Molecular and Materials Structure for New Technologies*, Vol. *360* (Eds: J. A. K. Howard, F. H. Allen, G. P. Shields), **1999**, p. 291.
[51] A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, K. A. Persson, *APL Mater.* **2013**, *1*, 011002.
[52] C. Draxl, M. Scheffler, *MRS Bull.* **2018**, *43*, 676.
[53] P. Z. Moghadam, A. Li, S. B. Wiggin, A. Tao, A. G. P. Maloney, P. A. Wood, S. C. Ward, D. Fairen-Jimenez, *Chem. Mater.* **2017**, *29*, 2618.
[54] Y. G. Chung, E. Haldoupis, B. J. Bucior, M. Haranczyk, S. Lee, H. Zhang, K. D. Vogiatzis, M. Milisavljevic, S. Ling, J. S. Camp, B. Slater, J. I Siepmann, D. S. Sholl, R. Q. Snurr, *J. Chem. Eng. Data* **2019**, *64*, 5985.
[55] D. J. Earl, M. W. Deem, *Ind. Eng. Chem. Res.* **2006**, *45*, 5449.
[56] D. Siderius, V. Shen, R. Johnson, R. D. van Zee, *NIST/ARPA-E Database of Novel and Emerging Adsorbent Materials, National Institute of Standards and Technology*, **2020**, https://doi.org/10.18434/T43882
[57] Metal Organic Framework Database, **2022**, https://mof.tech.northwestern.edu/
[58] G. Raman, *ChemistrySelect* **2021**, *6*, 10661.
[59] J. Ohyama, A. Hirayama, N. Kondou, H. Yoshida, M. Machida, S. Nishimura, K. Hirai, I. Miyazato, K. Takahashi, *Stem Cells Int.* **2021**, *11*, 2067.
[60] R. Wang, Y. Zou, C. Zhang, X. Wang, M. Yang, D. Xu, *Microporous Mesoporous Mater.* **2022**, *331*, 111666.
[61] M. Rupp, A. Tkatchenko, K. R. Muller, O. A. von Lilienfeld, *Phys. Rev. Lett.* **2012**, *108*, 058301.
[62] J. Behler, M. Parrinello, *Phys. Rev. Lett.* **2007**, *98*, 146401.
[63] D. Weininger, *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 31.
[64] M. Pinheiro, R. L. Martin, C. H. Rycroft, M. Haranczyk, *CrystEngComm* **2013**, *15*, 7531.
[65] A. P. Bartok, S. De, C. Poelking, N. Bernstein, J. R. Kermode, G. Csanyi, M. Ceriotti, *Sci. Adv.* **2017**, *3*, e1701816.
[66] N. Kim, K. Min, *J. Phys. Chem. Lett.* **2021**, *12*, 2334.
[67] I. B. Orhan, H. Daglar, S. Keskin, T. C. Le, R. Babarao, *ACS Appl. Mater. Interfaces* **2022**, *14*, 736.
[68] X. Yu, S. Choi, D. Tang, A. J. Medford, D. S. Sholl, *J. Phys. Chem. C* **2021**, *125*, 18046.
[69] X. Wu, S. Xiang, J. Su, W. Cai, *J. Phys. Chem. C* **2019**, *123*, 8550.
[70] C. M. Simon, R. Mercado, S. K. Schnell, B. Smit, M. Haranczyk, *Chem. Mater.* **2015**, *27*, 4459.
[71] B. J. Bucior, N. S. Bobbitt, T. Islamoglu, S. Goswami, A. Gopalan, T. Yildirim, O. K. Farha, N. Bagheri, R. Q. Snurr, *Mol. Syst. Des. Eng.* **2019**, *4*, 162.
[72] G. S. Fanourgakis, K. Gkagkas, E. Tylianakis, E. Klontzas, G. Froudakis, *J. Phys. Chem. A* **2019**, *123*, 6080.
[73] R. Gurnani, Z. Yu, C. Kim, D. S. Sholl, R. Ramprasad, *Chem. Mater.* **2021**, *33*, 3543.
[74] X. Zhu, D. C. W. Tsang, L. Wang, Z. Su, D. Hou, L. Li, J. Shang, *J. Cleaner Prod.* **2020**, *273*, 122915.
[75] A. Ahmed, D. J. Siegel, *Patterns* **2021**, *2*, 100291.
[76] M. Fernandez, N. R. Trefiak, T. K. Woo, *J. Phys. Chem. C* **2013**, *117*, 14095.
[77] S. Amani, A. B. Garmarudi, M. Khanmohammadi, F. Yaripour, *RSC Adv.* **2018**, *8*, 34830.
[78] M. Fernandez, A. S. Barnard, *ACS Comb. Sci.* **2016**, *18*, 243.
[79] H. Hotelling, *J. Educ. Psychol.* **1933**, *24*, 498.
[80] L. van der Maaten, G. Hinton, *J. Mach. Learn. Res.* **2008**, *9*, 2579.
[81] L. McInnes, J. Healy, J. Melville, *ArXiv* **2020**, 180203426v3.
[82] M. Suyetin, *Faraday Discuss.* **2021**, *231*, 224.
[83] Y. He, E. D. Cubuk, M. D. Allendorf, E. J. Reed, *J. Phys. Chem. Lett.* **2018**, *9*, 4562.
[84] S. M. Moosavi, A. Nandy, K. M. Jablonka, D. Ongari, J. P. Janet, P. G. Boyd, Y. Lee, B. Smit, H. J. Kulik, *Nat. Commun.* **2020**, *11*, 4068.
[85] B. A. Helfrecht, R. K. Cersonsky, G. Fraux, M. Ceriotti, *Mach. Learn. Sci. Technol.* **2020**, *1*, 045021.
[86] X. Yuan, L. Li, Z. Shi, H. Liang, S. Li, Z. Qiao, *Adv. Powder Mater.* **2021**, *1*, 100026.
[87] M. Rahimi, M. H. Abbaspour-Fard, A. Rohani, *J. Cleaner Prod.* **2021**, *329*, 129714.
[88] J. Abdi, M. Hadipoor, F. Hadavimoghaddam, A. Hemmati-Sarapardeh, *Chemosphere* **2022**, *287*, 132135.
[89] Z. Jensen, S. Kwon, D. Schwalbe-Koda, C. Paris, R. Gómez-Bombarelli, Y. Román-Leshkov, A. Corma, M. Moliner, E. A. Olivetti, *ACS Cent. Sci.* **2021**, *7*, 858.
[90] A. W. Thornton, D. A. Winkler, M. S. Liu, M. Haranczyk, D. F. Kennedy, *RSC Adv.* **2015**, *5*, 44361.
[91] A. W. Thornton, C. M. Simon, J. Kim, O. Kwon, K. S. Deeg, K. Konstas, S. J. Pas, M. R. Hill, D. A. Winkler, M. Haranczyk, B. Smit, *Chem. Mater.* **2017**, *29*, 2844.
[92] R. Ma, Y. J. Colón, T. Luo, *ACS Appl. Mater. Interfaces* **2020**, *12*, 34041.
[93] T. C. Nicholas, A. L. Goodwin, V. L. Deringer, *Chem. Sci.* **2020**, *11*, 12580.
[94] G. Panapitiya, G. Avendano-Franco, P. J. Ren, X. D. Wen, Y. W. Li, J. P. Lewis, *J. Am. Chem. Soc.* **2018**, *140*, 17508.
[95] V. L. Deringer, M. A. Caro, G. Csanyi, *Adv. Mater.* **2019**, *31*, 1902765.
[96] M. Eckhoff, J. Behler, *J. Chem. Theory Comput.* **2019**, *15*, 3793.
[97] K. Esfandiari, A. A. Ghoreyshi, M. Jahanshahi, *Ind. Eng. Chem. Res.* **2017**, *56*, 14610.
[98] Z. Yıldız, H. Uzun, *Microporous Mesoporous Mater.* **2015**, *208*, 50.
[99] D. Wu, Q. Yang, C. Zhong, D. Liu, H. Huang, W. Zhang, G. Maurin, *Langmuir* **2012**, *28*, 12094.
[100] H. Dureckova, M. Krykunov, M. Z. Aghaji, T. K. Woo, *J. Phys. Chem. C* **2019**, *123*, 4133.
[101] M. Moliner, J. M. Serra, A. Corma, E. Argente, S. Valero, V. Botti, *Microporous Mesoporous Mater.* **2005**, *78*, 73.
[102] S. Lee, B. Kim, J. Kim, *J. Mater. Chem. A* **2019**, *7*, 2709.
[103] J. Schmidhuber, *Neural Networks* **2015**, *61*, 85.
[104] Y. LeCun, Y. Bengio, G. Hinton, *Nature* **2015**, *521*, 436.
[105] O. C. Yolcu, F. A. Temel, A. Kuleyin, *J. Cleaner Prod.* **2021**, *311*, 127688.
[106] Y. Yu, W. Zhang, D. Mei, *J. Phys. Chem. C* **2022**, *126*, 1204.
[107] J. P. Torres, R. T. Codorniu, R. L. Baracaldo, H. C. Sariol, T. M. Peacok, J. Yperman, P. Adriaensens, R. Carleer, Á. B. Sauvanell, *SN Appl. Sci.* **2020**, *2*, 2088.
[108] A. Rabbani, M. Babaei, R. Shams, Y. D. Wang, T. Chung, *Adv. Water Resour.* **2020**, *146*, 103787.
[109] Y. Wang, Z. Cao, A. B. Farimani, *npj 2D Mater. Appl.* **2021**, *5*, 66.
[110] T.-H. Hung, Z.-X. Xu, D.-Y. Kang, L.-C. Lin, *J. Phys. Chem. C* **2022**, *126*, 2813.
[111] X. Lu, Z. Xie, X. Wu, M. Li, W. Cai, *Chem. Eng. Sci.* **2022**, *259*, 117813.
[112] A. Golbraikh, A. Tropsha, *J. Mol. Graphics Modell.* **2002**, *20*, 269.

[113] D. L. J. Alexander, A. Tropsha, D. A. Winkler, *J. Chem. Inf. Model.* **2015**, *55*, 1316.
[114] L. Breiman, *Mach. Learn.* **2001**, *45*, 5.
[115] S. M. Lundberg, S. I. Lee, in*Advances in Neural Information Processing Systems 30 (NIPS 2017)*, (Eds: I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, R. Garnett), Neural Information Processing Systems Foundation, Inc. (NeurIPS), Long Beach CA, USA **2017**.
[116] V. V. Korolev, A. Mitrofanov, E. I. Marchenko, N. N. Eremin, V. Tkachenko, S. N. Kalmykov, *Chem. Mater.* **2020**, *32*, 7822.
[117] S. H. Wang, X. Y. Xue, M. Cheng, S. C. Chen, C. Liu, L. Zhou, K. X. Bi, X. Ji, *Acta Chim. Sin.* **2022**, *80*, 614.
[118] M. I. M. Kusdhany, S. M. Lyth, *Carbon* **2021**, *179*, 190.
[119] B. Zhou, A. Khosla, À. Lapedriza, A. Oliva, A. Torralba, in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Las Vegas NV, USA **2016**, p. 2921.
[120] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, I. Polosukhin, presented at *31st Annual Conf. Neural Information Processing Systems (NIPS)*, XX, Long Beach, CA, December **2017**.
[121] M. Chai, S. Moradi, E. Erfani, M. Asadnia, V. Chen, A. Razmjou, *Chem. Mater.* **2021**, *33*, 8666.
[122] F. Oviedo, J. L. Ferres, T. Buonassisi, K. T. Butler, *Acc. Mater. Res.* **2022**, *3*, 597.
[123] E. Ren, P. Guilbaud, F.-X. Coudert, *Digital Discovery* **2022**, *1*, 355.
[124] G. H. Gu, J. Noh, I. Kim, Y. Jung, *J. Mater. Chem. A* **2019**, *7*, 17096.
[125] R. Ramprasad, R. Batra, G. Pilania, A. Mannodi-Kanakkithodi, C. Kim, *npj Comput. Mater.* **2017**, *3*, 54.
[126] T. C. Le, D. A. Winkler, *Chem. Rev.* **2016**, *116*, 6107.
[127] J. S. Smith, B. T. Nebgen, R. Zubatyuk, N. Lubbers, C. Devereux, K. Barros, S. Tretiak, O. Isayev, A. E. Roitberg, *Nat. Commun.* **2019**, *10*, 2903.
[128] J. M. Granda, L. Donina, V. Dragone, D. L. Long, L. Cronin, *Nature* **2018**, *559*, 377.
[129] H. Y. Huo, Z. Q. Rong, O. Kononova, W. H. Sun, T. Botari, T. J. He, V. Tshitoyan, G. Ceder, *npj Comput. Mater.* **2019**, *5*, 62.
[130] H. Furukawa, K. E. Cordova, M. O'Keeffe, O. M. Yaghi, *Science* **2013**, *341*, 1230444.
[131] Z. C. Hu, B. J. Deibert, J. Li, *Chem. Soc. Rev.* **2014**, *43*, 5815.
[132] B. Y. Xia, Y. Yan, N. Li, H. B. Wu, X. W. Lou, X. Wang, *Nat. Energy* **2016**, *1*, 15006.
[133] K. Adil, Y. Belmabkhout, R. S. Pillai, A. Cadiau, P. M. Bhatt, A. H. Assen, G. Maurin, M. Eddaoudi, *Chem. Soc. Rev.* **2017**, *46*, 3402.
[134] J. F. Olorunyomi, M. M. Sadiq, M. Batten, K. Konstas, D. Chen, C. M. Doherty, R. A. Caruso, *Adv. Opt. Mater.* **2020**, *8*, 2000961.
[135] C. E. Wilmer, M. Leaf, C. Y. Lee, O. K. Farha, B. G. Hauser, J. T. Hupp, R. Q. Snurr, *Nat. Chem.* **2012**, *4*, 83.
[136] D. Banerjee, C. M. Simon, A. M. Plonka, R. K. Motkuri, J. Liu, X. Y. Chen, B. Smit, J. B. Parise, M. Haranczyk, P. K. Thallapally, *Nat. Commun.* **2016**, *7*, 7.
[137] M. Witman, S. L. Ling, S. Jawahery, P. G. Boyd, M. Haranczyk, B. Slater, B. Smit, *J. Am. Chem. Soc.* **2017**, *139*, 5547.
[138] S. Henke, R. Schmid, J.-D. Grunwaldt, R. A. Fischer, *Chem. - Eur. J.* **2010**, *16*, 14296.
[139] Y. G. Chung, D. A. Gómez-Gualdrón, P. Li, K. T. Leperi, P. Deria, H. Zhang, N. A. Vermeulen, J. F Stoddart, F. You, J. T. Hupp, O. K. Farha, R. Q. Snurr, *Sci. Adv.* **2016**, *2*, e1600909.
[140] L. Grajciar, C. J. Heard, A. A. Bondarenko, M. V. Polynski, J. Meeprasert, E. A. Pidko, P. Nachtigall, *Chem. Soc. Rev.* **2018**, *47*, 8307.
[141] L. B. Vilhelmsen, K. S. Walton, D. S. Sholl, *J. Am. Chem. Soc.* **2012**, *134*, 12807.
[142] W. Li, X. X. Xia, S. Li, *J. Mater. Chem. A* **2019**, *7*, 25010.
[143] S. Chong, S. Lee, B. Kim, J. Kim, *Coord. Chem. Rev.* **2020**, *423*, 213487.
[144] R. Long, X. Xia, Y. Zhao, S. Li, Z. Liu, W. Liu, *iScience* **2021**, *24*, 101914.
[145] H.-C. Zhou, J. R. Long, O. M. Yaghi, *Chem. Rev.* **2012**, *112*, 673.
[146] M. Fernandez, P. G. Boyd, T. D. Daff, M. Z. Aghaji, T. K. Woo, *J. Phys. Chem. Lett.* **2014**, *5*, 3056.
[147] B. Smit, T. L. M. Maesen, *Chem. Rev.* **2008**, *108*, 4125.
[148] J. Burner, L. Schwiedrzik, M. Krykunov, J. Luo, P. G. Boyd, T. K. Woo, *J. Phys. Chem. C* **2020**, *124*, 27996.
[149] R. Anderson, J. Rodgers, E. Argueta, A. Biong, D. A. Gómez-Gualdrón, *Chem. Mater.* **2018**, *30*, 6325.
[150] M. Pardakhti, E. Moharreri, D. Wanik, S. L. Suib, R. Srivastava, *ACS Comb. Sci.* **2017**, *19*, 640.
[151] S.-Y. Kim, S.-I. Kim, Y.-S. Bae, *J. Phys. Chem. C* **2020**, *124*, 19538.
[152] R. Wang, Y. Zhong, L. Bi, M. Yang, D. Xu, *ACS Appl. Mater. Interfaces* **2020**, *12*, 52797.
[153] G. Anderson, B. Schweitzer, R. Anderson, D. A. Gómez-Gualdrón, *J. Phys. Chem. C* **2019**, *123*, 120.
[154] G. Borboudakis, T. Stergiannakos, M. Frysali, E. Klontzas, I. Tsamardinos, G. E. Froudakis, *npj Comput. Mater.* **2017**, *3*, 40.
[155] G. S. Fanourgakis, K. Gkagkas, E. Tylianakis, G. E. Froudakis, *J. Am. Chem. Soc.* **2020**, *142*, 3814.
[156] H. Liang, K. Jiang, T.-A. Yan, G.-H. Chen, *ACS Omega* **2021**, *6*, 9066.
[157] 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.
[158] A. Graves, J. Schmidhuber, in *Proc. 21st Int. Conf. Neural Information Processing Systems*, Curran Associates Inc, Vancouver, British Columbia, Canada **2008**, p. 545.
[159] X. Zhang, K. Zhang, Y. Lee, *ACS Appl. Mater. Interfaces* **2020**, *12*, 734.
[160] S. P. Collins, T. D. Daff, S. S. Piotrkowski, T. K. Woo, *Sci. Adv.* **2016**, *2*, e1600954.
[161] S. M. Moosavi, A. Chidambaram, L. Talirz, M. Haranczyk, K. C. Stylianou, B. Smit, *Nat. Commun.* **2019**, *10*, 539.
[162] K.-J. Kim, Y. J. Li, P. B. Kreider, C.-H. Chang, N. Wannenmacher, P. K. Thallapally, H.-G. Ahn, *Chem. Commun.* **2013**, *49*, 11518.
[163] A. A Talin, A. Centrone, A. C. Ford, M. E. Foster, V. Stavila, P. Haney, R. A Kinney, V. Szalai, F. El Gabaly, H. P. Yoon, F. Léonard, M. D. Allendorf, *Science* **2014**, *343*, 66.
[164] L. S. Xie, L. Sun, R. Wan, S. S. Park, J. A. DeGayner, C. H. Hendon, M. Dincă, *J. Am. Chem. Soc.* **2018**, *140*, 7411.
[165] D. Sheberla, J. C. Bachman, J. S. Elias, C.-J. Sun, Y. Shao-Horn, M. Dincă, *Nat. Mater.* **2017**, *16*, 220.
[166] D. Feng, T. Lei, M. R. Lukatskaya, J. Park, Z. Huang, M. Lee, L. Shaw, S. Chen, A. A. Yakovenko, A. Kulkarni, J. Xiao, K. Fredrickson, J. B. Tok, X. Zou, Y. Cui, Z. Bao, *Nat. Energy* **2018**, *3*, 30.
[167] L. S. Xie, G. Skorupskii, M. Dincă, *Chem. Rev.* **2020**, *120*, 8536.
[168] A. S. Rosen, S. M. Iyer, D. Ray, Z. Yao, A. Aspuru-Guzik, L. Gagliardi, J. M. Notestein, R. Q. Snurr, *Matter* **2021**, *4*, 1578.
[169] R. Batra, C. Chen, T. G. Evans, K. S. Walton, R. Ramprasad, *Nat. Mach. Intell.* **2020**, *2*, 704.
[170] P. G. Boyd, A. Chidambaram, E. García-Díez, C. P. Ireland, T. D. Daff, R. Bounds, A. Gładysiak, P. Schouwink, S. M. Moosavi, M. M. Maroto-Valer, J. A. Reimer, J. A. R. Navarro, T. K. Woo, S. Garcia, K. C. Stylianou, B. Smit, *Nature* **2019**, *576*, 253.
[171] J. A. Gustafson, C. E. Wilmer, *ACS Sens.* **2019**, *4*, 1586.
[172] D. K. Singh, K. S. Krishna, S. Harish, S. Sampath, M. Eswaramoorthy, *Angew. Chem., Int. Ed.* **2016**, *55*, 2032.
[173] M.-M. Titirici, R. J. White, N. Brun, V. L. Budarin, D. S. Su, F. del Monte, J. H. Clark, M. J. MacLachlan, *Chem. Soc. Rev.* **2015**, *44*, 250.
[174] Z. Zhang, J. A. Schott, M. Liu, H. Chen, X. Lu, B. G. Sumpter, J. Fu, S. Dai, *Angew. Chem., Int. Ed.* **2019**, *58*, 259.
[175] S. Wang, Z. Zhang, S. Dai, D.-E. Jiang, *ACS Mater. Lett.* **2019**, *1*, 558.

[176] S. Wang, Y. Li, S. Dai, D. E. Jiang, *Angew. Chem., Int. Ed.* **2020**, *59*, 19645.
[177] C. Zhang, D. Li, Y. Xie, D. Stalla, P. Hua, D. T. Nguyen, M. Xin, J. Lin, *Fuel* **2021**, *290*, 120080.
[178] M. E. Davis, *Nature* **2002**, *417*, 813.
[179] C. S. Cundy, P. A. Cox, *Chem. Rev.* **2003**, *103*, 663.
[180] N. E. R. Zimmermann, M. Haranczyk, *Cryst. Growth Des.* **2016**, *16*, 3043.
[181] M. Shamzhy, M. Opanasenko, P. Concepción, A. Martínez, *Chem. Soc. Rev.* **2019**, *48*, 1095.
[182] K. N. Pai, V. Prasad, A. Rajendran, *ACS Sustainable Chem. Eng.* **2021**, *9*, 3838.
[183] F. Göltl, P. Müller, P. Uchupalanun, P. Sautet, I. Hermans, *Chem. Mater.* **2017**, *29*, 6434.
[184] B. Kim, S. Lee, J. Kim, *Sci. Adv.* **2020**, *6*, eaax9324.
[185] E. H. Cho, L.-C. Lin, *J. Phys. Chem. Lett.* **2021**, *12*, 2279.
[186] Y. Sun, R. F. Dejaco, Z. Li, D. Tang, S. Glante, D. S. Sholl, C. M. Colina, R. Q. Snurr, M. Thommes, M. Hartmann, J. I Siepmann, *Sci. Adv.* **2021**, *7*, eabg3983.
[187] H. Mai, T. C. Le, T. Hisatomi, D. Chen, K. Domen, D. A. Winkler, R. A. Caruso, *iScience* **2021**, *24*, 103068.
[188] J. D. Evans, F.-X. Coudert, *Chem. Mater.* **2017**, *29*, 7833.
[189] D. Schwalbe-Koda, S. Kwon, C. Paris, E. Bello-Jurado, Z. Jensen, E. Olivetti, T. Willhammar, A. Corma, Y. Román-Leshkov, M. Moliner, R. Gómez-Bombarelli, *Science* **2021**, *374*, 308.
[190] M. Moliner, Y. Román-Leshkov, A. Corma, *Acc. Chem. Res.* **2019**, *52*, 2971.
[191] F. Daeyaert, F. Ye, M. W. Deem, *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 3413.
[192] S. Ma, C. Shang, C.-M. Wang, Z.-P. Liu, *Chem. Sci.* **2020**, *11*, 10113.
[193] A. D. Kiadehi, M. Taghizadeh, M. D. Rami, *J. Ind. Eng. Chem.* **2020**, *81*, 206.
[194] M. Fischer, *Sustainable Energy Fuels* **2018**, *2*, 1749.
[195] S. R. Lin, Y. K. Wang, Y. H. Zhao, L. R. Pericchi, A. J. Hernandez-Maldonado, Z. F. Chen, *J. Mater. Chem. A* **2020**, *8*, 3228.
[196] Z. Jensen, E. Kim, S. Kwon, T. Z. H. Gani, Y. Román-Leshkov, M. Moliner, A. Corma, E. Olivetti, *ACS Cent. Sci.* **2019**, *5*, 892.
[197] T. M. Davis, A. T. Liu, C. M. Lew, D. Xie, A. I. Benin, S. Elomari, S. I. Zones, M. W. Deem, *Chem. Mater.* **2016**, *28*, 708.
[198] T. Wu, C. Shu, S. Liu, B. Xu, S. Zhong, R. Zhou, *Energy Fuels* **2020**, *34*, 11650.
[199] A. Corma, J. M. Serra, P. Serna, S. Valero, E. Argente, V. Botti, *J. Catal.* **2005**, *229*, 513.
[200] P. Frontera, M. Miceli, F. Mauriello, P. De Luca, A. Macario, *Catalysts* **2021**, *11*, 1225.
[201] P. Neelamegam, B. Muthusubramanian, *Environ. Sci. Pollut. Res.* **2021**, https://doi.org/10.1007/s11356-021-15962-4.
[202] K. Muraoka, Y. Sada, D. Miyazaki, W. Chaikittisilp, T. Okubo, *Nat. Commun.* **2019**, *10*, 4459.
[203] A. P. Cote, A. I. Benin, N. W. Ockwig, M. O'Keeffe, A. J. Matzger, O. M. Yaghi, *Science* **2005**, *310*, 1166.
[204] Y. Lan, X. Han, M. Tong, H. Huang, Q. Yang, D. Liu, X. Zhao, C. Zhong, *Nat. Commun.* **2018**, *9*, 5274.
[205] C. Desgranges, J. Delhommelle, *J. Phys. Chem. C* **2020**, *124*, 1907.
[206] P. Yang, H. Zhang, X. Lai, K. Wang, Q. Yang, D. Yu, *ACS Omega* **2021**, *6*, 17149.
[207] G. S. Fanourgakis, K. Gkagkas, E. Tylianakis, G. Froudakis, *J. Phys. Chem. C* **2020**, *124*, 19639.
[208] D. Yuan, W. Lu, D. Zhao, H.-C. Zhou, *Adv. Mater.* **2011**, *23*, 3723.
[209] E. S. Sanz-Pérez, C. R. Murdock, S. A. Didas, C. W. Jones, *Chem. Rev.* **2016**, *116*, 11840.
[210] L. Tan, B. Tan, *Chem. Soc. Rev.* **2017**, *46*, 3322.
[211] M. Pardakhti, P. Nanda, R. Srivastava, *J. Phys. Chem. C* **2020**, *124*, 4534.
[212] J. Dean, M. G. Taylor, G. Mpourmpakis, *Sci. Adv.* **2019**, *5*, eaax5101.
[213] M. O. J. Jäger, E. V. Morooka, F. F. Canova, L. Himanen, A. S. Foster, *npj Comput. Mater.* **2018**, *4*, 37.
[214] M. Witman, S. Ling, D. M. Grant, G. S. Walker, S. Agarwal, V. Stavila, M. D. Allendorf, *J. Phys. Chem. Lett.* **2020**, *11*, 40.
[215] S.-Y. Kim, S. Han, S. Lee, J. H. Kang, S. Yoon, W. Park, M. W. Shin, J. Kim, Y. G. Chung, Y.-S. Bae, *Adv. Sci.* **2022**, *9*, 2201559.
[216] A. S. Rich, T. M. Gureckis, *Nat. Mach. Intell.* **2019**, *1*, 174.
[217] B. Weng, Z. Song, R. Zhu, Q. Yan, Q. Sun, C. G. Grice, Y. Yan, W.-J. Yin, *Nat. Commun.* **2020**, *11*, 3513.
[218] M. Andersen, K. Reuter, *Acc. Chem. Res.* **2021**, *54*, 2741.
[219] E. Kim, K. Huang, A. Saunders, A. McCallum, G. Ceder, E. Olivetti, *Chem. Mater.* **2017**, *29*, 9436.
[220] V. Tshitoyan, J. Dagdelen, L. Weston, A. Dunn, Z. Q. Rong, O. Kononova, K. A. Persson, G. Ceder, A. Jain, *Nature* **2019**, *571*, 95.
[221] M. Agrawal, R. Han, D. Herath, D. S. Sholl, *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 877.
[222] J. W. M. Osterrieth, J. Rampersad, D. Madden, N. Rampal, L. Skoric, B. Connolly, M. D. Allendorf, V. Stavila, J. L. Snider, R. Ameloot, J. Marreiros, C. Ania, D. Azevedo, E. Vilarrasa-Garcia, B. F. Santos, X.-H. Bu, Z. Chang, H. Bunzen, N. R. Champness, S. L. Griffin, B. Cheng, R.-B. Lin, B. Coasne, S. Cohen, J. C. Moreton, Y. J. Colón, L. Chen, R. Clowes, F.-X. Coudert, Y. Cui, et al., *Adv. Mater.* **2022**, *34*, 2201502.
[223] F.-X. Coudert, *Chem. Mater.* **2017**, *29*, 2615.
[224] S. C. Weston, B. K. Peterson, J. E. Gatt, W. W. Lonergan, H. B. Vroman, M. Afeworki, G. J. Kennedy, D. L. Dorset, M. D. Shannon, K. G. Strohmaier, *J. Am. Chem. Soc.* **2019**, *141*, 15910.
[225] J.-O. Kim, W.-T. Koo, H. Kim, C. Park, T. Lee, C. A. Hutomo, S. Q. Choi, D. S. Kim, I.-D. Kim, S. Park, *Nat. Commun.* **2021**, *12*, 4294.
[226] J. Li, Y. Tu, R. Liu, Y. Lu, X. Zhu, *Adv. Sci.* **2020**, *7*, 1901957.
[227] B. Burger, P. M. Maffettone, V. V. Gusev, C. M. Aitchison, Y. Bai, X. Wang, X. Li, B. M. Alston, B. Li, R. Clowes, N. Rankin, B. Harris, R. S. Sprick, A. I. Cooper, *Nature* **2020**, *583*, 237.
[228] W. Al Zoubi, M. P. Kamil, S. Fatimah, N. Nashrah, Y. G. Ko, *Prog. Mater. Sci.* **2020**, *112*, 100663.
[229] K. Zhang, X. Han, Z. Hu, X. Zhang, Z. Tao, J. Chen, *Chem. Soc. Rev.* **2015**, *44*, 699.
[230] A. Y. S. Eng, C. B. Soni, Y. Lum, E. Khoo, Z. Yao, S. K. Vineeth, V. Kumar, J. Lu, C. S. Johnson, C. Wolverton, Z. W. Seh, *Sci. Adv.* **2022**, *8*, eabm2422.



Haoxin Mai received his M.Sc. degree in Information Technology and Computer Science from the University of Technology Sydney, and Ph.D. degree in Materials Chemistry from the Research School of Chemistry at the Australian National University in 2019. He has worked at the Royal Melbourne Institute of Technology (RMIT) University as a research fellow from 2019. His research interests include perovskite photocatalysis and photoluminescence, ferroelectric thin films, controllable synthesis of inorganic colloid nanocrystals, and machine learning.

Tu C. Le is a Lecturer at the School of Engineering, STEM College, RMIT University. Prior to joining RMIT in 2017, she worked at the Commonwealth Scientific and Industrial Research Organisation. She completed her Ph.D. at Swinburne University of Technology in 2010. She is interested in material design and development using machine learning algorithms for a broad range of applications such as sustainable energy generation and storage, sensors, and therapeutics.

David A. Winkler is Professor of Biochemistry and Chemistry at La Trobe University, Professor of Pharmacy at the University of Nottingham, and Professor of Medicinal Chemistry at Monash University. He applies computational chemistry, AI, and machine learning to the design of drugs, agrochemicals, electro-optic materials, nanomaterials, and biomaterials. His is a recipient of an ACS Skolnik award, a Royal Australian Chemical Institute Distinguished Fellowship, and a CSIRO Medal for business excellence. He is ranked 174th of 88,000 medicinal chemists worldwide and has written >250 journal articles and book chapters (4 ISI Highly Cited) and is an inventor on 25 patents.

Rachel A. Caruso conducted her Ph.D. studies at the University of Melbourne, Australia in the School of Chemistry. She led research groups at the Max Planck Institute of Colloids and Interfaces, The University of Melbourne, the Commonwealth Scientific and Industrial Research Organisation and now at RMIT University, where she is a Professor in the Applied Chemistry and Environmental Science discipline. Her research interests include controlling the structure of metal oxides, perovskites, and carbon-based materials on the nanoscale, and investigating their properties for application in energy conversion and storage, photocatalysts, and adsorbents.

